UiT The Arctic University of Norway

Faculty of Health Sciences

Department of Psychology

The temporal development of cognitive behavioral therapy as treatment for unipolar depression

An evaluation based on three meta-analyses, focusing on time-trends, effect sizes, and associated moderators.

—

Tom Johan Johnsen

A dissertation for the degree of Doctor Philosophiae - September 2021

Table of contents

Acknowledgements

The starting point of this scientific journey would be around the spring of 2013. As a student entering the final year of my education as a psychologist, I had a deep interest in all clinical aspects of psychology. As such, I spent a lot of time reading and gaining knowledge about the current state and situation with regards to treatment for the largest diagnostic groups of anxiety and depression. Well underway, it gradually became apparent to me that the field was in need of new knowledge and updated approaches in selected areas. The aim was thus set, albeit with total unawareness of the countless hours with (inspired) labor that would follow!

The completion of this thesis owns gratitude to a vast number of people, all cherished and fondly appreciated. Although too many to mention specifically in a brief section such as this, I`d like to give a sincere “thank you very much” to you all!

Extra credit and appreciation goes to UiT the Arctic university of Norway, Department of Psychology. I am truly grateful for being given the opportunity to explore my initial ideas and thoughts at your venue, and for all the support along the way.

I`d also like to express my heartfelt gratitude to my co-authors, without whom this project would have faced an uphill battle; Oddgeir Friborg, for your know-how of the whole field of psychology, as well as vast knowledge of statistics and methods; Jens Thimm, my friend. Your support, wisdom, drive and expertise has been truly invaluable along the way.

This many-year scientific journey, consisting of numerous afternoons, holidays and weekends behind a desk, has co-occurred with the birth and life of my precious and beloved son, Tomas. Your inspiration has provided me with the energy, motivation and life-force necessary to complete a project as comprehensive as this thesis. Thank you for being you!

Abbreviations

BDI	Beck depression inventory
BDI-II	Beck depression inventory version II
BT	Behavior therapy
CBT	Cognitive behavioral therapy
CMA	Comprehensive meta-analysis
CT	Controlled trials
DD	Depressive disorders
ES	Effect sizes
GCBT	Group cognitive behavioral therapy
HDRS	Hamilton depression rating scale
ITT	Intention to treat
MBI	Mindfulness-based interventions
MCBT	Mindfulness-based cognitive therapy
RCT	Randomized controlled trials
RCT-PQRS	Randomized Controlled Trial Psychotherapy Quality Rating Scale
SD	Standard deviation
SMD	Standardized mean difference
US	United States

List of papers

Paper I. Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*, 747-768. doi: 10.1037/bul0000015

Paper II. Friborg, O., & Johnsen, T. J. (2017). The Effect of Cognitive Behavioral Therapy as an anti-depressive treatment is falling: Reply to Ljòtsson et al. (2017) and Cristea et al. (2017*). Psychological Bulletin, 143*, 341-345. doi: 10.1037/bul0000090

Paper III. Johnsen, T. J., & Thimm, J. C. (2018). A meta-analysis of group cognitive-behavioral therapy as an anti-depressive treatment: Are we getting better? *Canadian Psychology/Psychologie canadienne, 59*, 15-30. doi: 10.1037/cap0000132

Paper IV. Thimm, J. C., & Johnsen, T. J. (2020). Time trends in the effects of mindfulness-based cognitive therapy for depression: A meta-analysis. *Scandinavian Journal of Psychology* https://doi.org/10.1111/sjop.12642

Abstract

The present thesis` main objective was to examine time-trends related to CBT as an anti-depressive treatment. A meta-analytic approach was the tool for such investigations, in which the formats of individual, group (GCBT), and mindfulness-based cognitive behavioral therapy (MCBT) were explored from a temporal point of view, and the results subsequently summarized in four articles. The meta-analyses measured the development of effect sizes (ES) with time. Potential moderator variables were also investigated. The primary outcome measure was the BDI, while the HDRS was utilized as a secondary outcome measure.

The results revealed large ESs for all treatment formats, ranging from $g = 0.80$ to 1.89. A significant decline in treatment effects with time was observed across all statistical conditions for individual CBT, while GCBT showed a significant rise in ESs with the passing of time, based on the results from the primary outcome measure. For MCBT, there were found no trends related to the passing of time. Moderator analyses revealed that the fall in ES for CBT was more pronounced for trials where the original treatment manual had been utilized, while the rise in GCBT effect sizes was limited to trials were no specific treatment manual had been used (or reported).

The conclusion is that CBT is efficient in all formats when it comes to battling depression. However, the results also point to the importance of continuous updates or adaptations, in order to be in line with societal developments. This applies for treatment formats, manuals, and outcomes measures alike. Flexibility and frequent updating/adaptation are thus considered keywords to achieve the best possible treatment effects for CBT.

Introduction

Depressive disorders (DDs) can be highly disabling and are ranked third in terms of disease burden as defined by the World Health Organization (WHO, 2014), and first among all psychiatric disorders in terms of disability adjusted life years (Wittchen et al., 2011). In addition, DDs seem to be rising globally (Everyday Health, 2013), and a 20-percent annual increase in its incidence has been predicted (Healthline, 2012). Improvements in treatment methods and prevention measures, and the availability of community psychiatric services are, therefore, as important as ever before. In response, the WHO has prioritized the combating of depression by launching an action plan called "The Mental Health Gap Action Program," aimed at improving mental health services globally (WHO, 2012).

Psychotherapy is a critical asset for dealing with the future challenges associated with DDs: hence, the optimization of existing therapeutic methods and the development of new ones are important clinical research tasks. Cognitive behavioral therapy (CBT) has represented an innovative psychotherapy approach since its introduction more than 40 years ago; it has continuously developed to treat many forms of disorders and overall, it has been highly successful.

When it comes to specific treatment for depression, the cognitive approach has been largely the same for the duration of its existence. However, until 2015 there had been no comprehensive efforts made to investigate CBTs treatment effects with the passing of time. Essentially, no-one knew if contemporary therapy was more effective, or if the effects had waned with time. This is the question that paved the way for the present thesis. Gaining insight of the temporal developments within a treatment form would potentially generate valuable knowledge, not only for CBT, but also for psychotherapy in general. The findings such investigations reveal, could lead to new understandings regarding the mechanisms that influence treatment effects. An investigation and evaluation of treatment effects could thus

contribute to the direction both clinicians and researchers essentially strive for; making contemporary and future psychological treatments more effective.

Depressive disorders

Depressive disorders are characterized by a set of associated symptoms, consisting of affective, cognitive, emotional, and behavioral changes that occur when the disorder is active. The most prevalent and common symptoms associated with major depression are loss of pleasure, loss of interest in activities, depressed mood, loss of energy, and exhaustion/fatigue. It is also common for patients to experience a variety of other associated symptoms, the most usual being suicidal thoughts, sleep disturbances, negative evaluations about themselves and the future, reduced appetite, irritability, poor memory and concentration, depressive rumination and worry, loss of sexual interest, and feelings of guilt and/or punishment. Individual variations in the expressed symptoms are common. Depression without mania/hypomania is often referred to as “unipolar” because the mood remains in one emotional state, in contrast to bipolar disorder.

A significant minority of patients are afflicted by atypical depression, which shares many of the typical symptoms of major depression but is often characterized by improved mood in response to positive events, as well as weight gain or an increased appetite, and hypersomnia. Another frequently occurring variation in the expression of unipolar depression, is melancholic depression, characterized by a loss of pleasure in most or all activities, a failure of reactivity to pleasurable stimuli, a worsening of symptoms in the morning hours, early-morning waking, psychomotor retardation, and excessive weight loss. There are also other subtypes of unipolar depression, where the current situation and/or environment is associated with the condition. Seasonal affective disorder is a form of depression in which depressive episodes typically come on in the autumn or winter, and resolve in spring, while postpartum

depression refers to the intense and sustained depression experienced by women after giving birth or while a woman is pregnant. For premenstrual dysphoric disorder, the associated symptoms are periods of anxiety, depression, or irritability in the week or two before a woman's menstruation (DSM-V, 2013). Some patients experience a more chronic presence of depressive symptoms, typically in a milder form. This subtype is referred to as dysthymic depression. A substantial number of patients experience the recurrent nature of depression, as approximately 6 out of 10 have a relapse (Solomon et al., 2000). Many of these are diagnosed with the subtype "recurrent depression", characterized by frequently occurring episodes of depression, where full remission is usually achieved between episodes. There are other subtypes of depression as well (for example catatonic), but these are more rear.

Clinical unipolar depression is typically categorized according to severity, with the ranges being mild, moderate, and severe depression. The differentiation is made by both severity and the number of experienced symptoms. Psychotic symptoms are associated with severe depression, as is usually suicide attempts. The research underpinning the current thesis, and its associated papers, focus on patients with unipolar depression of all clinical ranges, without psychotic symptoms.

Major depression is among the most prevalent psychiatric disorders in the western world, with an estimated global prevalence of 4.4% (WHO, 2017). Fourteen percent of patients with major depression have the illness for over five years (Patten, 2006), with an average duration of almost 10 years (Friborg et al., 2014). A crippling consequence of depression is its recurrent nature, which means suffering is reinstated for the majority, many of whom do not return for more treatment (Andrews, 2001). Moreover, for those experiencing two or more episodes, depression may develop into a chronic condition (Blanco, 2010). The costs of depression are hence substantial for society (Sobocki, 2006; Mental Health

Foundation, 2010). Regular assessments of treatment effects, as well as efforts to identify the most efficient treatments, are therefore important.

Unipolar depression as a psychological diagnosis has generated a vast amount of research over the last decades (Cuijpers, 2017). This massive focus has not decreased with the shift towards requirements of evidence-based treatment forms. Rather, recent scientific evolution put emphasis on the invention of highly accurate tools for measuring levels of depression and treatment effects.

Diagnostic methods and criteria

The modern method of diagnosing depressive disorders according to a set of associated characteristics, or symptoms, had its origin with the invention of the Diagnostic and Statistical Manual of Mental Disorders, first edition (American Psychiatric Association [APA], 1952). In the same period, the World Health Organization (WHO) revised their diagnostic manual, the International Classification of Diseases, Injuries and Causes of Death (ICD) with the intention to incorporate an official, international classification of mental disorders. The result was the ICD-7 (WHO, 1955). However, the reception of these early classifications of mental disorders did not achieve widespread international acceptance, leading to the subsequent developments of the DSM-II (APA, 1968) and ICD-8 (WHO, 1965) in the late 1960`s (Gruenberg, Goldstein & Pincus, 2005). The revision processes resulted in two similar diagnostical systems. However, researchers in the field of depression were not satisfied, stating that the lack of reliable and clear diagnostic criteria was problematic. Thus, in the early 1970`s the development of explicit diagnostic criteria was led by American researchers, resulting in a classification system named the Research Diagnostic Criteria (RDC). These efforts eventually culminated with the much-changed DSM-III (APA, 1980).

It was against this background modern cognitive behavioral therapy, as we know it today, developed as an anti-depressive treatment. The diagnosis of depression was initially based on the criteria proposed by the RDC, which were subsequently incorporated in the DSM-III. The new version had major changes as compared to its predecessor, the DSM-II, with the implementation of clear diagnostic criteria for depression being one of the more significant differences. The new version also represented a move away from the traditional labels of neurotic depression and depressive reaction, replacing them with the term major depression. Thus, if patients met certain criteria, or symptoms, associated with unipolar depression, they reached the diagnosis of major depression, a term which is still in use as of today. To reach a diagnosis of depression, patients had to be experiencing a loss of interest or depressed mood nearly every day for a period of two weeks, in addition to experiencing four out of eight symptoms associated with major depression.

The diagnostic criteria for depression remained quite similar through the revisions made in 1987, with the DSM-III-R (APA, 1987), and in 1994 with the introduction of DSM-IV (APA, 1994). Relevant changes consisted of moving to a requirement of meeting five out of nine symptoms associated with depression, including the experience of dysphoric mood or depressed mood or loss of interest or pleasure nearly every day most of the day during a 2-week period. The other criteria consist of; weight change, sleep disturbances, psychomotor change, diminished ability to think or concentrate or indecisiveness, fatigue, feelings of worthlessness or guilt, and recurrent thoughts of death or suicidal ideation. The diagnostic criteria for unipolar depression remained intact for the latest revision of the DSM-V (APA, 2013), with the added emphasis on functioning (depressed patients must experience difficulties maintaining their everyday functioning) representing a minor adjustment.

The major changes implemented in DSM-III was not followed by similar changes to the ICD, which largely followed the previous templates with the publication of ICD-9 (WHO,

1975). The system still did not include definite clinical diagnostic criteria, although attempts in that direction was made with the clinical modification ICD-9-CM (WHO, 1978). One of the objectives were to make clinical descriptions for depression more precise, but the definitions were not as defined or specific as the DSM-III`s. The higher rate of precision and clinical focus offered by the DSM-III contributed to researchers and authors of clinical trials and methods to prefer this system when using diagnostic procedures. The fact that it originated in America probably also played a role, especially for cognitive behavioral therapy, which has its modern foundations in the US. In 1992 the ICD-10 (WHO, 1992) was released. It now incorporated similar descriptions of the depressive syndrome and characteristics as the DSM-IV, but still lacked the distinct clinical criteria. As such, most clinicians, and researchers sticked to the more user-friendly American counterpart. This impression is confirmed when reading and evaluating the published clinical CBT-trials for depression: The two most utilized methods for including participants in a trial, are either by following the criteria outlined in a version of the DSM, or to use the cut-off-scores collected from self-evaluation inventories measuring symptoms for depression.

In summary, it may be concluded that the diagnostic criteria for unipolar depression have not experienced radical changes since the mid 1970`s for the DSM, and that this system was (and still is) the preferred diagnostic tool for CBT-trials. More pronounced changes have occurred for the ICD, with significant changes implemented with version 10. There have also recently been made adjustments with regards to the depression diagnosis with the release of the ICD-11 (WHO, 2019). However, given the recency of its release it is considered highly unlikely that any of the trials included in the current meta-analyses are affected by these later changes.

The psychodynamic approach

The recognition of mood-disturbances afflicting humans dates to the very beginning of civil life. The term melancholia comes from the ancient Greek, who used such a terminology to describe the personality or traits of the typical depressed person. The Greek considered melancholia to be the result of having too much black bile, a hypothesized bodily fluid, in the system. And, while modern understanding of the depressive syndrome does not consider bodily fluids to be influential, the term melancholia stayed relevant also with the development of psychodynamic theory.

Historically speaking, recognized and systematic psychological treatment of depression started with Freud`s psychoanalytic theory. Freud (1917) argued that depression could have many causes, including biological factors, overactive demands from the super-ego, and inwardly directed anger. At the same time, he also believed depression (melancholia) and grief were both linked to the loss of an important relationship or rejection by a parent and was produced via a psychological reaction from the patients' part. The tool to treat the disorder was the psychoanalysis. In the following decades, psychodynamic theory dominated the field of depression treatment, with the work of, among others, Adler (1927), Jung (1964),) and Erikson (1950) developing and refining the psychodynamic approach.

The psychodynamic branch was founded on a comprehensive theoretical base. The ability to explain nearly every aspect of human emotional experience, and frame it within a theoretical understanding, was considered a major advantage with the approach. Psychodynamic theory offered a completely new, complex, and exciting view on psychiatric disorders and mood-disturbances, which was appealing to the public. The movement also gave hope to the many sufferers of such illnesses, as well as their closest family and friends. With the forthcoming of the psychoanalysis, the main therapeutical tool for psychodynamic

theory, there was finally a tangible way to get better, through an approach that did not include being locked up or sent away.

Psychoanalysis for the depressed patient largely consisted of helping the patient gain insight into the origins of their emotional problems, thereby gaining access to solve the issues. The process consisted of bringing unconscious experiences (thoughts, memories, urges, and episodes) into conscious awareness. Via this process catharsis, or emotional release, would occur, and the patient would find relief from psychological distress.

Although psychodynamic theory did not enjoy unanimous support in the field of psychology, the psychoanalysis remained the main psychological treatment for depression for decades. However, its position became increasingly challenged from the 1950`s. Objections related to the lack of scientifical evidence for its treatment effects, and to the theoretical and clinical foundations, which were largely based on case-studies. It also became apparent that the psychoanalytic approach, with its focus on hidden mechanisms, drives and the unconscious, did not suit every patient. The timespan of treatment would also be of concern, with therapy often going on for years. As such, researchers and clinicians started to investigate other possible approaches to the treatment of depression, and mood disorders in general. However, psychodynamic therapy today remains a relevant and frequently used treatment for depression, especially with the developments of the less time-consuming, short-term forms of psychodynamic therapy.

Behavioral therapy

The first serious alternative to psychodynamic therapy as a treatment for depression was offered through the developments of behaviorism. Behavior therapy (BT) became a dominant force in the late 1940`s and throughout the 1950s, drawing from the work of, among others, Skinner (1953), Wolpe (1958) and Eysenck (1960). In the aftermath of World War II,

there were many veterans experiencing trouble with their emotional adjustment. As such, there was a need for an effective short-term therapy for anxiety and depression, a task which the psychoanalysis was neither designed nor equipped for.

The role of environmental cues in influencing the acquisition and maintenance of behavior is emphasized in BT. Conventional BT models for treating depression focus attention on increasing access to pleasant events and positive reinforcers and decreasing the intensity and frequency of events and consequences deemed to be unpleasant/negative (Lewinsohn, 1972). Although more frequently used for the anxiety disorders, the clinical approach of BT for depression also consisted of exposure training. If any neutral (objectively deemed as non-threatening) situation created feelings of anxiety or depression, the aim was for the patient to stay in the situation, or to repeat exposure until the troublesome emotions waned.

Behavioral therapy offered a more pragmatic and simplistic view on psychological disorders, as compared to the more complex psychodynamic therapy. Essentially, behaviorism was based on the idea that behaviors can be measured, trained, and changed. It proposed that our responses to environmental stimuli determine our behavior. Unconscious mechanisms were not seen as relevant, neither were drives and urges. The method initially had a broad appeal with the public, as it was easy to understand, and did not involve large investments in the form of money or time. Typically, treatment lasted for less than 20 sessions. The progress of behavioral therapy is by many considered to be the first wave of CBT, since the latter form of therapy also has a significant focus on behavioral mechanisms.

While behavioral therapy in its purest form often led to acceptable results for patients suffering from anxiety, the evidence was not overwhelming when it came to treating depression. And, with the advances of cognitive therapy in the 1970s, BT approaches based purely on operant and respondent principles widely became regarded as insufficient in the

treatment of depression. A broader approach was sought by researchers and clinicians, one which could encompass both behavioral and cognitive principles in the treatment of psychological disorders.

Cognitive behavioral therapy

One major objection made towards both psychodynamic therapy and behavioral therapy was that the processes of conscious thinking, evaluation, appraisal and judgment of situations and stimulus largely were bypassed. From the early 1960`s researchers and clinicians gradually increased focus towards the influence rational thinking exerted on mood disturbances, including depression.

Clinical therapist Alfred Adler was perhaps the first to point towards cognition in psychotherapy, with his notion of the impact basic mistakes had in creating unpleasant emotions for clients. However, the earliest recognized form of cognitive behavioral psychotherapy was Albert Ellis` rational emotive behavior therapy (Ellis, 1957)). It`s underlying theory states that clients emotional distress arises from their thoughts about an event rather than the actual event itself.

In the same period as Ellis did his work, psychiatrist Aaron T. Beck was working at the University of Pennsylvania. Having studied and practiced the psychoanalysis, Dr. Beck designed and carried out several experiments to test psychoanalytic concepts of depression. Fully expecting the research would validate these fundamental concepts, he was surprised to find the opposite (Beck Institute, 2021).

As a result of his findings, Dr. Beck began to look for other ways of conceptualizing depression. He found that depressed patients experienced streams of negative thoughts that seemed to arise spontaneously. He called these cognitions “automatic thoughts.” He found that the patients’ automatic thoughts fell into three categories; The patients had negative ideas

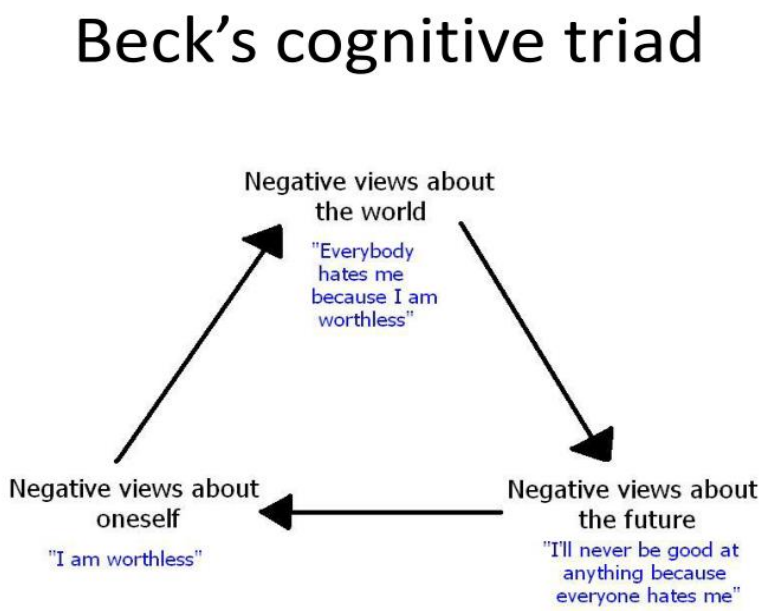
about themselves, the world and/or the future. Dr. Beck began helping patients identify and evaluate these automatic thoughts. He found that by doing so, patients were able to think more realistically. As a result, they felt better emotionally and were able to function more adequately. Further, when patients changed their underlying beliefs about themselves, the world and other people, therapy resulted in long-lasting change. Dr. Beck called this approach "cognitive therapy.", and it has also become known as "cognitive behavior therapy," or "CBT." As a result of his work, Dr. Beck is commonly regarded as the founder of CBT.

The pioneering work of Beck and Ellis paved the way for a new direction in psychotherapy. The empirical and clinical support for this "second wave" of CBT was solid and led to a change of paradigm in psychology, with the stage set for cognitive behavioral therapy to become the dominant force in the treatment of a range of psychological disorders. This is a standing CBT still enjoys in contemporary theory and treatment for depression, although other directions of therapy have challenged the cognitive approach, with behavioral activation (Lejuez et al, 2001), interpersonal therapy (Weismann et al, 1981) and short-term psychodynamic therapy (Luborsky et al, 1995) all proving to be effective in the treatment of depression (Cuijpers, 2017).

Theoretical and clinical principles

The CBT method refers to a class of interventions sharing the basic premise that mental disorders and psychological distress are maintained by cognitive factors or cognitive processes (Hofmann et al, 2012). As posited by Beck (1970) and Ellis (1962), maladaptive thoughts maintain emotional distress and dysfunctional behavior, for which alleviation or cure is realized by changing them. When it comes to depression, the cognitive model postulates three specific concepts to explain the psychological substrates of depression (Beck, 1979). The first concept is named the "negative cognitive triad" and has been the main theoretical

base for understanding and treating depression. The triad involves automatic, spontaneous, and seemingly uncontrollable negative thoughts about the self, the world or environment, and the future. These automatic thoughts largely exist outside conscious awareness and are major determines in how people think and behave. Often, people tend to think and behave in ways that enhance or maintain their negative automatic thoughts. This is especially the case for depressed patients.

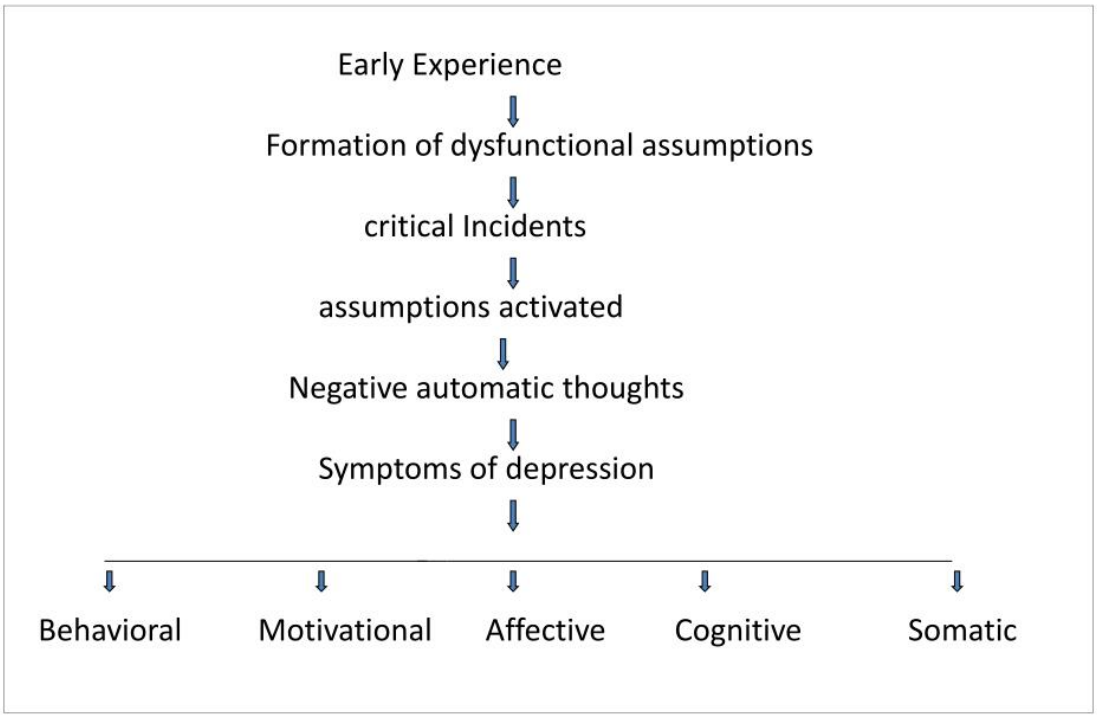


The automatic thoughts are thought to be the cognitive manifestation of fundamental assumptions embedded in a patient's belief system, constructs often defined as "schemas". Schemas are the second ingredient in the cognitive model of depression. These basic assumptions are viewed as either adaptive, in which they promote good functioning and wellbeing, or maladaptive, exerting a dysfunctional and dysphoric effect. Cognitive theory states that basic assumptions, or schemas, form a personal matrix of meaning and value, and is the backdrop against which everyday events acquire relevance, importance, and significance. The developmental period is thought to be the arena where people's basic

assumptions are shaped and formed, in which everyone learns rules or formulas by which they attempt to make sense of the world. That is, how they set goals, how they evaluate and modify behavior, and how they understand events in their lives.

The third key concept in the cognitive model is faulty information processing. This term refers to the depressed patient`s systematic errors in thinking, that maintain the persons maladaptive schemas, despite the presence of contradictory evidence. Such faulty processing consists of overgeneralization (drawing a general rule based on one incident), selective abstraction (focusing on a detail while ignoring the bigger picture), arbitrary inference (drawing a specific conclusion in the absence of solid evidence), personalization (the tendency to relate events to oneself), absolute thinking (black/white dichotomous view), and magnification and minimization, which is the tendency to make wrong estimation of the effects of events. Dr. Beck understood these faulty processes to be a thinking disorder in depressed patients and defined the concepts as “primitive” or “mature” modes of organizing reality. (Beck et al, 1979). A depressed person typically organize reality in accordance with the principles of a primitive organization of information.

Thus, in cognitive theory the patient’s thought processes are viewed as most important for emotional wellbeing. By identifying and evaluating automatic thoughts about himself, the world, and the future (the cognitive triad), patients can think more realistically. As a result, they feel better emotionally and can behave more functionally. Further, when patients change their underlying core beliefs (schemas) about themselves, the world and other people, therapy can result in long-lasting change. The ability to maintain and enhance objective, realistic thinking and adaptive schemas will be enhanced by moving to a mature mode of organizing reality. The aim of CBT is thus both to achieve remission from symptoms by changing the automatic negative thoughts, and to achieve deeper change and gain resilience from new episodes of depression, by adjusting the core belief system of the patient.



Treatment manual

Cognitive behavioral therapy in its current format has, by and large, been the preferred choice for individual treatment of depression since the late 1970`s. The timespan of its existence is thus large, with over 40 years. During this time, the cognitive wave has become a powerful unit, both within its field of profession, and politically. The specific treatment for depression has not changed much during this time span though, and still lends heavily to the original ideas of Beck, and his original treatment manual from 1979 (Beck et al., 1979). In fact, over half of the trials included in the meta-analysis for paper I in the current thesis, stated that they applied and followed the original manual.

The broad acceptance and use of a set manual yield an advantage when it comes to examining temporal trends in treatment effects based on CBT trials, as the degree of

standardization over the years is high. The net result is improved heterogeneity, and increased validity and reliability for statistical analyses. There are also potential large clinical benefits; Patients are receiving a well-tested and documented form of therapy, information about the treatment is easily accessible, and therapists have solid experience in delivering therapy. However, there also exist potential downsides. The use of a set manual could leave limited room for therapeutic freedom, creativity and adaptation to the needs of the specific patient and circumstances. Development over time do not occur naturally, or automatically, when following a set manual. Instead, practitioners and patients are reliant on officially implemented and published changes to the treatment manual, which in practice do not happen too often. The net result could be a more rigid form of treatment, which is unlikely to be helpful for every patient. Perhaps such factors are part of the reasons why clinical trials over the effect of CBT consistently show a considerable number of dropouts during treatment (Fernandez et al, 2015), and have a substantial number of patients classified as non-responders (Samaan et al, 2021).

Treatment description

The following section aims to provide a clinical and theoretical description of each step in traditional cognitive behavioral therapy. To assist such purposes, a typical course of treatment for the depressed patient entering cognitive behavioral therapy is presented. The illustrated case is an example of CBT for depression. However, it is important to bear in mind that cognitive therapists are encouraged to adjust their approach according to the needs of the individual patient. Such adjustments may, for example, relate to the amount of time devoted to the different steps in therapy. Some patients, especially the more severely depressed, may require an extended focus on functioning and behavioral techniques, while others may require more time on strengthening the alliance or the release of emotions. The most important

objective for cognitive therapists, is to adequately cover the different areas deemed important in CBT for depression, such as behavioral techniques, the working alliance, education, the review of automatic thoughts, and pinpointing maladaptive beliefs. The order or doses of each ingredient do not necessarily need to follow a set formula but may rather be tailored to the client's depression.

The illustrated case report (in italics) is based on an example from the original cognitive behavioral manual (Beck et al, 1979), and the intention is to provide the reader with a bridge from theory to clinical practice.

The initial interview

This is the first meeting between a therapist and the client, who is typically referred to psychological treatment by his/her general physician. The objective is to establish rapport (also coined the therapeutic alliance), perform a diagnostic assessment, and gain insight in the patient's background, medical history, social network, as well as personal assets for therapy. Further, the therapist aims to collaborate with the patient to identify which problems to focus on initially in therapy, also labelled as target symptoms. The symptoms are formulated in cognitive terms, defined as maladaptive thoughts or attitudes, and the therapist provides an initial education about the cognitive approach to therapy. The therapist also performs an assessment of suicidal risk, elicits feedback about the patients experiences and thoughts about the first session, and plans the next step in a therapeutic course.

Karen is a 36-year-old married woman, working part-time in a sales-company. She has 3 children aged 7,9 and 14 and has been in a stable relationship with her husband for 16 years. In the initial interview, she describes herself as a person who "can`t do anything right", "is a failure as a wife and mother", and "a burden to her family". Karen often thinks about suicide to unburden her surroundings. In the initial interview, the therapist and patient

decide it could be helpful for her to enter a standard, 22 session weekly treatment course of individual CBT for depression.

Sessions 1 to 3

The agenda for the first sessions typically involves gaining a thorough knowledge and awareness of the patient's depression and review the patients' symptoms, as indicated by ratings in the BDI, which is filled out before the first session (and often before every session). In addition, focus is on behavior and activity levels, which are reviewed. Education about the cognitive approach and ideas continue, and the relationship between thinking, behavior and affect is demonstrated. Homework is considered an essential component of CBT, and for the first sessions the tasks typically consist of filling out forms related to broader psychological functioning, reading passages or leaflets about depression and CBT, as well as following an established plan for appropriate activity levels. The patient is also encouraged to define problems she sees as contributing to her depression, and record her cognitions during periods of sadness, anger, or apathy, in order to elicit awareness of the relationship between thinking, behavior and affect.

In the initial sessions, Karen presented thoughts about herself as being "selfish", "thinking as a child" and "insignificant in the household". She also criticized herself for "not doing what I should be doing" and related this to her inability to get her youngest son out of bed. As part of the activity scheduling, Karen and the therapist decided a new approach which involved giving the boy the responsibility for getting out of bed himself. This behavioral intervention proved successful, and thus alleviated some of the concerns, and provided motivation for further adjustments in activity levels and behaviors. As activity levels and motivation seemed to be at satisfactorily levels, treatment started to focus more directly on the patients' cognitions.

Sessions 4 to 8

After behavioral activation has been addressed, the next step is typically to focus on the identification and discussion of specific cognitions that leads to negative affect. These cognitions consist of automatic and negative thoughts about the client, the world, and the future (the cognitive triad). The cognitions are reviewed and challenged, with the aim of recognizing and correct errors in thinking. Homework consists of recording negative cognitions (automatic thoughts), and, if possible, record alternative thoughts or explanations.

Karen thought she was "a bad mother, without any benefit to her children". She admitted these thoughts led to harsh levels of self-criticism. The therapist asked why she had such thoughts, what was her evidence? Karen replied, "the boys hate me", "my daughter doesn`t respond to me", and "my husband never gives me any feedback about my parenting". The evidence was reviewed and challenged, and alternatives were identified and discussed. The discussions revealed that her boys had never expressed any words of hate or strong disapproval towards their mother. Further, Karen recognized it was not unusual for teenage girls to occasionally not respond to their mother and admitted that her daughter most of the time was responsive. The patient also deemed it plausible that the lack of feedback from her husband could be because he had no larger objections towards her parenting. As a result of these more realistic reviews of her automatic thoughts, Karen felt less sad and critical of herself.

Sessions 9 to 17

After reviewing and challenging automatic thoughts, the next step in therapy is to address the underlying assumptions that assist in maintaining negative thoughts. The therapist helps identify such maladaptive attitudes, which consists of deeply embedded beliefs. Homework in this stage of therapy typically involves elements that may counter the effects of

the maladaptive attitudes, for example making lists where positive actions and thoughts are written down, memorized, and practiced in real life settings.

In therapy, Karen recognized that her tendency to criticize herself was driven by strict rules for behavior and functioning that she had embedded in her personal understanding of herself. These "should do`s" were almost impossible to satisfy, thus leaving her feeling worthless and inadequate. In addition, her strict internal rules led her to not focus on her own wants and needs, thus preventing activities and thoughts that may enhance positive feelings. Karen's rules of "I should prioritize other people's needs" and "I should perform all tasks perfectly" were thus identified as maladaptive beliefs and challenged and corrected over several sessions. As she gradually practiced not doing every task in a perfect manner and identified and prioritized doing some of the stuff she wanted, she gradually experienced less thoughts of self-criticism.

Sessions 18 to 20

At this stage of therapy, the main interventions have been well practiced, and the patient is becoming more adept in identifying and challenging maladaptive thoughts and assumptions by him/herself. Patients often feel more energetic, less depressed and are ready for functional changes in their lives. Thus, the focus of therapy turns to recognizing and implement new (and adaptive) skillsets, hobbies, occupations, educations, or other activities that are considered helpful. At the same time, it is important to be aware of a return of the problematic areas, as the endeavor of new tasks often makes challenges resurface. Homework may consist of identifying and discussing future plans with friends or families, and to follow a structured plan for implementing the desired actions.

As Karen was getting more acquainted to her own needs and preferences, she gradually recognized a desire to start playing football, a sport she had enjoyed in her youth.

She discussed this with her husband, who was supportive, and together they figured out the practical aspects of such an endeavor. As the first training approached, some of Karen's negative thoughts resurfaced, and she criticized herself for being selfish and not prioritizing more important tasks. These thoughts were addressed in therapy, and between sessions by Karen herself. Although not completely gone before the first training session, her maladaptive thoughts were reduced to an extent that allowed her to attend.

Sessions 21 and 22

The final sessions of therapy focus on consolidating the treatment progress and preventing relapse. This is achieved through revisiting the previous stages of therapy and make sure the patient has established a sound platform to continue monitoring and reviewing his/her maladaptive thoughts, should they resurface. Relapse is sought to be prevented via enhancing the patient's belief in a sustained outcome, reassure the patient that he/she has the availability of further boosting sessions if things get bad, as well as scheduling follow-up sessions at timely intervals. The tapering off the last sessions is also seen as helpful to prevent relapse, and the final sessions should thus be scheduled twice a month, rather than weekly.

Towards the end of therapy, Karen felt much better and reported not being depressed anymore. This was reflected in the BDI, where her score now was 7, which is at a non-clinical level of symptomology. She expressed concerns about no longer having the opportunity of frequent treatment sessions but was largely reassured with the scheduling of follow-ups at the intervals of 1, 2, and 6 months.

Follow-up

In follow-up sessions, focus is typically directed to assess the present situation, and evaluate if there has been regression with regards to any part of the previous problematic

areas. The patient`s continued ability to use treatment tools is also checked, and rehearsal implemented if deemed necessary. Treatment is typically terminated after two or three follow-up sessions.

Karen remained nondepressed throughout the follow-up period and noted with pleasure that she generally felt more confident and was more relaxed in her role as a parent and wife. Her score on the BDI was 5 at the last session of follow up. Her previous automatic thoughts still occasionally resurfaced, but Karen had now incorporated techniques that enabled her to fend of such thoughts quite rapidly. She felt ready for terminating therapy.

Treatment effects and further developments

A large amount of research has confirmed the efficacy of classical CBT in treating depression. Meta-analyses published in the 1980s (Dobson, 1989), the 1990s (Hollon, Shelton, & Loosen, 1991), and after 2000 (Cuijpers et al., 2008), concluded that CBT had a high treatment efficacy. Effect sizes (ES) for individual CBT have recently been proven to be strong at post-treatment, with ranges of g = 1.37 to 1.89 observed across different analytic conditions (Johnsen & Friborg, 2015; Cristea et al, 2017). Cognitive behavioral therapy has often (but not universally) been found to maintain substantial parts of its treatment effects. Follow-up at 1-year typically reveals relatively comparable scores as those measured at post-treatment (e.g., Kovacs et al, 1981; A-Tjak et al, 2021). The magnitude of CBT's enduring effects has also been found to be at least as large as keeping patients on continuation anti-depressive medication (Hollon et al, 2005; Dobson et al, 2008).

Later variations of the cognitive method have been developed, building on the foundations of CBT. The most well-known are CBT combined with mindfulness (Segal, Williams, & Teasdale, 2002), integrated cognitive therapy with elements of interpersonal therapy (Castonguay, 1996), and meta-cognitive therapy (Wells, 2000), which represent

further innovations in CBT. These newer forms of CBT, typically referred to as the third wave of cognitive therapy, have exhibited promising efficacy in clinical trials, however, few studies have demonstrated these innovations to be significantly more effective in treating DDs than classical CBT (e.g., Ashouri et al., 2013; Manicavasgar, Parker, & Perich, 2011).

Treatment fidelity and adherence

Therapists conducting CBT are expected to have good knowledge of the general cognitive principles, as well as having the necessary therapeutic skills to translate knowledge into practice when delivering therapy. These skills are usually achieved via specific training and guidance, or through formal education and courses. The main goal with such training is to ensure that therapy is conducted as closely in line with the cognitive principles as possible, and thus not consisting of any approaches, angles, interventions, or elements that are not cognitive oriented. Purity in treatment is the objective. A treatment purely consisting of elements derived from CBT is thus considered to have the highest degree of treatment fidelity, or integrity.

Treatment fidelity and treatment adherence are often used interchangeably as terms relating to the therapists` ability to follow the treatment manual as carefully as possible. When a therapist precisely follows each step of the manual, through each session of therapy, in the manner described by the manual, adherence is at its highest.

For researchers and clinicians affiliated with the cognitive branch of psychotherapy, treatment fidelity and adherence to the manual are traditionally stressed to be factors of high importance for therapeutical outcome. The general opinion is that higher degrees of fidelity and adherence correlates with better outcomes for the clients. As such, different scales have been developed to measure fidelity and adherence, which researchers are encouraged to use when conducting clinical trials. One of the earliest, and most frequently utilized scales, is the

cognitive therapy scale (CTS: Young & Beck, 1980). However, despite available scales, there has been a limited amount of research on the correlation between the degree of adherence/fidelity and treatment outcomes, or effect sizes. No meta-analyses have attempted to quantify and measure this relationship precisely. The most probable reason for this gap relates to the lack of a standardized procedure regarding which scales to use, and how to adequately report results from research trials.

The lack of quantifiable data was indeed the main reason why the meta-analyses in the current thesis could not precisely investigate the relationship, and potential interaction, between adherence, effects sizes and time for publication. However, for individual cognitive therapy, paper I (Johnsen & Friborg, 2015) found that there was no difference in outcome for trials that did use a measure of adherence and fidelity, versus trials that did not use one. This finding does however not give a clear answer as to whether higher degrees of fidelity are related to better outcomes.

Group cognitive behavioral therapy

Group psychotherapy is a common treatment modality for many disorders, including depression. Group therapy may be defined as a meeting of two or more people (usually six to eight persons) working toward a common therapeutic goal. Group cognitive behavioral therapy (GCBT) is based on traditional cognitive therapy (Beck et al., 1979) and typically includes elements such as case formulations, reviews of automatic thoughts and maladaptive beliefs, and analyses of antecedent events or situations and associated behaviors and cognitions with fellow group members, all of whom contribute with their personal experiences and point of views. In addition, homework, conceptualization of problems, and group attention to self-defeating beliefs are common elements within this format.

GCBT gained popularity as a treatment for depression in the late 1970s and showed a promising degree of efficacy (e.g., Beck et al., 1979; Shaw, 1977). Although the focus in modern psychotherapy, by and large, has been the individual format, the increasing number of studies during the last decade on group therapy as a viable anti-depressive treatment (e.g., Hans & Hiller, 2013; Huntley, Araya, and Salisbury, 2012) indicates a shift in interest towards the group format. Group cognitive behavioral therapy is considered short-term, with treatment courses typically ranging from 8 to 20 sessions.

Cognitive treatment methods and principles are clearly defined and well understood by researchers and clinicians, hence facilitating homogeneity with regards to the use of treatment techniques, ingredients, and assessment measures across studies. This methodological allegiance allows valid and reliable comparisons of effect sizes for GCBT interventions. The effectiveness of GCBT for depression has been demonstrated in several meta-analyses (e.g., Huntley et al., 2012; McDermut, Miller, & Brown, 2001; Okumura & Ichikura, 2014). Recent estimates of ESs have revealed ranges of g = 1.14 to 1.56 across statistical conditions (Johnsen & Thimm, 2018). Studies comparing individual versus group CBT for depression have typically reported the latter as less expensive (Tucker & Oei, 2007; Vos et al., 2005) but not necessarily less effective (Hans & Hiller, 2013; Khoshbooii, 2012) than the former. A recent meta-analysis (Burlingame et al., 2016) showed no differences in outcome between individual and GCBT for depression in studies in which the treatment, the patients, and the doses (number of sessions) were identical. Findings like these highlight the potential importance of implementing group psychotherapy formats in addition to the individual treatment format. However, although individual and group CBT share the same treatment philosophy and therapeutical techniques, there are differences between the two formats that may influence the effects of these treatment modalities, necessitating the need for separate

analyses to optimize statistical reliability and validity. For example, group cohesion and normalization are potential therapeutic factors that are specific to GCBT (Whitfield, 2010).

There exist large variations when it comes to using a treatment manual in GCBT. Trials have typically either invented their own manual, modified existing treatment protocols, or been performed without following a set manual. This does not necessarily mean that adherence to the general cognitive principles is compromised but implies that measures of adherence to a set manual are difficult to implement, or not applicable. Also, with such a varied approach to therapy within the field, thorough and precise statistical evaluations of the relationship between treatment fidelity and ESs are difficult to perform.

Treatment course

Although several researchers and clinicians have developed manuals for conducting GCBT, the ingredients and techniques are similar and follow cognitive theoretical principles. The variations usually concern how much time (number of sessions) that is devoted to each element, the order in which the cognitive techniques are presented, which themes to focus on, and how much time is allocated to the group conversations. The two latter variables largely depend upon the characteristics and setting of the group. In the following text, a typical course of GCBT for depression is described, based on a tested manual (Schaub et al, 2006).

Sessions 1-3 are predominantly of psychoeducational character. The ingredients here consist of an introduction to CBT, the cognitive principles, and the therapeutical concept, as well as education about symptoms, and a focus on how to behave and not to behave in the group setting. Information about antidepressants, medication and alternative forms of therapy is provided, and the vulnerability stress model is presented.

Sessions 4-6 focus on behavior and behavioral activation. Information is provided about the vicious cycle of depression and the depression spiraling effect, and the balance and implementation of activities are discussed. Positive activities are planned and structured.

Sessions 7-10 tunes into the cognitive part, with information provided regarding the cognitive triad of depression and the cognitive model. Cognitive distortions and the process to change them are discussed, and depressive rumination thematized in the group setting. Core beliefs are discussed, identified, and sought changed.

Sessions 11 and 12 focus on the prevention of relapse, through learning how to identify early warning signs of depression, how to deal with crises, and how to handle a depressive mood without the guidance of a group or a therapist. Finally, treatment is summarized and concluded.

Mindfulness-based cognitive behavioral therapy

Recently, mindfulness-based cognitive therapy (MBCT; Crane, 2009; Segal, Williams, & Teasdale, 2002; Segal, Williams, Teasdale, & Kabat-Zinn, 2013) was developed as a modification of CBT with the primary intent to prevent the relapse and recurrence of depressive episodes in individuals who had recovered from depression (Lau, 2016). The approach was well-received amongst clients and clinicians, and soon trials and treatments were extended to cover all clients and types of depression, except perhaps the most severe cases. This shift has led to a mixed approach for researchers and clinicians conducting MBCT, where some trials focus on preventing relapse and recurrence, while other trials` main goal is to treat unipolar depression in a comparable manner as the traditional therapies of CBT and GCBT. Based on recent published research and available literature, as well as clinical developments, the general impression is that MCBT is moving towards a more encompassing treatment of depression.

MBCT is a manual-based treatment that combines exercises in mindfulness training with cognitive behavioral techniques. Both therapies are more similar than different, working to help patients better control their thoughts and emotions and thus their responses to them. The mindfulness elements of MBCT, including breathing and meditation exercises, are thought to rebalance networks within the brain to help patients better control the body's automatic responses to the stresses associated with negative thoughts or emotions. The integration of mindfulness practice with cognitive interventions distinguishes MBCT from other mindfulness-based interventions (MBIs), such as mindfulness-based stress reduction (Kabat-Zinn, 1990). The overall goal of MBCT is to increase metacognitive awareness (Lau, Segal, & Williams, 2004) and, thereby, reduce cognitive and emotional reactivity (Gu, Strauss, Bond, & Cavanagh, 2015). Studies have shown that MBCT is effective in reducing the relapse and recurrence of depression (Kuyken et al., 2016; Piet & Hougaard, 2011).

The treatment and clinical trials of current unipolar depression with MBCT usually follows the original manual by Segal et al. (2013). This is in contrast with GCBTs more varied approach to using a treatment manual. MCBT is delivered in a group format with up to 12 participants and one or two instructors. After an individual pretreatment interview in which the participant's history of depression is discussed and information about MBCT is provided, the treatment consists of eight weekly two-hour sessions (Baer & Walsh, 2016).

Several meta-analyses have shown that mindfulness-based interventions (MBIs) in general (e.g., Goldberg et al., 2018; Goyal, Singh, Sibinga, & et al., 2014; Hedman-Lagerlöf, Hedman-Lagerlöf, & Öst, 2018; Wang et al., 2018), and MBCT in particular (Galante, Iribarren, & Pearce, 2013; Lenz, Hall, & Bailey Smith, 2016), are effective in reducing depressive symptoms. For example, a recent meta-analysis of randomized controlled trials (RCTs) observed effect sizes of d = 0.59 for MBIs vs. no treatment and d = .38 for MBIs vs. active control conditions (Goldberg et al., 2018). For MBCT specifically, similar, or higher

ESs for the reduction of depressive symptom severity have been reported. For example, Hofmann et al. (2010) observed an average ES of 0.85 (Hedges' g) in nine pre-post studies. Lenz et al. (2016) reported mean ES of g = 0.76 and 0.54 for MBCT vs. waitlist or no treatment and for MBCT vs. alternative treatments, respectively, in RCTs. Recently. Goldberg et al. (2019) found that MBCT was superior to non-specific control conditions (d = 0.71) at posttest but not more effective than other active treatments (d = 0.00).

Measuring devices for depression

Unipolar depression as a psychological diagnosis has generated a vast amount of research over the last decades (Cuijpers, 2017). This massive focus has not decreased with the shift towards requirements of evidence-based treatment forms. Rather, recent scientific evolution put emphasis on the invention of highly accurate tools for measuring levels of depression and treatment effects. From the beginning of the 1960s many measuring devices for depression have been developed, all with the intention of measuring levels of depression in the most valid and reliable manner possible. While the use of some instruments has faded with time, as is the case for the Zung Self-Rating Depression Scale (Zung, 1965), others, like the Montgomery-Åsberg Depression Rating Scale (Montgomery & Åsberg, 1979) have been utilized in relevant areas. The latter instrument has been a highly popular tool for general practitioners throughout the last decades. Despite efforts in development, few instruments were deemed to encompass both the necessary clinical and statistical qualities to meet the requirements of modern research methods. However, there were two noticeable exceptions: The Beck Depression Inventory and the Hamilton Rating Scale for Depression.

The Beck depression inventory

The Beck Depression Inventory (BDI; Beck et al., 1961), is a self-report rating inventory that measures 21 different attitudes, symptoms, and behaviors that characterize depression. The internal consistency is generally good with high alpha coefficients (e.g., .86 and .81 in psychiatric and non-psychiatric populations; Beck, Steer, & Garbin, 1988).

The Beck Depression Inventory originally took upon itself a huge task, being devised to measure as many as possible of the common symptoms related to major depression disorder. The mission also included a "one-size fits all" objective, where reliable measurements across diagnostic categories and levels of unipolar depression, could be achieved by a single self-evaluation form consisting of 21 items. The items themselves had a broad scope, ranging from cognitive and behavioral symptoms to somatic and emotional ones. Soon after the invention of the BDI, a plethora of researchers and clinicians put the tool to its test, via both clinical trials and statistical and methodological analyses. And, although some minor flaws could be identified from time to time, the conclusion was that this form was the real deal. The BDI proved itself as an easily accessible, low maintenance tool, with enough precision to make statistical comparisons valid, and clinical information highly relevant.

With time, the original BDI has been revised three times, with modest to moderate changes to each version. Compared to its predecessor (the BDI-1a; Beck, Steer, & Garbin, 1988), the latest version (the BDI-II) has incorporated an item measuring hypochondriasis, changed the timeframe of symptoms from one week to two weeks, and put even more emphasis on measuring all diagnostic criteria related to depression (Beck et al, 1996). This latest revision was largely inspired by changes in diagnostic procedures for depression, as outlined by the Diagnostic and Statistical Manual of Mental Disorders IV (APA: DSM IV, 1994). One of the stated aims from the inventors of the BDI, was to connect the inventory

closely to the diagnostic manual, with previous revisions being made shortly after a new edition of the manual has been released.

The updated version of the inventory has been shown to have solid psychometric qualities. For example, Beck et al (1996) found high alpha coefficients (.91), and high test-retest reliability (r = .93) across several samples. The BDI has moderate to high correlations with other inventories measuring depression. For example, the Hamilton depression rating scale (HDRS; Hamilton, 1960) has been found to correlate between .5 to .8 with the BDI (Beck et al., 1988; Beck et al., 1996). The criterion validity of the BDI has also been examined. Ambrosini et al (1991) investigated the concurrent validity on initial BDI-scores, and found high sensitivity, specificity, and predictive powers of 86%, 82%, and 83%, respectively. The predictive validity of the BDI was extensively investigated in a study by Burkhart et al (1984), where 600 students filled out the BDI with varying intervals. The results showed that scores on the BDI declined in a faster rate with repeated administration, leaving the authors to suggest that the BDI does not discriminate between stable and unstable forms of depression.

The BDI shares theoretical principles and foundations with cognitive behavioral therapy. As such, since the 1970`s almost all studies of CBT for depression have utilized the inventory as their primary outcome measure. This high degree of standardization offers a major advantage when it comes to examining time-trends and is thus the main reason why the BDI was the primary outcome measure for the papers and analyses included in the present thesis.

The Hamilton depression rating scale

While the BDI is generally held in high regard for its psychometric qualities, there are also challenges and potential problems linked with self-report inventories. Patients could have

difficulties understanding the general instructions or wording of the items correctly, or they could distort their answers (even lie). One way to eliminate the potential confounding effects connected with subjective self-reports, is to have an objective rating of the patients' symptoms. This is what the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960) offers. The HDRS is a clinician administered rating scale, measuring similar characteristics of depression as the BDI. The device has good psychometric qualities, with a generally high inter-rater reliability, where coefficients typically exceed .84 (Hedlund & Vieweg, 1979). The correlation between the BDI and the HDRS is in the moderate to high range, r = .5 to .8 (Beck et al, 1988; Beck, Steer, & Brown, 1996).

While the use of other depression measures in treatment trials have been varied, the HDRS has been utilized more frequently, in conjunction with the BDI or by itself. Thus, the HDRS was an obvious and natural choice as the secondary outcome measure in the relevant meta-analyses and papers building up this thesis.

Meta-analysis and meta-regression analysis

As the field of psychotherapy gradually moved towards a more prudent and rigorous scientific approach during the 1950`s and 60`s, researchers typically used traditional and trusted statistical methods based on the principles of correlation when investigating treatment effects. Research trials consisted of small samples with 4-10 participants being enrolled in a treatment program, where the intervention was the treatment method to be investigated. Outcome measures varied to a large extent, but the common idea was to gather clinical information at the start and finish of treatment, and then compare the registrations. This was typically done through a t-test, and the results indicated to which degree an intervention was effective in easing the participants suffering. The approach quickly gained approval, especially from the adherers of the modern psychological fields of behavior therapy and

cognitive therapy. The rising popularity contributed to an increase in treatment trials, which increasingly were conducted with similar methods, and were investigating similar interventions (for example BT). However, the outcomes of such trials often revealed differences, despite very similar approaches and research designs. The need of a method to combine and summarize the results become apparent. Thus, in the late 1970`s, the research field of psychology experienced a breakthrough, via the work of Gene V. Glass and others. Glass and his associates' efforts culminated with the first recognized study where results from different treatment trials were combined, and the outcome was a summarized result derived from the included trials (Smith & Glass, 1977). The term meta-analysis was selected to describe the method, and the procedure soon became the gold standard for summarizing research findings. This is a standing the method still largely enjoys in contemporary health-sciences.

In the beginning, meta-analytical models and programs were designed and utilized to summarize effect sizes, which allowed comparisons between treatments to be made on a higher level, with more included patients, and vastly enhanced validity and reliability. However, the earlier models were quite crude. The ability to process and analyze several variables simultaneously and check their potential moderating influence on effect sizes (treatment outcome) was not possible or did not yield accurate results. This changed with the invention of meta-regression analysis, a statistical procedure largely inspired by the works of Cathrine S. Berkey (Berkey et al, 1995). The method allows researchers to investigate the degree of influence different variables exert on the outcome measure of choice and has gradually become implemented in the field of psychological research from the early 2000`s and onwards. One of the computerized programs based on the method of meta-regression analysis, is the CMA (Comprehensive Meta-Analysis), which is the statistical program utilized for the meta-analyses comprising the present thesis.

Advantages and statistical considerations

The benefits of using meta-analytic methods to summarize clinical results are well-known (Borenstein et al., 2009). By accessing a large pool of studies and assigning the individual studies different weights according to their sample size, the potentially troublesome role of individual studies indicating weak or even contradictory results is minimized. A meta-analysis is also preferable in situations where studies are well-defined or similar in terms of patients, diagnoses, intervention procedures, and the measurement instrument used (e.g., the Beck Depression Inventory), thus simplifying the quantification of the effect size considerably. Moreover, meta-regression approaches may be used to identify potential sources of co-variation between study-related factors and treatment effects.

There are several available methodological and statistical options when performing meta-analyses, all of which could influence effect sizes (ES). After determining which variables and phenomenon to be investigated, and which outcome measure to be used, the procedure of calculating ESs typically utilizes either the pre-post within-group, or the between group (controlled trial). For studies that do not include a no-intervention control group, a standardized mean difference (SMD, also denoted Cohen's *d*) is calculated for the intervention group (*M*pre – *M*post, divided by the standard deviation -SD- of the change score). Often, the Hedges *g* correction is made to the SMD, which reduced the SMD for studies having small sample sizes (Hedges & Olkin, 1985).

For the controlled trials (CT), or the between group condition, effect sizes are calculated from the difference between the pre- and post-test scores on the outcome measure (for example, the BDI and the HDRS) for the intervention group and the control group, respectively, and then standardized using either the change scores, or the post scores. When examining trials consisting of few participants, standardization using change scores is usually preferred, because studies including a smaller number of participants might contain pre-

intervention differences despite randomization. The change score variant is less sensitive to such differences compared to standardization using post scores. Another potential advantage of using the SD for change scores is that the effect sizes for CT studies are estimated similarly as studies without a control group (within-study designs). Standardization by change scores also is recommended when the objective is to assess change relative to pre-intervention scores (Kulinskaya et al., 2002), and has frequently been the preferred method to quantify treatment effects in meta-analytic reviews of psychotherapy (e.g., Abbas et al., 2013; McGuire et al., 2014; Zoogman et al., 2014). However, one limitation is that change scores require knowledge of the pre-post correlation, which typically is not reported in the individual trials. As such, this index often needs to be imputed. The utilization of a controlled trials condition is usually the preferred method when performing a meta-analysis on treatment effects. This method yields significantly lower scores of heterogeneity compared to within group designs, and typically shows lower (and less inflated) treatment effects (Pallesen et al., 2005).

Another important consideration is whether to apply either a random effects model, or a fixed effects model. As the random effects model has the assumption that the true effect sizes would vary among studies due to the study-related factors, this is most often the choice when comparing and calculating treatment effects. Employing a random-effects model increases the generalizability of the results (Field, 2003).

Disadvantages and alternative models

Although the standard meta-analysis and meta-regression analysis offer precise methods for summarizing data, the designs also have inherent disadvantages, as compared to other modern methods such as network meta-analysis and individual participant data meta-analysis (IPD). The perhaps biggest issue relates to the level of data processing. The raw data for meta-regression analyses consists of the aggregated data for treatment trials. This means

that data from the individual participants are not considered. However, aggregate data are sometimes not available or poorly reported in articles and are more likely to be reported when statistically or clinically significant, amplifying the threat of publication bias and within study selective reporting (Riley et al, 2010). On the other hand, individual participant data facilitates standardization of analyses across studies and direct derivation of the information desired, independent of significance or how it was reported. Individual participant data may also include a longer follow-up time, more participants, and more outcomes than were considered in the original study publication.

Another common issue with aggregate meta-analyses, is the vulnerability for small samples to achieve effect sizes that may not be representative. This issue is not relevant for the IPD-approach. The method may thus offer a more detailed, precise, and robust approach for summarizing data, which is why contemporary researchers are increasingly moving in that direction. However, it should be noted that for many study-designs the IDP meta-analysis is not a realistic option. One case in example, are the analyses building up the current thesis. The temporal focus means that available data must be collected from studies across a large timeframe, with the oldest articles exceeding 40 years. It is thus not conceivable that all the relevant data is still intact or available, and there would most likely be difficult to get the necessary approvals for gaining access to the individual participants' data.

When it comes to comparing different treatments simultaneously, the conventional meta-analysis is limited. It can only compare two interventions at a time, and mainly those evaluated directly in head-to-head trials. Thus, if the objective is to analyze the effectiveness of several treatments in one go, the network meta-analysis is the preferred option. This method offers a potential more efficient and comprehensive way of comparing treatments, and its appliance is therefore rising within the field of medicine and psychology. However, this method is not appropriate for some designs or research questions. If the main objective is to

evaluate potential moderators influence on any measured phenomenon (for example an effect size), the meta-regression analysis and the IPD meta-analysis are considered better options.

Researcher allegiance is a variable that may influence the design, procedure and outcome of any research trial or study. Meta-analyses should not be considered an exception, even though they, historically speaking, may seem to have been regarded as exempt from this phenomenon. As such, the amount of literature and studies on the topic is very limited, bordering non-existent. The lack of interest in such a phenomenon seems strange, as there are many obvious manners in which allegiance could influence a meta-analysis, ranging from the selection procedure itself, to methodological and statistical choices, culminating in if/how to communicate any findings. This is a topic where future attention is welcomed.

Treatment effects and effect sizes

In the context of data analysis, relationship typically refers to the correlation between two characteristics or attributes for a set of analysis units. This also apply for trials in psychotherapy, where the correlation between the variables "treatment" and "outcome" is the relationship of interest. Outcome may be defined in many ways, including self-report inventories, and/or ratings made by a clinician. The invention of the term correlation is ascribed to Sir Francis Galton, and his work in the late 1880`s (Huberty, 2002). Pearson (1905) developed the concept further and labelled the correlation ratio as "Pearson`s "*r*, or simply "*r"*, as is the most usual prefix today. The correlation ratio was the underpinning of further and more complex statistical developments, such as *t* and *f* tests of significance.

In the two-group mean-comparison situation, the typical effect size index considered is a standardized mean difference (SMD). Such an index was first proposed by Cohen (1962), with the term "*d*" representing effect size as the difference between two groups, as a consequence of any moderating variable (for example, implementation of cognitive

behavioral therapy). Effect sizes below 0.2 are considered low, moderate effects are observed up to 0.5, large effects up to 0.8, and very large effects from 0.8 and upwards. Further refining of Cohen`s *d* was done by Hedge (1981), who corrected the *d*`s inhered bias towards small samples, with his index Hedge`s *g*. For trials and meta-analyses measuring the efficacy of CBT, the expression of the SMD as *d* or *g* has been common practice.

Although being an easily understandable and accessible expression of treatment effects, the measures of *d* or *g* are not perfect, and are specifically vulnerable for small sample sizes, where effect sizes are at risk for being inflated. This typically happens when all, or the vast majority of, members of a small sample (2-20 participants) report good effects from a treatment trial, without any significant variability within the group. For example, if all 10 clients included in any depression trial had a reduction pre-post treatment from 30 to 3 points on the BDI, ESs would be in the range of 7-8 *g,* which is a value that could be hard to make any statistical or clinical meaning from. This phenomenon is unlikely to happen with large samples, in which the principle of random distribution usually means there will be a larger spread in scores and outcome among participants within a treatment trial. Standard deviations will increase, and the observed ESs decrease. When performing a meta-analysis based on trials which include small samples, this is important to be aware of. Modern computerized programs have got sophisticated statistical procedures that are designed for adjusting the possible bias associated with smaller sample sizes, such as the Hedges` g calculation. However, in common practice researchers still come across ESs that remain very high, and thus possibly are inflated. Considerations should in such instances be made as whether the study should be excluded or not. This is especially important if the resulting analysis leads to any significant findings. In such circumstances, a control-mechanism should always be implemented, usually in the form of a re-analysis on the same dataset without the outlying study included.

Another factor that may exert influence on ESs, is the stringency in which the conditions of the treatment trials are conducted. It is wise to be aware of a potential inverted relationship; when stringency increases, ESs are vulnerable for a decrease. Higher demands can, for example, occur as requirements of a control group, implementing the same exact circumstances for each participant/therapist/rater in a trial, and incorporating a strict selection procedure. While traditionally difficult to measure precisely, a rating of study quality is thought to provide an index which is relevant for pinpointing the level of stringency. When calculating aggregated effect sizes, it is thus recommended to utilize a reliable measure of study quality. In this way, statistical checks can be performed to determine if study quality (and stringency) in any way covariate with ESs. If such a relationship does indeed exist, caution must be implemented when interpreting the results.

Clinically meaningful change

The term effect size is a statistical expression of treatment effect. As such, it is not possible to translate ESs directly into an accurate number of patients who feels subjectively better or have recovered from a diagnosis. Further, to describe the robustness of any findings in a research trial, the preferred term is "statistically significant". Unfortunately, the term statistical significance does not automatically equate to a *meaningful* or *practical effect*. Some statistically significant effects are meaningful, yet others are not. Because statistical significance and practical significance are often conflated when one interprets research findings (i.e., statistical significance is assumed to establish practical significance), researchers are now encouraged to explicitly interpret the practical import of statistical results by providing estimates of effect sizes (Schuele & Justice, 2006). Effect-size estimates are values that characterize the magnitude of an effect or the strength of a relationship. They do not necessarily give information about meaningfulness. Thus, readers of research articles must

consider two issues to decide whether research results are sufficient to have an impact on clinical practice: First, are the results statistically significant? Second, are the results clinically meaningful or relevant?

To make informed recommendations about to which degree a treatment is of benefit to patients, clinicians and researchers should thus figure out what constitutes a clinically significant treatment effect. This can be done by establishing the minimal clinically important difference (MCID) on the outcome measure, i.e., the smallest difference in score considered clinically worthwhile by the patient. MCID is a patient-centered metric that captures both the magnitude of improvement and the value the patient places on that improvement (McGlothlin & Lewis, 2014). For the BDI, the main outcome measure for depression, MCID has traditionally been established by clinical consensus or recommendations. A change of three points on the inventory is typically considered to indicate that MCID has occurred (NCCMH, 2004). Although frequently utilized, this way of determining clinical meaningful change has not been empirically tested, nor validated. Furthermore, utilizing recommended thresholds have the disadvantage of ignoring the subjective point of view of the patient. This matter has been addressed in recent research, where the patient's subjective feelings of global improvement has been measured against improvements on the BDI-II (Button et al, 2015). This comprehensive study indicated that MCID is best measured on a ratio scale as a percentage reduction in symptom score. The researchers estimated that the MCID on the BDI-II corresponds to a 17.5% reduction in scores from baseline. For individuals with longer duration depression who had not responded to antidepressants, the corresponding estimate was 32%. As such, baseline severity is a major factor when determining MCID. For patients with a baseline score of 20 on the BDI, the recommended 3-point reduction does indeed represent clinical meaningful change. However, if baseline is a score of 40 points, the clinical recommendations are not entirely accurate, nor applicable.

The Hamilton Rating Scale for Depression has also recently been evaluated for clinically meaningful change (Rush et al, 2021). Here, the researchers found that a reduction in symptom score > 4 constitutes a clinical meaningful change for patients with non-psychotic major depressive disorder. A change in symptom scores of 7 or more was determined to be clinically substantial.

What impact on effect sizes will a change in three points on the BDI, or 4 points on the HRSD, exert? The answer is "it depends". This is due to differences in the factors determining ESs, which will vary from study to study according to the individual characteristics of a trial. However, estimates based on the datasets for the meta-analyses performed in the current thesis, indicate that a change in 3-4 points for the BDI and HRSD roughly translates to a difference of 0.2-0.4 *g* in ESs. These estimates show that a clinically meaningful change only qualify as a rather weak effect size. Thus, if patients in a research trial experience reductions from pre-treatment scores of 23, to post-treatment scores of 19, this is considered clinically meaningful, but will most likely equate to statistically low effect sizes. It should also be observed that such a reduction does not imply any diagnostical change; the patient still (most likely) meets the diagnostic criteria for major depression, at the same level (moderate) as before treatment started. These considerations indicate that for CBT anti-depressive treatment trials, even minor differences in statistical effect sizes equates to a clinical meaningful change.

Specific moderators for treatment outcome (effect size)

Effect sizes may vary according to the characteristics of the included studies, their samples (clients) or the setting of the studies. As such differences may influence treatment outcome (quantified as effect sizes) it is considered highly informative to investigate if any specific variables have an undue, or confounding, effect on the outcome of a psychotherapy

trial. These variables, often denoted as moderators, typically consists of easily identified characteristics that are reasonably unproblematic to categorize. For clients entering a trial investigating the effect of psychotherapy for depression, the following variables are typically registered and investigated; gender, age, degree of psychiatric comorbidity, use of psychotropic medication and severity of depression. Previous studies have typically not revealed any significant differences in treatment effects related to gender and age (Joutsenniemi et al., 2012; Wierzbicki & Pekarik, 1993), while a higher degree of psychiatric comorbidity often implies a worse course of illness or treatment prognosis. The most common Axis I comorbidity is anxiety disorders (Kessler, et al., 2003), which usually imply a higher degree of severity at intake (Kohler et al, 2013), as well as a poorer natural course (Penninx, et al., 2011). The presence of comorbid Axis-II disorders, of which the Cluster C diagnoses, particularly, avoidant personality disorder, are the most prevalent (Friborg et al, 2014), heightens the risk of a worse outcome following treatment (Newton-Howes, Tyrer, & Johnson, 2006). The relative efficacy of psychotropic medication versus CBT has been subjected to many clinical trials; however, a meta-analysis of 21 studies found no differences between the two treatment modalities in alleviating depression (Roshanaei-Moghaddam et al., 2011). The addition of medication to CBT has been studied to a lesser degree; however, a meta-analysis consisting of seven studies found that CBT plus medication was slightly better ($d = 0.32$) than CBT alone (Cuijpers et al, 2009). With regard to the severity of depression, previous research has found that patients who were more severely depressed, reported larger treatment effects than less severely depressed patients. This phenomenon is also described as regression to the mean (Garfield, 1986; Lambert, 2001).

For characteristics related to the therapist, the variables most typically investigated are type of therapist (psychologist, psychiatrist, general practitioner, student, etc.), and ratings of the competence of the therapist. Previous research has not indicated any significant difference

related to type of therapist, although more therapeutic experience has been found to relate to a shorter time to remission (Okiishi et al., 2006).

When it comes to trials investigating CBT as an anti-depressive treatment, the technique-specific and methodological factors investigated typically consists of number of therapy sessions, application of a CBT manual (typically Beck`s original manual from 1979), and checks of adherence to the treatment protocol (including subsequent feedback to the therapists). A dose-response relationship has been documented, in that additional sessions of therapy usually lead to a higher treatment efficacy (e.g., Howard et al., 1986). Adherence to a treatment manual, which ensures correct implementation of CBT, has been found to improve the outcome (Shafran et al, 2009).

Common methodological moderators for psychotherapy trials in general, are type of statistical analyses (intention to treat or completers), and ratings of the methodological quality of the study. Studies using stricter criteria for methodological quality generally yield lower treatment effects (Gould, Coulson, & Howard, 2012). A quite recent meta-analysis (Hans & Hiller, 2013) showed slightly larger effect sizes in depression treatment trials using a statistical design requiring treatment completion (d = 1.13), as compared to an intention to treat (ITT) design (d = 1.06).

When it comes to the general surroundings and environment of any trial (the common factors), there has not been a great deal of systematic investigation, except for perhaps ratings of therapeutic alliance. Patients experiencing a stronger alliance with their therapist have been reported to show better effects of their treatment (Rector, Zuroff, David, & Zindel, 1999).

Researcher allegiance has increasingly received attention as a potential factor that could significantly moderate the reported effect sizes in randomized controlled trials, and, as such, also in meta-analyses. However, historically there are very few RCTs in which the researchers disclose their allegiance. A recent and comprehensive study investigated how

many authors of RCTs in the field of psychotherapy reveal their allegiance (Dragioti et al, 2015). The results showed that from 793 trials where obvious researcher allegiance was found, only three percent of authors stated their allegiance in the article text itself. This finding does have a large bearing on meta-analyses in the next instance, where only 17 percent of 146 papers included information about allegiance. Further, only four percent of meta-analyses employed a proper method to control the effects of allegiance in the individual trials. The meta-analyses comprising the current thesis do not investigate or report researcher allegiance, and the lack of relevant information in the individual trials are the main explanation behind this choice.

Another common factor for trials may be country, or region, for where the trial is investigated. A recent meta-analysis found no differences in ES for studies performed in the US vs the rest of the world (Christea et al, 2017). From a wider perspective, common factors could also consist of type of political system, religious views, and other cultural and societal characteristics. The influence these variables may have on treatment effects have not, to the best of the doctoral candidate`s knowledge, been previously systematically investigated.

The underpinning of effect sizes

When a treatment is efficacious, psychotherapy research trials point to four sources to explain the observed improvements: (1) client factors, (2) therapist factors, (3) the so-called common factors, and finally, (4) technique-specific factors. Client factors represent the characteristics of the patient, such as personality traits, temperament, motivation for treatment, or important life events experienced by the patient during the course of therapy. Therapist factors are the characteristics of the therapist, which can include anything from gender, age, and education to personal style and appearance. Clinical training, competency, and skills in establishing a therapeutic alliance and using therapeutic techniques are of

particular importance (Crits-Cristoph, 1991). The two latter components may also be denoted as common and technique-specific therapy factors, which influence the outcome of CBT.

The common factors represent characteristics of the treatment setting that are important and common to all therapy models. These characteristics may include the context of therapy; the client, the therapist, and their relationship (usually coined as the therapeutic alliance); how expectancies for improvement develop; a plausible rationale explaining the patient's illness; or even therapeutic techniques that are not specific to a therapy model.

The technique-specific therapy factors represent those elements that are specific to a particular therapy model, and typically are described thoroughly in therapy manuals, indicating specific topics to be addressed during therapy, how they should be conveyed, the implementation of structure, the number of therapy sessions, the degree of exposure, and/or the schedule of homework tasks.

The use of experimental designs has been the choice of researchers when it comes to providing insight regarding which of the four variance components contribute most to the treatment effect. Such procedures are an attempt to isolate and define the influence of the different factors, to identify which ones are the most important to improve. Previous studies have shown that a major part of the treatment effect seems to be caused by the client-related and common factors, which explain between 30–40% and 30–50% of the total treatment effect, respectively (e.g., Horvath & Greenberg, 1986; Luborsky et al., 1988). The therapist-related factors have been found to explain 5–15% of the treatment outcomes (Huppert et al., 2001; Wampold & Brown, 2005). Such calculations leave approximately 10–20% of the effect attributable to the specific therapy (Lambert, 1992; Duncan, Miller, & Sparks, 2004).

Since the common factors have been shown to be so important for attaining improvement following therapy, psychotherapy researchers have increasingly become concerned how to integrate them into the therapy (Imel & Wampold, 2008). An important line

of support for the common factors model comes from meta-analyses showing that different treatment modalities produce relatively comparable treatment effects (e.g., Smith & Glass, 1977; Wampold et al. 1997); hence, the assumption that elements common to all therapies underlie the lack of marked differences among them (Lambert & Bergin, 1994; Seligman, 1995). The most important ingredient in the common factors have been the working alliance between the therapist and the patient. A more solid alliance is associated with quicker and larger treatment effects (Rector, Zuroff, David, & Zindel, 1999), and a reduction in the number of early dropouts (Kegel & Fluckiger, 2014).

Therapists who use CBT are trained to establish rapport by, for example, socializing the patient to the cognitive therapy process (thus, being explicit about how the therapy will progress, which may reduce uncertainty), communicating to the patient how CBT might be helpful (instilling hope and positive expectations), and educating the patient about the disorder per se (helping patients to understand their problems). Moreover, CBT therapists set an agenda in collaboration with the patient in order to avoid spending the limited amount of time they have on irrelevant topics. They actively invite the patient to provide feedback, to ensure a mutual understanding and provide opportunities for quick adjustments. They construct and continuously refine their conceptualization of the case, further facilitating and deepening the understanding of the patient's problems. They collaborate actively with the patient in making plans for between-session tasks that may help eliminate negative personal beliefs and behaviors. The latter may help the patient to attribute positive changes to their own efforts, thereby increasing self-efficacy. For this reason, improvements in self-efficacy may be mediated by using specific techniques aimed at improving self-efficacy, in addition to an effective integration of the common factors.

A modern holistic approach

As the common factors are universally recognized as having the largest influence on treatment effects, a further investigation of their true constituents is warranted. The framework of the common factors is our immediate world; the "here and now" for every citizen on planet earth. The main elements in this frame are society, culture, geographical position, political state, technological advances, school systems and environmental developments. Attitudes, and ethical, ideological, philosophical, and religious views are also higher-order common factors. Moving down the hierarchy, we observe the background, upbringing, and immediate environment of each client (and therapist), as well as their neighborhood, class of school and workplace. Even further down the line, we find the common factors that previous research has identified as contributing to treatment effects, such as the context of therapy, expectancies, the placebo-effect, and the working alliance.

Every specific treatment, like CBT, has come to life in a specific set of common factors, represented by the leading ideas and views present in the society of the therapy`s conception. The influences of the common-factors also have their bearing on client and therapist-factors. Both parties are embedded in, and influenced by, the context of their own, specific environment. It is thus essential to keep in mind that also every specific factor (technique or treatment form) is embedded within the larger framework of the common factors. This notion suggests that there is a hugely complex, advanced, dynamic and continuously developing interaction between what has been previously coined as the common factors, and the other factors (client, therapist and technique). They are inextricably and inherently connected to each other. If there is development on one part, there will invariably be effects for one (or more) of the other.

Previous research has provided examples of the interaction between factors related to treatment effects. The role of specific versus non-specific factors in CBT seem to shift with

the provision of an increasing number of therapy sessions (Honyashiki et al, 2014). Intuitively, this makes sense, as (the lower order) common factors (e.g., alliance) should be more important in the beginning of therapy, while efficient implementation of treatment-specific factors is increasingly important as therapy progresses – and gains more effectivity by having a solid foundation from the start. In addition, the role of common factors has also been shown to depend on the mental disorder of the patient. For example, patients with borderline personality disorder may respond much more favorably to the relationship and alliance-building skills of a therapist (Bienenfeld, 2007) compared with patients with bipolar disorders. These are but two minor examples of the interaction between common factors and specific factors. However, in the treatment room the factors are always and inevitably linked in various and complex manners. This also holds true for the valued and important working-alliance between patient and therapist. The alliance is also highly influenced and determined by the overriding factors, consisting of culture, attitudes, ideas, general knowledge, and background on each parts behalf.

When it comes specifically to CBT as a depression treatment, we have an intervention (technique, specific factor) developed in the United States (US) of the 1970`s (common factor; society, geography, knowledge etc.), delivered by therapists in all shapes and forms, to an even larger variation of patients (therapist, client-factors), throughout the passing of more than 40 years of time. As specific techniques dictated by a therapy model, according to the current understanding, represent a small part of the overall treatment effect, one would theoretically expect that refinements or improvements of CBT approaches over the past 40 years would have little impact on treatment efficacy, or reported effect sizes. This lack of impact would be even more pronounced for therapy that adheres to the original CBT-manual. However, it is at this junction it becomes essential to keep in mind that the implementation of any specific treatment component is always embedded within a common factors model. This

interaction strongly implies that it is entirely plausible, and expected, to observe significant changes in treatment effects with the passing of time – even (and perhaps more so) if the technique remains unchanged. In fact, the opposite view implies a bizarre scenario: For no development in treatment effects to occur, the world would essentially have to stand still, with every aspect of life being the same – as the years pass on. This especially holds true the longer a treatment paradigm, or technique, is in existence.

Temporal developments and time-trends

A hallmark of our modern society has been the rapid development in many domains, particularly in science, technology, and health. Old procedures and methods have been replaced with safer and more effective solutions. For example, in somatic health care, cruciate ligament surgery currently takes considerably less time, requires fewer resources, and has a better long-term prognosis than it did 30 years ago (Cirstoiu et al., 2011). Another example is a percutaneous coronary intervention (formerly known as coronary angioplasty), which uses a catheterization technique to insert a stent in the groin or arm to improve blood flow in the heart's arteries. The technique is quick and presently requires minimal rehabilitation (an overnight hospital stay); hence, it represents a huge improvement compared to older techniques (Knapik, 2012).

Despite the large number of clinical trials and reviews of both CBT and other forms of therapy, the author of this thesis did not find any previous publications or attempts made to systematically evaluate how the efficacy or effectiveness of CBT has evolved over time. The publication of paper I in this thesis did, however, ignite activity in this area of research. The subsequently published, relevant studies will be evaluated and critically reviewed in the discussion part of the present thesis.

The reasons why researchers have not previously focused on temporal developments of treatment effects are largely unknown. One factor probably relates to methodological challenges. Comparable improvements/developments in somatic healthcare are clearly defined and have standardized procedures, with the outcomes being more readily observable. Another likely issue relates to statistics. Having knowledge of, and access to, the right analytical tools is imperious for any complex calculation of effect sizes and its potential moderators. A third factor probably lies within the attitude of clinicians and researchers; "if it isn't broke, why fix it", the notion being that it is not worth investing time and effort into scrutinizing a seemingly highly functioning and effective treatment. After all, any therapy that still exists after 5 decades, is obviously not ineffective. And the clinical form of "natural selection" (RCTs) would probably sooner or later get rid of the ineffective approaches anyway. However, when considering the vast number of patients worldwide afflicted by depression, even the smallest loss or gain in treatment effects would translate into huge numbers when it comes to overall monetary and health wise considerations.

Aims and hypotheses

Paper 1

The first article, "The effects of Cognitive Behavioral Therapy (CBT) as an anti-depressive treatment is falling: A meta-analysis" offers a meta-analytical examination of treatment effects and effect sizes connected to individual CBT. The primary *objective and aim* was to examine whether published clinical CBT trials aimed at treating unipolar DDs demonstrate a historical change in treatment effects. The secondary purpose was to examine the role of various moderators of the reported effect sizes. Further, in order to examine if the client, therapist, technique, and common factors were related to treatment effects differently across time, all of the available data related to these components in the CBT studies were

included in the meta-analysis. Finally, an examination as to whether these moderators influenced the regression slopes describing the time trends in the treatment effects was performed.

The hypotheses were set a priori and based on knowledge available at that time (2014). The assumption was that although CBT treatments have focused less on the common factors, it was still deemed plausible that CBT therapists had become increasingly aware of the importance of integrating common and specific techniques to take full advantage of the therapy. Therefore, the expectancy was for contemporary CBT treatments to show better treatment outcomes when compared to older clinical trials. It was also predicted that diagnostic severity and type of therapist (psychologist or student therapist), and therapist competency would be associated with better treatment effects, while the variables age and gender were not expected to co-vary with therapeutic outcome.

Paper II

The second article in this thesis, "The effect of cognitive behavioral therapy as an anti-depressive treatment is falling: Reply to Ljottson et al (2017) and Cristea et al (2017)", was both an extension of the analyses made for paper I, as well as a critical reassessment of two subsequent studies challenging its original results and conclusions. This paper`s primary *objective and aim* was to reanalyze the data set for nonlinear time trends, examining the conclusion from Ljottson et al, where they stated that the drop in effect sizes for CBT had been halted from 2001 and onwards. The secondary purpose of this article was to examine and critically assess and evaluate the findings and conclusion from the meta-analysis performed by Cristea et al (2017), where the authors essentially denied the existence of a true drop in treatment effects with time for CBT. The expectancy was to achieve a more nuanced

explanation of their findings. *The hypothesis* was that the reanalysis would affirm the original findings from paper I.

Paper III

The third article, "A meta-analysis of group cognitive behavioral therapy as an anti-depressive treatment: Are we getting better?" followed in the footsteps of paper I and used a similar approach to investigate the temporal developments of treatment effects for GCBT. The study had an *exploratory design*, and no hypotheses were thus set in advance. *The aim* was to examine whether there was a similar decline in effect sizes of GCBT for depression with advancing publication year, as was observed for individual CBT. A second purpose was to investigate moderators of the reported effect sizes of GCBT. As meta-analyses are considered the best available method for such evaluations, as well as for examining time-trends, this was again the preferred statistical choice. The rationale for paper III was thus to provide information and updates with regards to the development of effect sizes for group cognitive behavioral therapy over time.

Paper IV

The fourth article, "Time trends in the effects of mindfulness-based cognitive therapy for depression: A meta-analysis", followed in the footsteps of the previous articles, with its focus on CBT-related treatment and time-trends. This time *the aim* was to investigate ESs and explore if there were any significant time-trends related to a modern development of CBT, the mindfulness-based cognitive behavioral therapy. The main interest concerned if a form of therapy with a much shorter lifespan, 10 years for the included papers in the meta-analysis, also would show a decrease or increase in treatment effects with time. Potential moderators`

influence on effect sizes were also investigated. As the article covered previously unchartered territory, *the design was exploratory*, with no specific hypothesis set in advance.

General aims and hypotheses for the thesis

The main *aim* of this thesis was to examine whether any changes in CBT`s effects as an anti-depressive treatment have taken place since the introduction of this form of therapy five decades ago. Contemporary cognitive behavioral therapy is found in varying designs and formats, all of which share the basic cognitive assumptions and theoretical framework. As such, the time-trends for individual CBT, group CBT, and mindfulness-based CBT were all investigated in separate analyses and papers, with a common meta-analytical design. The secondary objective was to investigate potential moderators of effect sizes and time-trends. Subsequently, data for selected variables were registered and analyzed in subgroup conditions for all meta-analyses.

Broadly speaking, the designs of the meta-analyses were *exploratory*. This is because a limited amount of research and knowledge regarding time-trends existed prior to the publication of the articles comprising this thesis. However, the initial *hypothesis* was that CBT as a treatment for depression would show an improvement with time.

Methods

The present thesis comprises work from three different, but thematically linked projects. The first two papers utilized identical sample and data set, while papers III and IV have separate datasets. All papers incorporated a similar meta-analytical approach and design, in which individual trials were systematically searched for, and their results collapsed and summarized through meta-regression analyses.

Although the current section provides a description of methodical aspects, information is kept at a level deemed appropriate for summarizing purposes. For a complete and detailed description and overview of all methodical aspects, flow-charts, figures, and tables, please observe the relevant sections in the original and published manuscripts for the individual studies. As the proposition of a thesis provides an arena to expand the rationale behind selected methodical choices, relevant parts of the current section provide a more detailed insight in the decision process as compared to the original articles.

Individual CBT (Paper I)

Data collection, studies & selection criteria

A thorough search procedure was conducted in January 2015. Following a review of relevant abstracts, 489 articles were obtained via the University library. The following exclusion criteria were then applied: (1) the implemented therapy was not pure cognitive behavioral therapy. Among the studies comparing CBT with other treatment forms (interpersonal therapy, for instance), we included only the CBT treatment arm; (2) a unipolar DD (either mild, moderate, severe, or recurrent) was not the primary psychiatric diagnosis; (3) participants were not adults (mean age < 18); (4) therapy was not implemented by a therapist trained in CBT; (5) the psychotherapeutic intervention was not intended to treat depression; (6) the outcome was not measured with the BDI or the HRSD; (7) patients had acute physical illnesses or suffered from bipolar or psychotic disorders; (8) treatment was not implemented as individual face-to face therapy; and (9) the patients had a BDI score lower than 13.5. If a study assigned patients to different sub-groups based on diagnostic severity (usually based on the pre-test BDI scores), only the most severe sub-group was included to avoid inflating the number of independent studies. The selection procedure yielded a final study pool of 70 studies.

Coding of study information and moderator variables

The following data from the studies were coded: demographic information (gender and age), year of implementation of the intervention, duration (number of sessions), type of therapist (psychologist, trained psychology-student, or other/unknown), therapist competence (as measured by the Cognitive Therapy Scale), information about the severity of the diagnosis (mild, moderate, severe, or recurrent depression) along with the proportion (%) of the sample having comorbid psychiatric diagnoses, whether the patient population had any special characteristics, and the proportion (%) of patients using psychotropic medication.

We investigated whether the effect sizes co-varied with any of the following moderator variables: type of statistical analysis (intention-to-treat versus completers analysis), gender (as % men), age, proportion of patients using medication, proportion of comorbidity, use of the Beck CBT treatment manual versus no manual, checks (and subsequent feedback) of therapist adherence to the treatment manual, version of BDI (I or II), severity of the depressive disorder, diversity of the study populations, number of therapy sessions, type of therapist, therapist competency, and the publication year of the CBT intervention (the moderator of most interest). We also examined whether the latter variable co-varied with the effect sizes in the waiting list control groups. The competence of the therapist was, in a few studies, rated using the Cognitive Therapy Scale (CTS; Dobson et al, 1985), and it was included as a moderator. The Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS) was used to rate the methodological quality of the published studies (Kocsis et al., 2010).

Effect sizes & statistical calculations

We used two procedures when calculating the effect sizes based on the BDI and the HRSD pre-/post-intervention scores: a pre-post within-study design, and a controlled trial

(CT) design. For studies that did not include a no-intervention control group, a standardized mean difference (SMD, or Cohen`s *d*) was calculated for the intervention group. A Hedges *g* correction was then applied to the SMD.

The effect sizes for the treatment remission rates were coded as an event rate (rate = number of patients achieving remission/sample size). Remissions were, by the authors of the individual studies, deemed to have occurred for patients who ended treatment with a BDI score below a pre-defined clinical cut-off score, < 10. This is in accordance with common interpretations of the BDI, where scores from 9 and below are considered to be under the clinical threshold for a depression diagnosis.

The Comprehensive Meta-Analysis software, version 2 (Borenstein et al., 2005) was used for all statistical analyses, except for the two-way interaction analyses between the moderator variables, which had to be analyzed in SPSS 21. Meta-regression analyses were used to analyze the role of the continuous moderator variables (e.g., publication year), and were based on the unrestricted maximum-likelihood method, as it assumes an underlying random distribution of effect sizes. The moderator analyses for the categorical variables were based on a similar Q-test statistic to examine whether the variability between categories (subgroups in the study) was larger than the variability within studies. The influence of the time variable on ES was examined based on both the BDI and the HDRS, in addition to the remission rates. The associations of the other moderator variables and ES were examined based on the BDI measure, which had the largest number of studies.

Publication bias and heterogeneity

Heterogeneity was calculated as I^2. This is an intuitive and simple expression of the inconsistency of the studies' results (Higgins et al., 2003). To identify any publication bias or

undue outliers, visual inspections of the funnel and forest plots were performed, and Duval and Tweedie`s trim and fill method was used.

A non-linear re-analysis (Paper II)

In addition to encompassing a broad theoretical discussion of the findings in paper I, this study consists of additional statistical analyses performed on the same dataset as paper I.

For analyses regarding the BDI, we first visually inspected the scattering of the weighted ESs in the regression line and noticed that the decline was fairly stable until the year 2001. Moreover, the reported ESs between the years 2001 and 2014 seemed slightly inverse u-curved rather than completely flat, as the piecewise model suggests. To examine this possibility, we specified a segmented model consisting of two parts: a linear part describing the whole time period (1977-2014) and a quadratic part describing the time trend following the breakpoint. The breakpoint was empirically chosen by searching for the publication year that could render both parts of the model to be statistically significant. This only happened if publication year was centered at the year 2001. The nonlinear model also yielded the highest model fit in terms of the R-square index.

For the HDRS, the nonlinear model (similar as above), which fit best when centered at the year 2001 (R2 =.298), was not better than the best piecewise model centered at 1998 (R2 = .290). The effects for the HDRS were thus best described by a piecewise model.

Group CBT (Paper III)

The design and procedure for this article was similar to paper I. However, some methodological and statistical adjustments were made in order to further improve the reliability and validity of the results. The relevant adjustments are reported in the following sections.

Data collection & selection criteria

All Search queries returned 26,479 studies. By examining their titles, the abstracts of 934 papers were read by the first author to judge their relevance. Following that review, 181 papers were obtained from the university library. The same exclusion criteria as for paper I were then implemented, except now the delivered therapy had to be GCBT, and the included trials had to be randomized controlled trials (RCT`s). The selection procedure was conducted by the first author and yielded a final study pool of 37 papers.

Coding of study information and moderator variables

The following data were coded from the papers: demographic information (gender and age), year of implementation of the intervention, country in which the intervention took place (U.S. vs. the world), duration of treatment (number of sessions), type of therapist (psychologist or trained student), and information about the severity of the diagnosis (mild, moderate, severe, or recurrent depression). The depression diagnoses of the patients were set according to the original authors' definitions. If unreported, we categorized the diagnoses based on the BDI pre-scores as mild (13–19.5), moderate (20–29.5), or severe (>30). The moderator "manual" was coded dichotomously, according to whether a set manual was followed or not.). Effect sizes were examined as to whether they co-varied with any of the moderator variables listed above. As the number of trials using HDRS-scores was low, no

sub-group analyses were performed for this outcome measure. The RCT-PQRS was used to rate the methodological quality of the published studies. Coding of all variables was completed by the first author.

Effect sizes and statistical calculations

For the first analysis, which included all 35 RCTs using the BDI, a standardized mean difference (SMD, or Cohen's *d*) was calculated for the intervention group (Mpre–Mpost, divided by the standard deviation of the change score). A Hedge's g correction was applied to the SMD. All interventions consisted of GCBT that used a RCT design.

For the second analysis, which consisted of the 16 RCTs that included a no-treatment control group, the effect sizes were calculated from the difference between pre- and postscores on the BDI for the GCBT group and the no-intervention group, respectively, and then standardized using the change scores. For the third analysis, the same procedure as for the first analysis was used (a within-study design), but this time the outcome measure was the HDRS.

Although all studies in the present meta-analysis were originally randomized controlled trials (RCT), several of the RCT studies only included other treatment comparison groups (e.g., another psychotherapy group or medication treatment arm), and hence did not include a no-treatment control group (e.g., a wait-list group). These studies were coded as within-study designs. The remaining studies, which included a no-treatment control group, were thus coded as controlled trials in our analysis. In addition, these studies were also coded as within-studies and added to the separate within-study design pool. Thus, the two statistical conditions in this meta-analysis were kept separate from each other when calculating results.

Mindfulness-based CBT (Paper IV)

Design and procedure

This paper largely followed the templates from paper I and paper III. It had an exploratory meta-analytic design, with a focus on identifying time-trends with relevance to effect sizes. A systematic search was conducted in 2018. The following inclusion criteria were applied: 1) MBCT was given in a group format aimed at reducing depression; 2) participants were adults (≥ 18 years of age) diagnosed with depression or showing elevated scores on the BDI (> 13) or the HDRS (> 8), as a group; 3) a version of the BDI or the HDRS was used as an outcome measure; and 4) publication was in English and was in a peer-reviewed journal.

Studies were excluded when 1) MBIs other than MBCT were examined, 2) no treatment effects for MBCT were investigated or reported, 3) depression was not the principal problem of the participants; 4) partial or complete sample overlap with a study already included in the meta-analysis was observed, 5) information necessary to calculate ES (i.e., means and standard deviations) was lacking, or 6) only dichotomous outcomes (e.g., relapse) were reported.

Moderator variables & coding

For each study included in the meta-analysis, the following information was extracted: 1) year of publication; 2) sample size of the MBCT group and the control group; 3) mean age and percentage of females in the MBCT group; 4) number of sessions; 5) modification of the treatment manual by Segal et al. (2002; 2013); 6) use of the BDI or BDI-II as outcome measure; 7) no treatment vs. active treatment comparison groups; 8) randomization of participants; and 9) reporting results of intent-to-treat (ITT) analyses. For the meta-analytic calculations, means and standard deviations of the BDI and/or the HDRS at pre-treatment and post-treatment were extracted for the treatment group and, if present, for the control group(s).

To assess the methodological quality of the studies included in the meta-analysis, the Jadad scale (Jadad et al., 1996) was used.

Effect sizes & statistical calculations

To obtain the ES for each study, the SMD between the intervention group and control group, and/or the pretest and the posttest was calculated correcting for bias (Hedges' *g*). This procedure is intended to prevent the inflation of small samples` effect size, but visual inspections of the funnel plots revealed that one study in the within-group pre-post score design (Alboghasemi et al, 2015) had much larger ESs as compared to all the others studies, across statistical conditions. However, the funnel and forest plots did not seem to indicate large skewness, so the decision was made to include the trial in the specific analysis.

The analyses were conducted separately for controlled studies (between-group) with and without active treatment comparisons and pre-post differences (within-group) and for the BDI/BDI-II and HDRS as outcome measures. When data for ITT samples were available, these were preferred over data from completer samples. This choice reflects a methodological shift over time for treatment trials and RCTs, in the sense that ITT scores are more likely to be reported than scores based on completers in contemporary studies. To examine publication year as moderator for the pooled ES, a meta-regression analysis was used.

Ethical considerations

As papers I – IV in the present thesis consisted of summarizing and analyzing the data and results from previously published articles in recognized journals, no specific ethical considerations were deemed necessary. Caution were being made for interpreting and communicating the findings in an objective and prudent manner, while still maintaining a distinct and informative presentation of the results.

Summary of papers

The current summary presents and discusses the results for the individual studies in a prudent and pragmatic manner. For a detailed observation and discussion of all results, statistics, tables, and figures, please observe the relevant sections in the original and published manuscripts for the individual articles.

Paper I*: The effect of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis.*

Results

The average weighted effect size for the BDI (k = 67) was g = 1.58 (CI.95 = 1.43 to 1.74). For the HDRS (k = 34), the average ES was 1.69 (CI.95 = 1.48 to 1.89).

The CBT effect sizes based on the BDI had a significant negative relationship with time, i.e., publication year ($p < .001$). According to a sub-group analysis, a similar negative relationship was evident among studies using within-group designs ($p < .001$), and controlled trial (CT) designs ($p < .05$). Please observe figure 1.

The effect sizes for the HDRS showed a comparable picture, as ES decreased with time ($p = .01$). See figure 2. The significant negative relationship was evident for the within-group design studies ($p < .01$). The ES in the CT studies also showed a declining direction, but it was not significant ($p = .51$). The remission rates (defined according to percentage of patients achieving sub-clinical levels of symptoms, as measured by the BDI) were negatively related with publication year ($p < .01$).

Several extensive subgroup-analyses confirmed the main finding of a decrease in effect sizes with time, while, in contrast, the waiting list control group condition exhibited no significant changes in effect sizes across time ($p = .48$).

Publication bias for ESs was thoroughly investigated, leaving the conclusion that there did not exist any undue effects to the time-trend results.

A separate analysis for each moderator variable was conducted. For client-related variables, the moderator gender variable was significant ($p < .05$). Studies that included a higher percentage of women demonstrated a better treatment effect than studies consisting of more men.

For therapist-related variables, the effect size differences between types of therapists were significant ($p < .01$), indicating that trained psychologists achieved better treatment effects ($g = 1.59$) than did psychology students ($g = 0.98$).

The weighted correlation coefficients between *time* (publication year) and the moderator variables showed the following statistically significant relationships: Pre-intervention score BDI ($r = .26$, $p = .04$), severity (mild-moderate-severe) of depression ($r = -.04$, $p = .78$), methodological quality ($r = .43$, $p < .001$), and BDI (I vs. II) version ($r = .59$, $p < .001$). These analyses indicate that the methodological quality has improved significantly over the years. Newer trials also include more patients with higher initial BDI scores as compared to the older trials. None of the other moderators specified in the "methods" section showed a statistically significant result.

Finally, we examined whether the observed decline in the treatment effects depended on any of the investigated moderators by conducting two-way interaction tests (*time* × *moderator*). None of the moderators showed a significant interaction effect. Although the moderator "manual use" was not significant, it is interesting to note that studies using the Beck manual showed an even steeper decline than studies that did not use it.

Figure 1. The plot portrays the negative change ($p < .001$) in BDI effect sizes across time ($k = 61$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

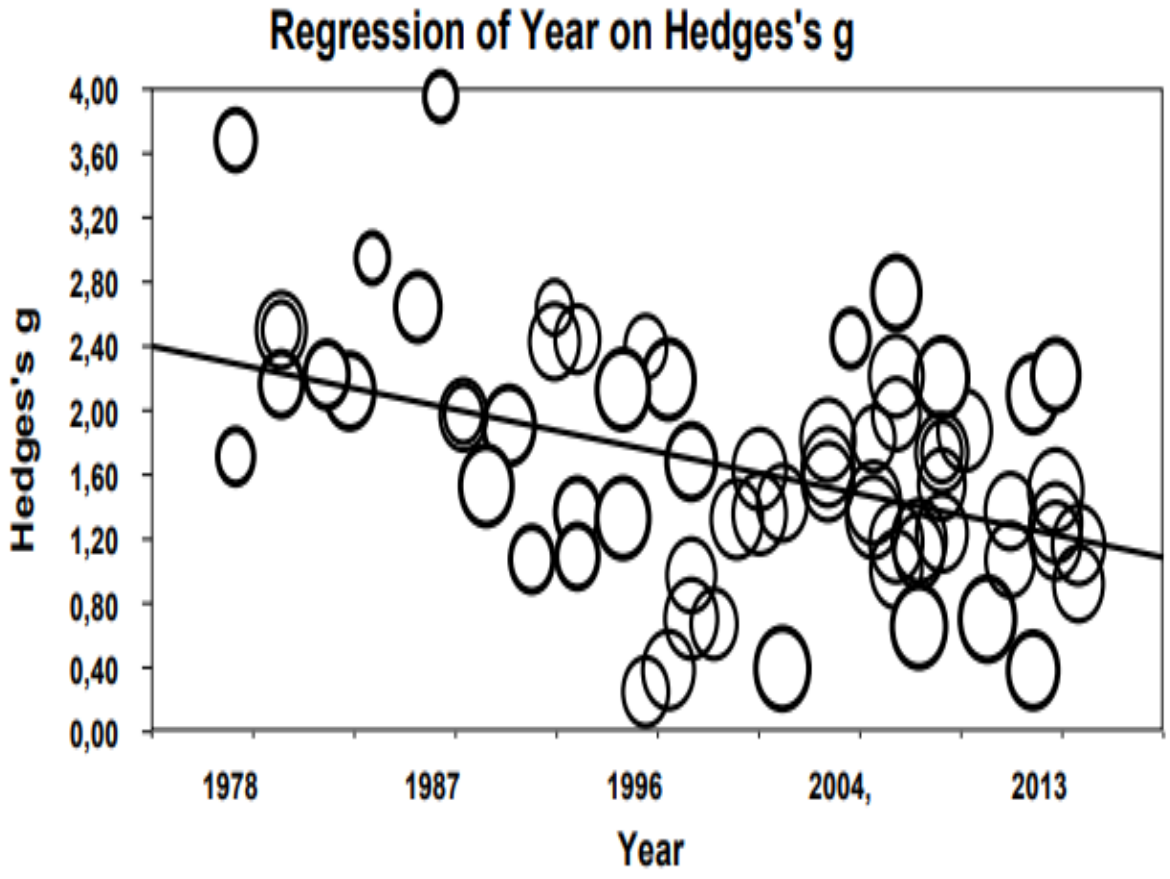
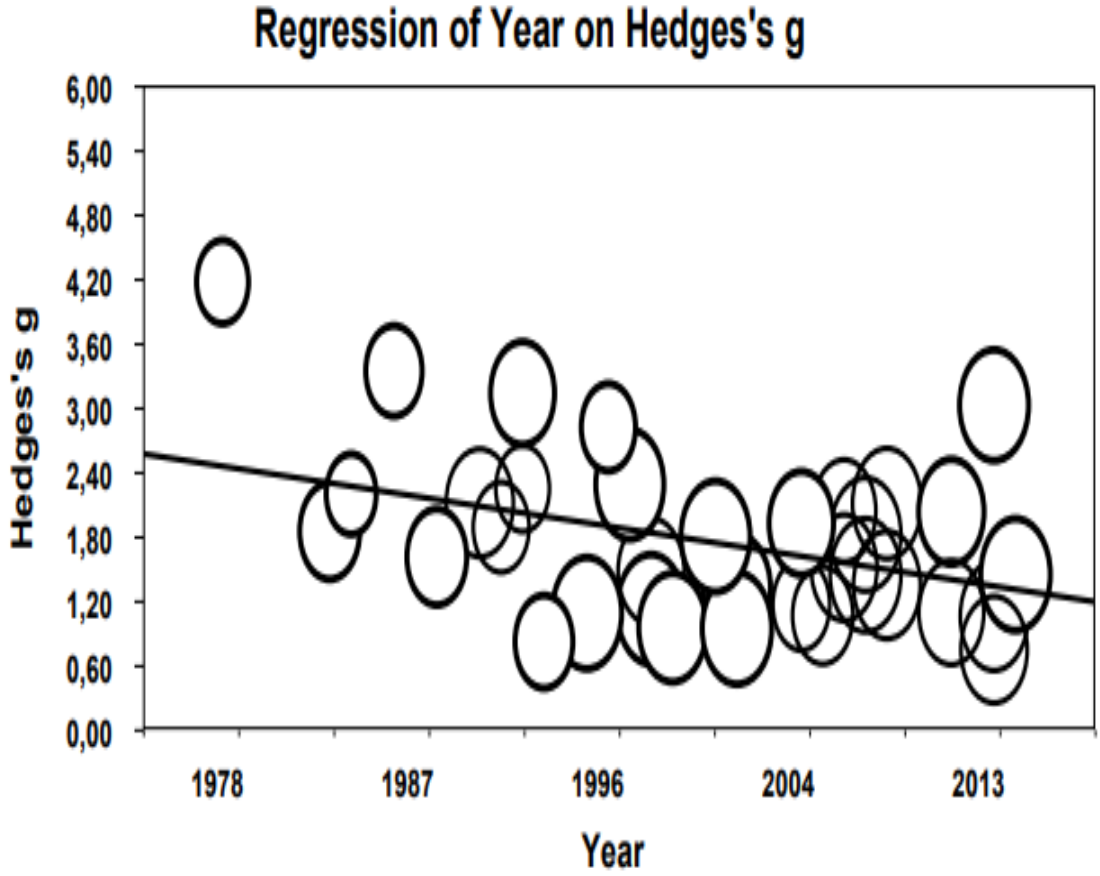


Figure 2. The plot portrays the negative change ($p < .01$) in HDRS effect sizes across time ($k = 34$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.



Discussion

The main finding was that the treatment effect of CBT showed a declining trend across time and across both measures of depression (the BDI and the HDRS). Contemporary clinical treatment trials for individual CBT therefore seem to be less effective than the therapies conducted decades ago. The declining effect of treatment over time seems robust.

We did not find evidence of significant differences in the treatment effects resulting from the use of the Beck manual (Beck et al, 1979) or not. However, the interaction analyses showed a slightly steeper decline for the CBT trials that used the manual compared to those that did not. This finding was rather surprising given that the original manual had a reputation among clinical researchers as one of the best ways to implement CBT. To the best of our knowledge, there have been no thorough investigations of how different ways of conducting CBT for depression may influence the outcome. Our findings indicate that further investigations regarding this matter are warranted.

The major practical impact of this study would be to heighten the awareness among practitioners and clinical researchers of the trends in modern psychotherapy and stimulate research on the highly relevant topic of time-trends for effect sizes. If psychotherapy of today has a lower efficacy than that conducted 30 to 40 years ago, this threatens the validity of current comparative studies. If we compare the efficacy of a new psychotherapeutic approach with the current best standard, which, for example, may be CBT, we risk concluding that the newer approach is preferable even though it may have a weaker effect than the seminal CBT trials of the 1970s. Researchers conducting randomized placebo-controlled trials today, thus, risk implementing newer treatment approaches that are relatively better than the current best CBT**.** Yet, what is the benefit of doing so if the absolute change is minor or even negative compared to the seminal studies?

Paper II: *The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: Reply to Ljòtsson et al.*

Results

The analyses indicate that CBT treatment effects, as measured by the BDI, have fallen linearly from 1977 until 2001, and not until 1995 as proposed by Ljótsson et al. (2017). The

fall has been going on for about 24 years, which encompasses half of all studies (33 of 67). From 2001 onward, the treatment effects have not declined further, but stability in the effects cannot be claimed due to the significant segmented quadratic model. This model shows a temporary rise followed by another fall, which may or may not be ongoing. Whatever is true, the major point is that a flattening in the treatment effects of CBT, or that the CBT effects now vary around their "true" value, as Ljótsson et al. (2017) conclude, is not well supported by the available data. The segmented nonlinear model with the publication year 2001 as the breakpoint also explained 2.4% more of the variation in the treatment effects than the piecewise model with year 1995 as the breakpoint.

Regarding the HDRS effects, the piecewise model did indeed fit the data best, as suggested by Ljótsson et al.

Discussion

The present reanalysis does not change the basic message stating that CBT effects have fallen considerably across two and half decades. In fact, the predicted ES for the year 2014 even comes out slightly worse for the segmented nonlinear (g = 1.12), as compared to the linear, model (g = 1.18). Nevertheless, the current ESs are strong, hence CBT is still to be considered as an effective anti-depressive treatment.

Since the current reanalysis shows that the nonlinear time trend explains a considerable portion of the between-study treatment variance (almost 30%), future meta-analytic summaries of treatment effects should not dismiss potential time trends.

To conclude, since the BDI effects during the last 13 years do not follow a flat trend but rather are in decline again, we believe a weather-climate analogy is an apt comparison: although weather varies across decades, the long-term climate changes (as projected by a linear model) may be regarded as the most reliable indicator.

Paper III: *A meta-analysis of group cognitive-behavioral therapy as an anti-depressive treatment: Are we getting better?*

Results:

The average weighted BDI effect size for all studies, based on within ESs (k = 35), was g = 1.33 (CI.95 = 1.16 to 1.50. For the HDRS, using a within-group design (k= 14), g was 1.56 (CI.95 = 1.20 to 1.90).

For the first analysis (a within-study design), the GCBT effect sizes based on all 35 RCT studies improved significantly with time as measured by the BDI (p = .02). See figure 3. For the second analysis (the between-group condition with the BDI as outcome measure), the ESs based on the 16 studies that included a no-intervention control group also improved significantly across time (p < .001). For the third analysis (a within-group design), effect sizes based on the 14 studies using the HDRS showed a non-significant trend towards a decline with the passage of time (p = .07, see figure 4).

Publication bias was inspected and deemed to be without any significant effect on the meta-regression line. Heterogeneity scores were in the moderate-to-high range for all analytic conditions, ranging from 46 to 84 percent. Separate moderator analyses were based on the BDI scores and revealed higher ESs for studies where no specific treatment manual was followed (p = .03). Further, additional analyses showed that for trials not using a manual, there was a significant improvement in ES with time (p < .01). This relationship between ES and time was not found for trials that used (or reported) a treatment manual.

For the variable country, there was a tangible but non-significant tendency towards higher ESs for studies performed outside the U.S. compared to studies performed in the U.S. (p = .06, see Table 3). Separate follow-up analyses revealed that trials conducted exclusively in the U.S had a non-significant tendency towards improvement in ES with time (p = .06), while trials conducted in the rest of the world showed no signs of improvement with time.

Figure 3. The plot portrays the positive change ($p < .02$) in BDI effect sizes across time ($k = 35$). The size of the circles indicates the relative contribution (random weight) of each study.

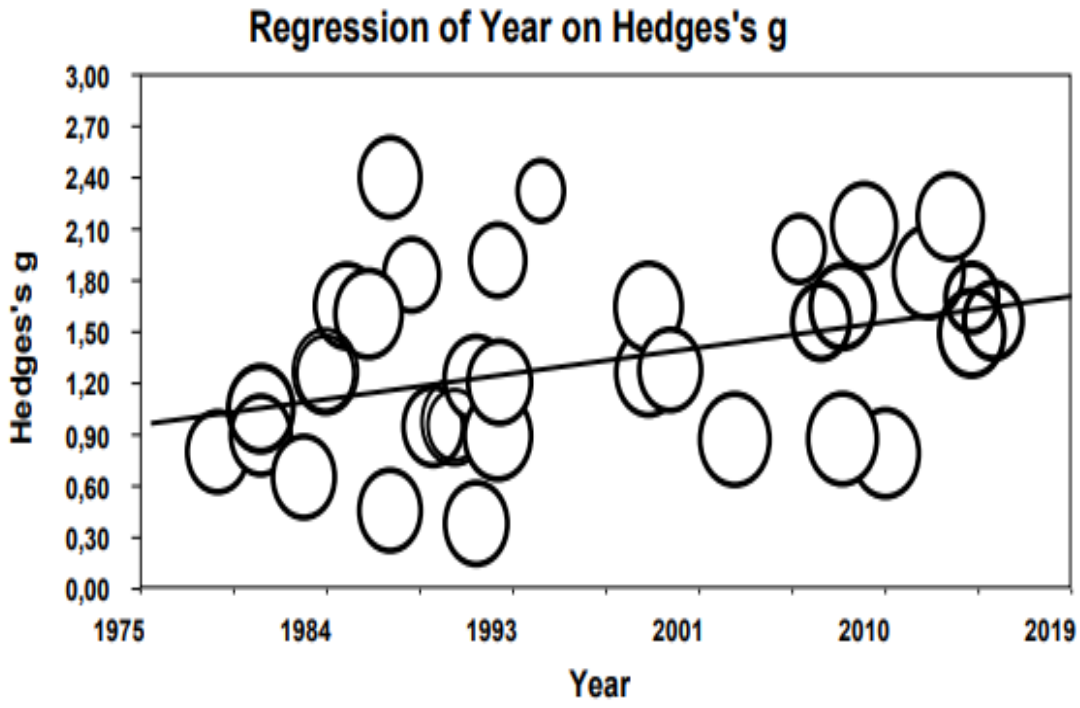
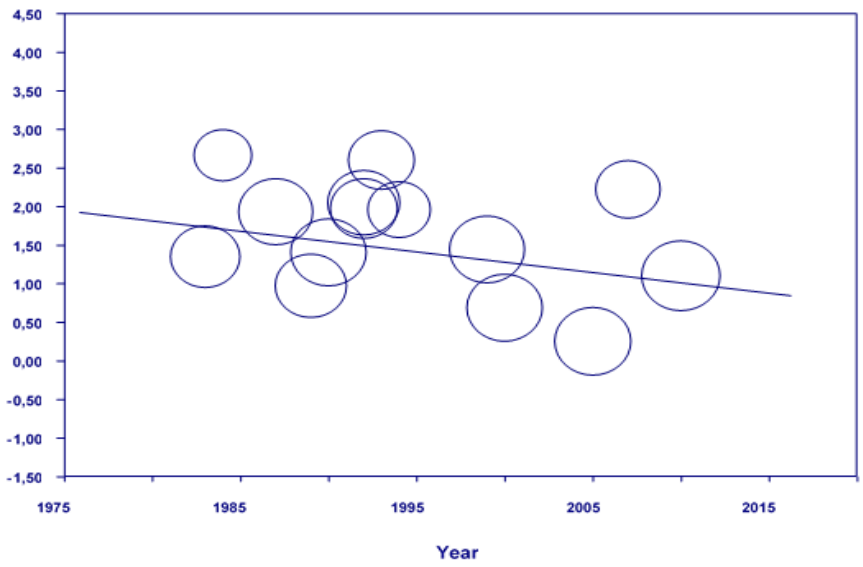


Figure 4. The plot portrays the tendency ($p < .07$) in HDRS effect sizes across time ($k = 14$). The size of the circles indicates the relative contribution (random weight) of each study.



Discussion

The analysis showed a significant increase in effect size across time, as measured by Beck's Depression Inventory. For the HDRS, the analysis did not confirm the findings of larger ESs with time. The discrepancy between the results from the two outcome measures was surprising, given the moderate to high correlation previously found between the BDI and the HDRS. Sub-group analyses and investigations over heterogeneity and potential bias were performed. The results suggested that the non-significant relationship between time and ES, as found in the initial HDRS-based analysis, were not as reliable as the significant relationship found by the BDI-based analyses.

The present analysis found a negative relationship between the application of a treatment manual and outcome. Trials not following (or reporting) a manual had a significantly larger ES than those that did report use of a manual. This finding is consistent with a previous study that have found negative effects from the use of treatment manuals (Duncan & Miller, 2006). Furthermore, sub-group analyses showed that for trials where no manual had been used, there was an increase in treatment effect sizes with the passage of time. However, for trials using a treatment manual, there was no increase with time. This finding, not explained by differences in study quality, could be interpreted in several ways, but indications are that with regard to treatment efficacy for GCBT, the specific factors (techniques applied in treatment) do have a substantial bearing on treatment outcome; when following a set routine, treatment effects do not improve with time. The results of this analysis indicate that a shift in perspective should be considered regarding the most efficient way of implementing GCBT, as there seems to be a beneficial effect in varying and adapting the interventions, perhaps according to the patient group. The highly structured manuals could thus be replaced by more adaptable forms of GCBT, or at least by frequently updated manuals.

Paper IV: *Time trends in the effects of mindfulness-based cognitive therapy for depression: A meta-analysis*

Results

The average weighted ES for between-group studies using a no-treatment control group and the BDI as an outcome measure (n = 11) was g = 0.92 (95% CI [0.70, 1.14], *I2* = 42.7). For the between-group studies using the HDRS and a no-treatment control group (n = 7), the ES was g = 0.80 (95% CI [0.61, 0.99], *I2* = 16.04).

Analyses of time trends showed no significant relationships between year of publication and ES.

None of the chosen moderator variables were found to moderate the observed ESs of between-group studies. For within-group comparisons using the BDI, sample size (b = 0.00, 95% CI [-0.01, 0.00], p = 0.013), age (b = -0.08, 95% CI [-0.12, -0.03], p < .001), and baseline depression (b = - 0.04, 95% CI [0.01, 0.06], p < .01) were significant moderators. Thus, smaller sample size, lower mean age, and higher baseline depression wes associated with higher ESs.

Discussion

The results suggests that treatment outcomes were stable over time. It can only be speculated whether the reported ESs of MBCT for current depression already represent the upper limit of its effectiveness or whether factors such as insufficient therapist training and supervision (cf. Waltman et al., 2016) inhibit an increase of the effects. Another possible factor is that there simply has not yet been a large enough timespan to detect any significant trends related to time.

A potential confounding factor could relate to the methodological choice of mixing scores from completers with scores from ITT. However, an inspection of the included studies`

plots revealed that the nine trials coded as ITT are evenly spread across the time range, and ESs did not seem to deviate from that of completers. Previous investigations have typically revealed rather small differences in effect sizes between ITT and completers. An extensive meta-analysis (Hans & Hiller, 2013) showed largely comparable effect sizes in depression treatment trials using a statistical design requiring treatment completion (d = 1.13), and those using an ITT design (d =1.06). This finding was replicated in a later study (Johnsen & Friborg, 2015), where no significant differences between ESs of completers vs ITT were found. Nor did a two-way interaction test (completers x ITT) reveal any significant impact on the developments of ESs with time.

As to the overall effects of MBCT for acute depression, the results of the present study are consistent with previous meta-analytic studies (e.g., Lenz et al., 2016), suggesting that MBCT is effective in reducing symptoms of depression. Applying Cohen's (1992) criteria, the average ESs for between-group studies comparing MBCT to no-treatment control conditions and pre-post studies were large for both outcome measures.

Based on the previous robust findings of the effectiveness of MBCT for current depression, it has been proposed that MBCT should be offered as a first-line treatment for depression on equal terms with other evidence-based treatments (Strauss et al., 2014). However, more research is needed to support this claim. It should be noted that the average ES observed for MBCT when compared to no treatment comparisons is lower than those for other psychological treatments. For example, for pre-post comparisons, the average ES for MBCT observed in the present study (g = 0.90) is smaller than those for individual and group CBT in clinical trials (g = 1.65 and g = 1.33, respectively; Johnsen & Friborg, 2015; Johnsen & Thimm, 2018) and in routine clinical practice (d = 1.06; Hans & Hiller, 2013).

General discussion

Comparing CBT, GCBT and MCBT

Effect sizes

As the constituting meta-analyses in this thesis utilized identical ways of calculating effect sizes, comparisons across papers are allowed. When using a within-group calculation, the average ES for GCBT and individual CBT were comparable. Based on the HDRS, *g* was found to be 1.56 for GCBT and 1.69 for CBT. Average ES`s based on the BDI was 1.33 (GCBT) and 1.65 (CBT) for the within-group format, and 1.14 (GCBT) versus 1.37 (CBT) utilizing a between-group design. The slight difference in effect sizes could partly be explained by somewhat larger pre-scores for patients entering trials with individual CBT. They averaged a score of 26.1, versus 24.2 on the BDI for patients in treatment with GCBT.

The similarities in treatment outcome between CBT and GCBT highlights the question regarding which format of therapy to be prioritized in battling depression. Taking into consideration the beneficial cost-effectiveness of group therapy, this avenue deserves further investigation. Recent meta-analyses have shown that the monetary benefits involved in the application of group CBT does not seem to compromise treatment efficacy significantly, as the outcome differences between individual and group CBT trials are only slight to moderate (Hans & Hiller, 2013; Huntley et al., 2012). However, these reviews do not take into consideration the fact that group CBT seems to have increased in effect size over the years, while the opposite is found for individual CBT.

Mindfulness-based CBT revealed somewhat smaller effect sizes, as compared to the more classical formats of CBT. For the BDI, the within-group and between-group conditions showed $g = 0.90$ and 0.92, respectively, while *g* based on the HDRS was 0.80. Although a firm conclusion would be hasty, there is little evidence to suggest this newer form of therapy is more effective as an anti-depressive treatment than its elder siblings. A more thorough

discussion on the topic of effect sizes and treatment outcomes follows in the section "explaining the results".

Moderator variables

The last couple of decades has generated a substantial amount of interest and research towards revealing the underpinnings of effective treatment. Such efforts have yielded a commonly accepted notion stating that there are four overriding factors contributing to ESs in psychological treatment. As these factors also could be significant for understanding temporal developments, a discussion of relevant findings associated with this thesis is considered informative. The four major factors consist of variables and characteristics linked with the therapist, the client, the treatment technique itself, and factors common for every treatment-format.

Therapist-related

When it comes to type of therapist as a moderator, the different analyses revealed different outcomes. For GCBT there was no difference in effects observed as a consequence of having a trained student therapist rather than a psychologist, but for individual CBT the results favored the more experienced psychologist. This relationship was not investigated for MCBT. Although these findings are interesting in themselves, there are no indications suggesting these results do have a moderating effect on the general time-trends regarding treatment effectiveness. This especially holds true as a two-way interaction test (Time x Type of therapist) for individual CBTs showed no relevant interaction effects.

Another variable connected to the therapist-related factor, is the competency of the therapist. Unfortunately, very few studies rated or reported this moderator, preventing an informed discussion to be performed.

Client-related

The studies within this thesis collected and applied demographic information in their analyses. The only exception was for paper II, where of course the sample was the same as for paper I. Consistency was observed across studies, with the average age of the clients ranging from 39 years in the sample from paper IV, to 40.5 years in paper I. The proportion of men in the studies ranged from 24.3 to 40.2 percent. This is in accordance with general statistics revealing that significantly more women do indeed receive professional treatment for unipolar depression (WHO, 2020). Diagnostic severity, as measured by pre-treatment scores on the BDI, were also quite similar across the studies, ranging from 23.3 to 26.1. Thus, the distribution of age, gender and severity was largely comparable for clients entering trials for CBT, MCBT and GCBT across decades.

In concordance with previous research there were not observed any consistent or large effects related to the variables gender and age when it comes to moderating ES. For individual CBT, the analyses demonstrated that woman profited more from therapy than did men. This pattern was not evident for GCBT or MCBT. When examining the variable age, the meta-analysis based on results from MBCT revealed that younger clients benefited more from therapy than did elder ones. This pattern was not observed for CBT or GCBT.

For the moderator diagnostic severity, the study based on MBTC found that higher levels of baseline depression yielded larger ESs. A similar pattern was showed for CBT, were patients with milder levels of baseline depression tended to achieve lower ESs. For GCBT there was no such statistically significant result, although the data do indicate a potential difference in ES between mild depression (g = 1.18), versus severe depression (g = 1.38). The relatively low number of trials consisting of patients with higher rates of severity (k = 5) most likely limited the statistical power.

For individual CBT, paper I investigated whether levels of diagnostic comorbidity (for both physical illness and/or psychological disorders) and psychotropic medication affected treatment effects. The results showed no significant moderation of ESs. The paper also investigated correlations between the different moderators and time. Here, one significant correlation appeared, revealing that newer studies had higher severity levels on the BDI compared to older ones. As higher levels of severity are associated with larger ESs, it is interesting to observe that the fall in treatment effects with time show little sign of being halted. This is another robust evidence for a genuine fall in ES with time for individual CBT.

The above findings are interesting in the broader scope of research on treatment effects, and some of the results could also provide valuable information when it comes to interpreting time-trends in treatment effects, as well as support the conclusions. This is discussed more thoroughly in the section "explaining the results". However, for the main purpose of this thesis, the major factor of concern is whether any of the client-variables show an interaction-effect with time as a moderator. If that were the case, this would indicate systematic changes in the moderator itself (for example, more women entering therapy in later years than previous ones) being accountable for differences in ES as a consequence of time passing by. However, across all the included meta-analyses, such interactions did not appear. This leads to the conclusion that client related moderators generally do not exert any confounding influence on the temporal development of effect sizes.

Common factors

Perhaps surprisingly, there were very few studies that investigated or reported variables of interest related to the common factors. Even the variable working alliance, which has achieved a vast amount of attention for the last decades and is recognized as the most influential moderator for ES, lacked the necessary number of studies reporting a quantifiable

measure for any statistical comparison to be useful. This is a finding that points to the need for developing, and utilizing, a universally accepted scale for measuring the working alliance between therapist and client. Such a scale should aim to provide useful information that could aid both clinical and statistical purposes. There have been developments made in recent years, for example the working alliance inventory (Horvath & Greenberg, 1986). Still, these developments have yet to be translated into practice when it comes to clinical trials measuring the effects of anti-depressive treatment.

Encompassing a broader perspective, study quality is also a common factor for published papers investigating treatment effects. For this moderator, perhaps surprisingly, none of the included meta-analyses found a significant impact on ESs caused by variations in study quality. For the MBCT-study, a likely explanation is the short timespan between studies (10 years), in which time it is implausible that large developments in study quality has occurred. The studies on CBT and GCBT showed a tendency indicating smaller ESs with better study quality, however, interaction-analyses featuring *time vs effect size* revealed no relationship of statistical significance. Summarized, the results do lend some support to the notion that better-quality studies achieve lower ES, although this effect is not very pronounced when it comes to CBT-based anti-depressive treatment trials.

Technique-related

As shown by the papers included in this thesis, cognitive behavioral therapy is considered and utilized as a short-term therapy for depression, with sessions typically ranging from 12 to 20 per treatment period. Within this range, none of the included meta-analyses found variations in treatment effects because of number of sessions provided. These findings support the use of CBT as a short-term therapy; it`s main effect is achieved regardless of

having 12 or 16 sessions, and the results do not indicate any increase (or decrease) in efficiency with extended sessions of therapy.

The use of, and adherence to, a structured treatment manual, has been a major part of the foundations for CBT. The commonly accepted view in the field has been that diligently following a manual, and preferably the original manual developed by Beck, will yield the best outcome for the patients. This notion is heavily challenged by the findings from the analyses included in this thesis. For GCBT, there was a significant difference in ESs, favoring trials that did not use (or report) a set manual. For the meta-analysis investigating individual CBT there were no differences in overall effect sizes related to the use of the original Beck manual or not. For MBCT, all included trials were based on the original treatment manual. Summarized, the findings lend support to the notion stating that the use of a set manual is not doing effect sizes any favors when it comes to CBT-based treatment for depression.

Closely connected to the variable "use of a manual", adherence checks were reported to have been used by many individual CBT-trials. Although not explicitly investigated in our studies, indications are that the trials reporting ratings of adherence, are also those who use a set manual. The results of the analyses showed no differences in effect-sizes between trials utilizing adherence checks versus those who did not utilize such checks. This finding is thus in accordance with other results, suggesting that the use of a structured manual is not associated with higher ESs.

Geographical differences

One article in the present thesis investigated geographical differences. For GCBT, there was a trend approaching significance ($p = .06$) related to country of origin, with trials conducted outside the U.S. showing higher ESs. Follow-up sub-group analyses also showed that the ESs of studies conducted in the U.S. increased with time, a trend that was not evident

for trials conducted in the rest of the world. With time, studies performed in the U.S. have improved, reaching a present level of efficacy comparable to that of trials conducted in the rest of the world. This development is interesting, particularly when considering a similar, yet opposite, finding from another recent meta-analysis concerning individual CBT (Cristea et al., 2017). In that study, ESs from studies performed in the U.S decreased with time, while ESs for trials from the rest of the world remained unchanged. However, the authors concluded that the observed effects most likely were spurious, mainly due to high levels of heterogeneity. Nevertheless, the common connotation for both CBT and GCBT seems to be that the largest temporal developments in treatment effects, regardless of direction, are limited to the U.S. One explanation for this finding could be that the range of publication years is larger for trials conducted in the U.S., thus increasing the potential for significant results to occur. Other explanatory factors are being discussed in the section "explaining the results".

Time-trends for treatment effects

The common theme and main topic of investigation for this thesis and its associated meta-analyses and articles, was to discover, describe and explain relevant time trends in the treatment of depression, specifically for the different formats of cognitive behavioral therapy. Although the foundations are similar for individual CBT, group CBT and mindfulness-based CBT, the investigations revealed large variations in temporal developments and trends. For individual CBT there exists a robust and significant decline in ESs, while group CBT shows signs of improvement with time. For the newer mindfulness-based CBT, there is no development with time, as ESs have stayed at the same level since its inception. The completely flat regression line exhibited for MBCT is most likely due to two factors, one being the standardization of treatment as described earlier. The other contributor to the nil-finding, is most likely the span of publication; all included studies were published between

2007 and 2017. This ten-year range represents a limited period for clear trends or tendencies to emerge.

Clinical meaning of the results

The primary objective of the meta-analyses included in the current thesis was to investigate and explore the temporal development of treatment effects, quantified as statistical effect sizes. The secondary objective was to identify any potential moderating variables to the observed ESs, which could assist in explaining the results. However, the statistical and methodological considerations do not pinpoint the practical significance or clinical meaning, which are the factors of most interest for therapists and clinicians. It is therefore important to assess the clinical implications as closely as the available data allows.

First and foremost; the three anti-depressive treatments investigated in the meta-analyses all enjoy a substantial treatment effect. CBT, GCBT, and MCBT will help patients get better from their depressive symptoms. However, the analyses also indicate that more patients will have a larger reduction of symptoms if they are treated with CBT or GCBT, as compared with MCBT. When considering that a difference of 0.2 - 0.4 *g* (which represents the difference in ESs between CBT and GCBT, as compared to MCBT) equates to a clinically meaningful change, these differences do not seem neglectable. Thus, clinicians should be aware that more patients will be likely to experience clinically meaningful change with the traditional forms of cognitive therapy.

When observing the temporal development of MCBT, the clinical implications are limited. No change has been observed with time, as treatment effects have been stable. Stability is thus the main message; therapists know what to expect and have little reason to suspect any weakening or strengthening of the clinical outcomes for the immediate future.

The temporal development for GCBT conveys another picture. Treatment outcomes have steadily increased with time, and ESs can now be expected to be roughly 0.5 - 0.6 *g* higher when compared to the trials of the 1980`s, as measured by the BDI. This statistically moderate difference in ES equates to a clinical meaningful change for patients, who on average can be expected to enjoy a 5-6 points larger reduction in depressive symptoms with contemporary treatment. Moreover, the data indicate that the improvement is ongoing, which should be of interest for therapists, treatment clinics and institutions, as well as policy makers. As the average pre-treatment BDI score for patients in GCBT trials for depression treatment was 24, a gain in symptomatic reduction of 5-6 points would mean that a substantial number of patients are expected to reach sub-clinical levels after treatment, thus being cured for major depression.

This is not the case for individual CBT, where ESs have fallen by roughly 50 percent since the 1970`s. The decline equates to a strong effect size of $g = 1.1$, which approximately equates to a 10-12 point difference in scores pre/post, as measured by the BDI and HRSD (!). When taken into consideration that a change of 4 points on these scales indicate that a clinical meaningful change has occurred, the potential clinical implications are huge. An example could serve to illustrate the impact: If a patient entered therapy with a diagnosis of severe depression, and a pre-treatment score of 31 on the BDI and HRSD, the reasonable expectation with contemporary treatment would be to achieve a post-treatment score of 15-22, and a change in diagnosis to mild depression. While such a reduction may seem quite impressive at face value, it must be noted that had the same patient entered therapy in the 70`s, the expected outcome from treatment would be a post-treatment score of 5-12, with clinical remission achieved from the depressive disorder. As individual CBT is the most widespread and utilized anti-depressive treatment format, it should be of profound clinical importance to invest efforts

in arresting the declining tendency in the treatment effects for CBT, and search for methods to regain the treatment outcomes of decades past.

The meta-analyses comprising the present thesis have not measured treatment outcome at any follow-up intervals, and thus can make no firm conclusions regarding eventual time-trends beyond post-treatment. However, as scores on symptom checklists for depression typically are comparable at post treatment and follow up, it does not seem plausible to expect considerable deviations from the time-trends found at post-treatment.

Explaining the results

Although the exact underpinnings of the current findings were not thoroughly or specifically investigated in the different articles building up this thesis, there is enough information in the results and analyses to make several qualified reflections and hypotheses regarding what causes the observed effects.

The placebo effect

The placebo effect is a recognized ingredient for all forms of treatment, both somatic and psychological ones. Typically, it accounts for between 20 and 35 percent of treatment effect. An important variation in the placebo effect happens when a new form of treatment is available, and a spike in effect is observed in the beginning of its lifespan. This effect is especially pronounced the more hyped and build-up the new treatment form is, and the more praise and plaudits it receives. When it comes to CBT in the 1970`s, there was no shyness in communicating and showcasing this new and revolutionary form of therapy. Often branded as the "gold standard" for treatment of depression, paired with a powerful shift in paradigm towards a cognitive era, the expectancies and hopes for suffering patients were probably through the roof in the early years of CBT. In recent times, however, an increasing number of

studies (e.g., Baardseth et al., 2013; Wampold et al., 1997, 2002) have not found this method to be superior to other techniques. Coupled with the increasing availability of such information to the public, including the Internet, it is not inconceivable that patients' hope and faith in the efficacy of CBT have decreased somewhat, in recent decades.

Arguments such as the above, lead to the inevitable question; could the observed fall in ESs for CBT be completely explained by wearing of the placebo effect? The answer is probably "no". First, the observed fall in treatment effects for individual CBT is not halted or stabilized after 10-12 years in existence, as would be expected if there was purely a "new-therapy" related spike. Second, when scrutinizing the meta-regression line, the observed fall in treatment effects has seemingly gone through phases. The decline was halted around the turn of the century but was again evident towards the end of the period under investigation. Although there is a possibility for these detailed analyses to be mere statistical artefacts, the main result stays the same; the ESs of CBT has been, and most likely still is, in decline.

A further question is whether a potential early spike in placebo effects could have influenced the temporal trends for GCBT or MBCT. However, there appears to be factors making this effect less significant. First, when comparing the studies on CGBT and individual CBT, the first article did not include any studies published before 1980, while the latter included five articles from the 1970`s. As both formats are cognitive behavioral, the brunt of an eventual spike would be connected to the earliest trials. In the same vein, MCBT would not be significantly affected by any initial spike connected to the brand CBT. Further, as all studies for MBCT was conducted in a ten-year timeframe, there is probably not a large enough range yet to detect any potential "new-treatment" spikes related to the invention of the term "mindfulness". Whether such an effect truly exists connected with mindfulness as a therapy form, remains largely unknown.

Study quality and sample size

When comparing trials that have been published in a timeframe approaching 40 years, there are specific procedural factors that may be suspected to contribute to the findings. One of these relates to the level of stringency connected to the individual trials, where a reasonable assumption is that earlier trials have a lower degree of methodological requirements, and thus are less robust. If this is a systematic finding, there could be a confounding factor to the results.

The best indicator we have for observing stringency, or how robust trials are, is study quality. This is a measure which is implemented and controlled for in all the meta-analyses building up the current thesis. The results do not univocally confirm that study quality has improved with time. For paper I there was indeed observed a significant increase in methodological quality with time. However, for paper III, which was covering a similar timespan as paper I, no such relationships were found. This was also the case for paper IV, though here the time range is limited. An important factor to consider when interpreting these results, is that paper III (GCBT) included a stricter selection procedure for which trials to be included in the meta-analysis, as compared to paper I (CBT). The main difference related to the design of the studies. For paper III, all trials had to be randomized controlled trials, while the requirements for paper I were more including. Based on this information, the most prudent conclusion seems to be that study quality has not increased dramatically with time for RCTs. However, trials without a randomized and controlled design are less likely to be published with the passing of time.

Correlation does not imply causation, so the observation of lower study quality in earlier trials do not necessarily explain the observed differences in effect sizes. To investigate any possible causation, a two-way interaction test was performed (time x study quality), to check for any influence on the temporal development of ES. The statistical analysis revealed

that no such interaction existed (p = .60), leaving the conclusion that temporal ESs did not systematically vary because of variations in study quality. Considering the evidence, it does not seem plausible that increased stringency for modern trials bear any significant part of the reason for the observed findings. This impression is enhanced by the findings of an increase in ES with time for paper III.

Sample size is another factor commonly thought to have a bearing on ESs and is a factor to consider when comparing trials from different eras. Papers I and III did not include a separate analysis for sample size`s influence on ESs. The assumption was that performing extensive checks for publication bias, implementing the Hedge`s *g,* and carefully investigating the funnel plots would adequately deal with concerns related to small samples. Indeed, the indicators from the tests of publication bias, and the forest plots, did point towards a potential confounding effect in the dataset for paper I, with smaller samples having larger effect sizes. As the main analyses showed significant findings, these indications were controlled for in a follow-up sub-analysis. Here, the 30 trials with the smallest samples were excluded, but the results of a falling effect with time remained significant. In paper IV (MCBT), one of the statistical conditions did find a significant, positive relationship between ES and sample size. However, when it comes to time-trends, the regression line was identical to the other statistical conditions in the paper.

Based on the combined results from the three meta-analyses, it cannot be firmly concluded that sample size did not exert any effect at all on the outcomes. However, it is not considered likely that this factor explains a significant proportion of the temporal development of treatment outcome.

Attitudes and knowledge towards mental health issues and treatment

For the past decades, there has been a positive development in western societies when it comes to openness for treatment for psychological disturbances, and attitudes towards mental illness. The tendency is that more afflicted people seek help, and that they are encouraged to do so by their friends, colleagues, employers, and families. At the same time, people with mild psychological disturbances are not subjected to similar levels of stigma and discrimination. This is probably a contributing factor for the gains in treatment effects observed for GCBT. In a group setting consisting of typically 6-10 people, the potential effects of stigma, including feelings of shame, discomfort, and anxiety, would be significant. Thus, if most clients involved in the group setting felt less stigma, ESs would be higher. This effect would not be as salient when it comes to individual therapy, as there is no-one else in the therapy room besides the client and therapist. Such a notion could help explain why the declines in ESs for CBT have seemingly not been prevented by a more open-minded society.

The article investigating MBCT also lend support to the hypothesis of a tendency towards a more open-minded society. For this analysis it was found that younger clients profited more from treatment than did the elder ones. It is considered plausible that this effect is mainly due to younger people (and their environment) being more open-minded towards treatment and psychological issues.

Internet and social media

The last three decades have seen the introduction and growth of internet and social media (SoMe). With the internet being generally available for the population, an unlimited source of potential information is but a few keystrokes away. The invention of the internet could have a significant effect on the temporal development in ES for individual CBT. Cognitive behavioral therapy is essentially an information-based form of therapy, where the

main objective is to use rational and objective thoughts and evidence to counter maladaptive feelings. Now, with the plethora of such information readily available for the population, it is likely that many patients have covered this section of treatment long before entering the room of therapy. As such, the measured treatment-effects would likely diminish. At its birth, CBT was being hailed for its simplicity, and user-friendly format. The clients could understand and participate in their own treatment, directed by books or leaflets handed from the therapist. In present time this information is freely available at home. As such, it is conceivable that the very expansion of CBT could contribute to its lower effects - as measured by treatment trials. Furthermore, as CBT is considered easier to understand and less complex than many other forms of therapy, there is a possibility that the development of internet and SoMe would have a larger negative effect for CBT, compared to other forms of (less known) therapies, for example intensive short-term psychodynamic therapy (Davanloo, 2000). This is an avenue where future research is necessary.

It is plausible that the growth of internet and social media would have a bigger impact on individual therapy than group therapy, as the focus on psychoeducation and theoretical knowledge is more pronounced for the first format. This factor could help explain some of the variations behind the different trends in temporal developments.

Standardization and the use of a manual

When it comes to interpreting the reasons for the observed fall in ESs associated with CBT, there has been widespread opinions among researchers and practitioners. Some point to treatment fidelity and adherence, strongly suggesting that a probable cause for the observed decrease in effect sizes could be found in a decrease in treatment fidelity, and/or a lack of adherence to a treatment manual, or cognitive therapy`s principles (Waltman, Creed & Beck, 2016; Waltman, Sokol & Beck, 2017). The authors call for the cognitive field in

psychotherapy to become more aware of following the principles of CBT, and for treatment fidelity to be implemented as a standard measure in research trials. The underlying assumption is that for the earlier trials for CBT, therapists and treatments had higher levels of adherence and fidelity. This notion is hard to measure in any quantitative analysis though, because most trials are lacking the relevant information, or use differing methods for measuring such phenomenon. It is thus very hard to conclude that poor fidelity or adherence are responsible for a decline in ESs.

What we do know, is that for GCBT a wide variety in regard to applying a manual was observed in paper III, with no single manual being universally embraced or utilized across time. Researchers frequently develop their own manuals before the start of treatment, and many do not follow a set manual at all. The result is both strong effects sizes, and improvement in effects with time. The analyses show a significant difference in ESs, favoring trials that did not use (or report) a set manual. Further investigations also showed that the statistically significant improvements with time were exclusively tied to the trials featuring no set manual (or not reporting one). The tendency of utilizing different manuals contrasts with individual CBT, where a large number of studies use the original Beck manual for CBT for depression (Beck et al., 1979). However, as illustrated in this thesis, this approach is not beneficial to the individual format either. Interaction-analyses consisting of trials for individual CBT showed that the steepest decline of ESs with time was tied to the trials using the original Beck manual. For the meta-analysis investigating MCBT, all included trials were based on the original treatment manual, and the timespan for investigation was limited. As previously noted, this study showed a virtually flat regression-line, with no changes in ESs with time. Thus, any conclusions concerning the use of a manual are hard to reach for this treatment format. However, there are indications that the widespread use of a set manual could have a detrimental effect for temporal improvements also for MCBT.

It is important to keep in mind that all the included trials adhere to the cognitive principles and use traditional cognitive and behavioral techniques and approaches. One of the keys to further improvements in ESs for all formats of CBT probably lies in moving away from using highly structured treatment manuals (which could lead to rigid treatment), thus paving the way for a more creative, adaptable, and intuitive way of conducting CBT. The future hallmarks of an excelling cognitive behavioral therapist could very well be the ability to use therapeutic creativity and adaptability, yet within a framework consisting of reasonable levels of competency, treatment fidelity, and adherence to general cognitive behavioral principles. This notion is consistent with previous research regarding the association between flexibility and the use of a manual (Kendall et al., 1998). More recent findings also suggest that therapist flexibility is associated with better treatment outcomes (Owen & Hilsenroth, 2014). The strong importance of "manual" as a moderator for ESs becomes apparent when observing that the only moderator identified as exerting a significant influence on ES for GCBT, was whether a set manual was utilized, or not.

Complex relationship between factors

The outcomes from the different analyses, as well as the theoretical aspects discussed in the previous sections, point towards an identification of two robust factors associated with a rise in effects sizes with time: the format of group therapy, and the flexibility afforded by not utilizing a set treatment manual – one which could be outdated, or conducted with too much rigidity. The first factor is most likely a reflection of the general development of society, with less shame and fear of psychological issues, combined with a higher degree of openness. The identification of two variables facilitating positive developments of ESs with time is supported by the empirical and theoretical analyses, and could explain why individual therapy, largely consisting of two factors with a negative impact on the temporal development

of ES (widespread use of a set manual, in an individual format), experience a fall in ESs. Further, MCBT, which is largely characterized by consisting of one detrimental, and one facilitating factor for temporal development of ESs (widespread use of a set manual, in a group format), shows a standstill with time. While GCBT, consisting of two facilitating factors for the increase of ESs with time (flexible/varied use of manuals, in a group format) shows an improvement in treatment effects.

The results thus suggest that a substantial reason for the observed decline in ES over time for individual CBT could be related to the use of a manual/procedure from the 1970s for a population from the 2000s and onwards. The variables "time" and "manual" seem to have a complex effect on each other, where the importance of updating the latter (and/or being flexible as a therapist), becomes more evident with the passage of time. The standardization of treatment likely results in less variability in treatment effects across time. On the other hand, when varying the approach, ESs increase. The findings building up this thesis strongly lend support to the importance of updating treatment procedures and manuals to ensure they are in touch with general developments in society.

Support for the above hypothesis is also available through investigating geographical differences. Effect sizes from studies performed in the U.S. decreased with time for individual CBT, while ESs for trials from the rest of the world remained unchanged (Christea et al, 2017). Although not specifically reported, it is reasonable to assume that trials performed in the U.S. utilize the original Beck manual to a higher extent, thus contributing to the observed fall in treatment effects, as the original manual gets more outdated for every passing year.

This hypothesized cultural and social "goodness of fit" for psychotherapy is a concept that is attracting increased interest from both researchers and commentators (e.g., Burkeman, 2015). As cultural shifts and developments have occurred at a rapid pace during the last decades, updates for treatment manuals should happen at a similar rate. Such a procedure

could also facilitate another contributor to treatment effect, the placebo effect. Hypothesized as a possible reason for the decline in individual CBT (Johnsen & Friborg, 2015), perhaps the loss of a placebo effect with time could be alleviated if manuals were updated and adjusted at timely intervals.

The relationship between the factors "time" and "adherence to a treatment manual" could also be part of a complex explanation for a decline in treatment effects. The original treatment manual by Beck et al (1979) stresses that high degrees of adaptation and flexibility should be afforded according to the individual needs of the patient, if the adaptations do not deviate from the cognitive principles, and the implementation of the techniques follow the descriptions in the manual. It is conceivable that the therapists in earlier treatment trials were more aware, or attentive, to the importance of adaptation or flexibility. The first treatment trials for CBT as an anti-depressive treatment were led by close associates to Dr. Beck, and sometimes led or supervised by Beck himself. As such, the probability of following the general principles of the manual was probably at the highest level. With time however, it is possible that the more abstract concepts of flexibility and adaptation in the original manual have become understated and replaced with a firmer focus on how to properly implement treatment techniques, such as identifying and challenging automatic thoughts, measuring and addressing maladaptive beliefs, and provide precise psychoeducation. Such a phenomenon would be reflected also in measures of adherence, which are more likely to emphasize, and put value on, the specific ingredients of treatment. Thus, a stronger focus on adherence to the treatment manual may exert a detrimental effect for the crucial elements of adaptation and flexibility. The result could be a more rigid treatment, with less room for individual variations when it comes to, for example, how many sessions should be devoted to the different parts of CBT, or how much time should be dedicated to building the alliance, or to which extent emotional release should be prioritized or supported. Researchers and clinicians of future

clinical trials should thus be aware of this potential pitfall, where stricter demands of adherence may imply moving away from highly important general principles of CBT.

Another factor that will have a significant effect on the development of therapy in years to come, is the COVID-19 situation. Although the net impact is hard to predict at this stage of the pandemic disease, some effects are probable to occur. For example, it is likely that the development in the world would lead to a more favorable view in the population towards internet-delivered treatment programs. This shift would translate into friendlier attitudes, and a higher belief and assuredness towards the internet-format, thus having a beneficial impact on the placebo effect. On the other hand, face-to-face meetings could suffer from the nocebo-effects stemming from peoples fear and reluctance to enter such meetings, conceivably exercising a detrimental effect on treatment effects. Such a scenario, likely to negatively affect group-therapy the most, is a potential relevant and potent example of the complex interaction of factors determining treatment effects; as society changes, so does the requirements for therapy. What constitutes the most beneficial form of therapy today, is not necessarily the same as 2-3 years ago – let alone 40 years ago! Thus, the present pandemic, as horrible as it may be, could also serve to highlight the main message within this thesis: To ensure improvement in our treatment efforts, it is essential to continuously update treatment procedures and manuals. Otherwise, we run the risk of not being aligned with society and the people we are treating.

As previously introduced, there could be a complex relationship regarding the amount of available information, the format of therapy, and the effect on treatment outcomes. For GCBT, information and knowledge of the treatment is probably not as readily accessible and common as for the individual format of CBT. This is partly since there are many variations for conducting GCBT, and because researchers and clinicians frequently develop their own manuals/procedures for trials and general treatment. By doing so, they are essentially

adhering to the theory of being in touch with societal developments. The procedures are being developed at the same time and place as the intervention is utilized. As such, information about the intervention would not become common knowledge before therapy starts, thereby preventing any limiting effect on either the placebo effect, or because of clients making use of the interventions beforehand. Further, the interventions and procedures utilized in GCBT are inherently much harder to practice before entering the therapy-room. After all, it requires a surrounding of a group consisting of similar clients. Although offering a potential explanation to the observed developments in ESs for CBT over time, there could be an argument made towards the observed associations merely revealing a shift in treatment outcome in general. If the patients do gain a substantial amount of helpful knowledge before entering therapy, could this quantify as a "real life" effect of treatment, and should it thus be considered in addition to the effects shown in treatment trials, or as a part of it? If so, what would the influence be on the "true" effect size? This is a complex question, where future research is most welcome.

Variables affecting meta-analyses on time-trends

Heterogeneity

Heterogeneity often poses a concern with the meta-analytic approach. This especially holds true for meta-analyses based on pre-post within-group ESs, where an inherent risk of bias in the form of data dependence, also known as type-1 error, exists (Cuijpers et al., 2017). This risk would also apply for several of the analytic conditions building up the present thesis. However, data dependence is not an issue with regards to between-group calculated ESs. The present articles lend themselves to both procedures, calculating effect sizes through both between-group and within-group approaches, with the results revealing similar outcomes. Overall, this suggests that data dependence do not have an undue effect on the results. Furthermore, even if an issue of data-dependence did exist, this would mainly affect the

overall estimated ES of treatment, and not the targeted time-trends. The underlying assumption is that heterogeneity for within-groups do not systematically change with the passing of time. Rather, the distribution of heterogeneity is considered to randomly vary with the passing of time. Thus, there is little reason to assume that the time-trend results, which are the focus of this study, would be differentially affected by the issue of type-1 error.

In general, high scores of heterogeneity is not an unusual phenomenon for published meta-analyses in the field of social and medical sciences, where about a quarter have *I2* scores above 50% (Higgins et al., 2003). As for psychology specifically, a recent study examining rates of *I2* in published meta-analyses in Psychological Bulletin between 1990 and 2013 (van Erp et al, 2017) revealed that over half of the between-study meta-analyses showed $I2 > 70$. Furthermore, it is important to emphasize that quantification of heterogeneity is only one component of a wider investigation of variability across studies, the most important being diversity in clinical and methodological aspects. This especially holds true for analyses based on small samples. Thus, one should be careful before assuming statistically significant findings as random or spurious purely on the grounds of large *I2* scores. There should be other clear indications present before disqualifying findings.

Methodical and statistical procedures

There are several methodological and statistical choices to be made when performing a meta-analysis. The aim is to utilize the most precise procedure for measuring the phenomenon in question, in a pragmatic and prudent manner.

Typically, current meta-analyses estimating effect sizes for psychotherapy utilize an intention to treat (ITT) format. With such a procedure, the estimated ESs are thought to be similar to what we see in real life therapy, where people not completing the course of therapy are also part of the equation. However, when performing meta-analyses on temporal

development, the ITT procedure carries a larger risk of unexplained variation with the passing of time. Even if the underlying assumption is that the distribution of dropouts and non-completers is randomly distributed over the years, there is still the possibility that some trials, performed in particular years, have a different rate of non-completers as compared to other specific studies performed in other specific timeframes. Thus, for two of the meta-analyses building up this thesis, the active choice to calculate ESs based on completers of a trial were made. With such a procedure we can be certain that the measured effect sizes are based on patients completing therapy, regardless of the year being 1979 or 2018, leaving less room for unwanted variability to exist.

As treatment for depression is an important field of psychology, much research has historically been performed in this area. However, a significant chunk of manuscripts never gets to enjoy life outside the drawer of publishing journals, leading to the issue of publication bias. The risk here is related to the higher probability of significant findings (i.e large effect sizes) to achieve publication, as compared to nil-findings (typically with low effect sizes). A larger meta-analysis on the issue of publication bias, found that ESs in some circumstances may be inflated by 40 percent (Cuijpers et al, 2010). Again, such a phenomenon can be very problematic when the main aim is to estimate an overall ES for treatment, as any risk for inflating the true effect-size is unwelcome. However, for meta-analyses on time-trends, this issue carries less threat. It does not really matter how many specific trials have been left in the drawer, as long as there is no systematical distribution of unpublished trials connected to certain periods of time. Fortunately, there are no indications of this being the case. Nevertheless, publication bias has been thoroughly examined in the present meta-analyses and corrected for if indicated. Such a procedure ensures that the results are dependable both from the perspective of estimating an overall effect size, and from the perspective of measuring time-trends.

Diagnostic Criteria

If relevant diagnostic criteria had been subjected to significant changes during the timeframe under investigation, it is reasonable to suspect such movements could have a bearing on the results. Major changes in diagnostic criteria implies that the current clinical understanding of a disorder has shifted. Such a shift would have an impact on several levels for research trials, with selection procedures, measures of the disorder, and the therapy itself being affected. However, within the time ranges for the studies under investigation in the current thesis, there appear to be limited moderating effects resulting from changes in diagnostic criteria. As the implementation of modern clinical diagnostic criteria for depression most likely preceded the trials included in the current meta-analyses, it is not considered a factor deemed plausible to have a contributing effect in explaining the observed time trends. After the implementation of stricter criteria in the mid to late 70`s, the DSM has only been subjected to rather minor diagnostic adjustments, and the clinical understanding of the depressive syndrome has largely remained intact. Also, the observed adjustments in the DSM over the last four decades do not apply for the more cognitive symptoms of guilt, suicidal ideation or thoughts of worthlessness, which have stayed the same from DSM-III to the current DSM-V.

Other perspectives and recent research

The findings of a temporal decline in ES for individual CBT has been challenged by a couple of re-analyses (Christea et al, 2017; Ljótsson et al, 2017). Yet, the field of depression treatment has, by and large, accepted the findings as reflecting a true phenomenon – albeit with a plethora of hypotheses with regards to the underpinnings of the results. For example, a recently published paper tied some of the decline to possible differences in the control groups for newer versus older trials of CBT (Dobson, 2016). Another study linked the decline to a

lack of training, competency, treatment fidelity and/or adherence to cognitive principles by modern cognitive behavioral therapists (Waltman, Creed, & Beck, 2016). This stance is largely supported by Dobson (2016), while also noting that earlier trials to a larger extent were characterized as measuring efficacy, versus effectiveness for newer studies.

The findings from paper I regarding the achievement of higher effect sizes for psychologists vs trained students, have been expanded in subsequent research. A new meta-analysis substantiated that students did indeed achieve poorer ESs from therapy than did trained clinicians (Goldstein et al, 2020).

The focus on change across time, and the development of methods for quantifying and measuring temporal trends of treatment effects, displayed by the meta-analyses in this thesis, have served as a model for further investigations in other areas of psychotherapy. For example, a comprehensive meta-analysis found that the rates of self-injuring thoughts and behaviors have not improved because of specific treatment for the last 5 decades (Fox et al, 2020). This finding could seem counterintuitive, given the number of efforts investigated in understanding and preventing such behaviors. Another meta-analysis found that cognitive behavioral therapy for psychosis had seen an improvement with the passing of time for the symptom of delusions, but no significant change across time for other symptoms (Sitko et al, 2020). Further, an extensive meta-analysis examined whether the effects of blood pressure lowering treatments had improved with time (Sekizawa et al, 2018). No change was observed for the chosen parameters.

The view presented in this thesis, regarding the need of systematical changes in how modern treatment for depression is conceived and implemented, enjoys substantial support by researchers in the field of depression treatment, and psychological treatment in general. A recent paper pointed towards a potential shift in paradigm, where the focus no longer should be on which therapy is the most efficient – but rather towards which kind of therapy is best

suited for the individual patient, in his/her situation in the here and now (Leichsenring et al, 2018;). This point is developed further in a subsequent article by the same authors. Here, a suggested change in funding policies is recommended, moving on from a "more and more of the same" philosophy, towards more variation in which therapies receive funding (Leichesenring et al, 2019). The main message being that plurality is the future for psychotherapy, as one uniform does not fit all clients. The general idea of choosing the best treatment format and ingredients for the individual patient is also discussed and supported in other recent articles on the topic (McCormack & Chadler, 2018; Mulder et al, 2017). In a similar vein, another recent article investigated the potential beneficial effects of implementing a personalized allocation of patients to therapists (Delgadillo et al, 2020). The hypothesis was that by matching patients with specific characteristics to specific therapists, outcomes would improve. The results suggest this was indeed the case.

A combined view encompassing both the need for plurality in psychotherapy, while also taking into consideration the contemporary developments of the world, leads to thoughts regarding therapy delivered via mobile phone applications and internet interventions. For the latter, a recent meta-analysis indicated that tailoring could be effective also for anti-depressive therapy delivered via the internet (Twomey et al, 2020). When it comes to therapy delivered via mobile phone applications, a recent study showed there were considerable positive effects on a range of mental health issues when using a CBT-based intervention (Rathbone et al, 2017). These findings should be considered as specially promising given the state of the world as of the year 2021, with the coronavirus still being highly active.

Limitations and strengths

The meta-analytic papers included in this thesis are exploratory in nature, with few hypotheses posed beforehand. This may reduce the certainty with which one can draw

conclusions. Further, reliable measures of therapist adherence or treatment fidelity were almost nonexistent for the included studies, making it impossible to perform highly informative analyses or discussions with that regard. On a more detailed level, the self-report inventory BDI comes in two forms, the BDI-I and the BDI-II. For some analyses, these were treated as interchangeable. Although the two versions are very similar, there is still a possibility of a confounding effect resulting from this mixture. Additionally, as the BDI and the HDRS do not measure improvement on a more general or global level, one cannot reliably generalize the effects to areas beyond symptom relief. Further, none of the included meta-analyses cover follow-up scores, which means that such results and information, with potential high clinical value, are not available. For papers I and III, this choice was made based on large variations in which individual trials included follow-up data, with especially several older trials lacking. Finally, the meta-analyses building up the current thesis do not check or control for researcher allegiances. However, the main problems associated with allegiance are not as pronounced in a temporal design, as in a comparative design. When comparing different forms of treatment, a potential inflation of ESs may obviously exert a large effect on the conclusion. For the current temporal analyses though, the underlying assumption is that any research allegiance will vary in a random manner with time, and thus most likely will not exert any systematical or confounding effects on the results. As far as the author is aware, there are no literature or research indicating that researcher allegiance was more pronounced in seminal or modern trials.

The methodical and statistical allegiance and rigidity are considered major strengths for this thesis. The included articles perform meta-analyses with a highly similar template, allowing comparisons across studies. Further, the current concept of reliably defining and measuring time-trends represents an innovation for the field of psychotherapy. This thesis also encompasses and investigates the major branches of CBT (individual, group, and

mindfulness-based CBT), thus representing a broad evaluation of the cognitive-based treatment of depression. A final aspect worth mentioning as a strength relates to the affiliation of the main author. The doctoral candidate does not have any links, ties, interests, or connections to any specific branches of therapy. The present thesis should thus be considered to meet the highest standards of objectivity.

Implications and future research

The background and main purpose for the present thesis and its articles was to investigate a previously largely unknown factor; the temporal development of CBT-related treatment for depression. The focus and findings discussed in this thesis are important and pave the way for future research in psychology in general. Efforts to investigate temporal developments are spreading throughout the field and will gain us further knowledge and insight for years to come, driving the profession forward. As a contributing factor to better understand the relationship between therapeutic flexibility/rigidity and treatment outcome, future trials in the field of depression treatment would benefit from implementing a common scale for measuring adherence to the treatment manual. In addition, such a scale should proviode checks for both adherence to the general principles of the treatment manual, as well as to the implementation of specific techniques.

From a clinical point of view, future research investigating potential temporal treatment effects at follow up intervals is important. The current meta-analyses do not give any precise indications as to whether the presented results and trends are similar 1 or 2 years after completion of treatment.

As for the specific findings related to CBT as an anti-depressive treatment, this thesis has provided strong evidence towards the importance of continuously updating, or adapting, treatment procedures and treatment manuals. Further research should thus emphasize efforts

in this direction. Here lies an important challenge for CBT specifically, and for clinical therapy in general. Further, the development of methods for identifying which patients (and individual characteristics) are suited for the different forms of therapies, are hugely welcome.

The results from the present articles provide a potential start on such a journey, by indicating certain characteristics that could help differentiate which clients would profit most from CBT, and the formats of individual or group therapy. For example, clients who are highly enlightened and updated on psychology and cognitive principles in advance of starting therapy, are probably more likely to not get the full beneficial effects of a standard cognitive behavioral therapy as it is conducted today. Many of the steps in the manual will be well-known and practiced by the clients beforehand, thus diminishing initial response. Such patients could be allocated to other forms of (less known) therapies. Further, higher levels of open-mindedness most likely indicate a more beneficial outcome from the less standardized, more flexible, and more social format of group cognitive therapy, while the more conscientious and closed personality type probably would get better gains from the more structured and organized individual CBT.

The general importance of being adaptive in society is probably at its most pronounced in the year of 2021. The spread of COVID-19 has had a huge impact on people's lives, and as such also for the way therapy is delivered. Many regions and countries have seen an increase in digitally delivered therapy. Therefore, research efforts for the immediate future should be directed towards finding the best way of adapting CBT to the highly relevant formats of mobile applications and the internet. Such a focus could also have the additional benefit of making help via anti-depressive therapy available for even more struggling people. The modern society consists of a significant amount of (more or less) data-bound individuals. This group's opportunity to get into a treatment program would be immensely enhanced by further developments on the digital and technological front.

Conclusions

Cognitive behavioral therapy is effective as an anti-depressive treatment. This applies both for the more recent adaptation in the form of mindfulness-based CBT, and even more so for the traditional individual and group CBT. However, the analyses showed vast differences in time-related effects for the three formats of CBT: The treatment effects for individual therapy are decreasing with the passage of time, GCBT has an increase in treatment effect, while MCBT show no tendencies in either direction. The main explanation for the different outcomes is probably due to two factors: the application of a set (and perhaps outdated) treatment manual, and the format of therapy (group vs individual). Individual therapy and the utilization of a set manual seem to be factors associated with a fall in treatment effect with time, while group therapy, and a more flexible approach to the treatment program are factors facilitating an increase in ES. Although the factors "group format" and "no set manual" may facilitate an improvement in ES with time, one cannot draw the conclusion that group therapy is (or will become) the most efficient modality of treatment for depression. Nor is it possible to state that a treatment without any form of manual would be better.

However, the results strongly indicate that a new approach is needed for traditional CBT to maintain its standing and treatment effects: Moving towards a more adaptive and flexible implementation and mindset would ensure even more clients recover from depression through the aid of CBT. It is thus considered to be of pivotal importance to frequently and continuously update and adapt treatment manuals to ensure they meet the dynamic and ever-changing requirements of modern society. What was considered the most efficient treatment 10-15 years ago, is most likely lagging today. Further, the beneficial effects provided by adopting a more adaptive and flexible approach towards therapy also extends to therapy forms and formats. Having a variety of different forms and formats for anti-depressive treatment would be beneficial for the patients and provide psychological treatment clinics with the tools

to help even more clients. Plurality, flexibility, and adaptation are thus keywords in our efforts to help more clients recover from depressive disorders. Perhaps we soon could witness the development of a more efficient system, with effective screening tools for accurately distributing the individual patients to the therapy form and the therapist which suits their needs the most.

References

Abbass, A. A., Rabung, S., Leichsenring, F., Refseth, J. S., & Midgley, N. (2013). Psychodynamic Psychotherapy for Children and Adults: A Meta-Analysis of Short-Term Psychodynamic Models. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 1241-45.

Adler, A. (1927). *Understanding human nature*. New York: Greenburg.

Ambrosini, P. J., Metz, C., Bianchi, M. D., Rabinovich, H., & Undie, A. 1991. Concurrent Validity and Psychometric Properties of the Beck Depression Inventory in Outpatient Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*, 51-57.

American Psychiatric Association. (1952). *Diagnostic and statistical manual: mental disorders*. Washington, DC.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. revised). Washington, DC.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC.

Andrews, G. (2001). Should depression be managed as a chronic disease? *British Medical Journal, 322*, 419–421. doi: 10.1136/bmj.322.7283.419

Ashouri, A., Vahid, M. K. A., Gharaee, B., & Rasoulian, M. (2013). Effectiveness of Metacognitive and cognitive-behavioral therapy in patients with major depressive disorder. *Iranian Journal of Psychiatry and Behavioral Sciences, 7*(2), 27-34.

A-Tjak, J. G. L., Morina, N., Topper, M., et al (2021). One year follow-up and mediation in cognitive behavioral therapy and acceptance and commitment therapy for adult depression. *BMC Psychiatry 21,* 41. doi: 10.1186/s12888-020-03020-1

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., . . Wampold, B. E. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review, 33*(3), 395-405.

Baer, R. A., & Walsh, E. (2016). Treating acute depression with mindfulness-based cognitive therapy. In A. Wells & P. L. Fisher (Eds.), *Treating depression* (pp. 344-368). New York: Wiley.

Beck, A. T. (1970). Cognitive therapy: Nature and relation to behavior therapy. *Behavior Therapy, 1*, 184–200.

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

Beck, A. T., Steer, R., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of beck depression inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment, 67*, 588-597.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II, Beck depression Inventory: Manual* San Antonio, TX: Psychological Corporation.

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

Beck Institute. (2021). History of Cognitive Behavior Therapy. Retrieved 24/07-2021. *https://beckinstitute.org/about-beck/history-of-cognitive-therapy/*

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine, 14*(4), 395–411.

Bienenfeld, D., 2007. Cognitive Therapy of patients with personality disorders. *Psychiatric Annals, 37*, 133-139.

Blanco, C., Okuda, M., Markowitz, J. C., Liu, S.-M., Grant, B. F., & Hasin, D. S. (2010). The epidemiology of chronic major depressive disorder and dysthymic disorder: Results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry, 71*, 1645–1656. doi: 10.4088/JCP.09m05663gry

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis (Version 2) [Computer software]*. Englewood, NJ: Biostat

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis.* New York: Wiley.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2017). *Comprehensive Meta-Analysis (version 3)*. Englewood: Biostat.

Burkeman; The Guardian (2015). Why CBT is falling out of favor. Downloaded 03/04.2016. *https://www.theguardian.com/lifeandstyle/2015/jul/03/why-cbt-is-falling-out-of-favouroliver-burkeman*

Burkhart, B. R., Rogers, K., McDonald, W. D., McGrath, R., & Arnoscht, O. 1984. The measurement of depression: enhancing the predictive validity of the Beck Depression Inventory. *Journal of clinical psychology, 40*, 1368-72.

Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., & Strauss, B. (2016). Outcome differences between individual and group formats when identical and nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy, 53*, 446-461. doi: 10.1037/pst0000090

Button, K., Kounali, D., Thomas, L., Wiles, N., Peters, T., Welton, N., . . . Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory - II according to the patient's perspective. *Psychological Medicine, 45*(15), 3269-3279. doi:10.1017/S0033291715001270

Castonguay, L. G. (1996). *Integrative cognitive therapy for depression treatment manual*. Unpublished manuscript, The Pennsylvania State University.

Cirstoiu, C., Circota, G., Panaitescu, C., & Niculaita, R. (2011). The advantage of arthroscopic anterior cruciate ligament reconstruction with autograft from tendons of the Semitendinosus - gracilis muscles for the recovery of the stability of the knee. *Maedica, 6*(2), 109-113.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology, 65*(3), 145–153. https://doi.org/10.1037/h0045186

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Canadian Psychology/Psychologie canadienne, 58*, 7-19. doi: 10.1037/cap0000096

Cuijpers, P., Dekker, J., Hollon, S. D., & Andersson, G. (2009). Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: A meta-analysis. *The Journal of Clinical Psychiatry, 70* (9), 1219–1229.

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010) Efficacy of cognitive behavioral therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry, 196 (3),* 173–8.

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.

Crane, R. (2009). *Mindfulness-based cognitive therapy: Distinctive features*: Routledge.

Cristea, I. A., Stefan, S., Karyotaki, E., David, D., Hollon, S. D., & Cuijpers, P. (2017). The effects of cognitive behavioral therapy are not systematically falling: A revision of Johnsen & Friborg (2015). *Psychological Bulletin, 143,* 326-340. . doi: 10.1037/bul0000062

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K.,…. Ztrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91.

Davanloo, H. (2000). *Intensive Short-Term Dynamic Psychotherapy.* New York: Wiley.

Delgadillo, J., Rubel, J., & Barkham, M. (2020). Towards Personalized Allocation of Patients to Therapists. *Journal of Consulting and Clinical Psychology* (advanced online publication). doi: 10.1037/ccp0000507.

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

Dobson, K. S. (2016). The efficacy of cognitive-behavioral therapy for depression: Reflections on a critical discussion. *Clinical Psychology: Science and Practice, 23*, 123-125. doi: 10.1111/cpsp.12151

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,* 295-300.

Dobson, K. S., Hollon, S. D., Dimidjian, S., et al. (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *J Consult Clin Psychol, 76*, 3, 468–477.

Dragioti, E., Dimoliatis, I., & Evangelou, E. (2015). Disclosure of researcher allegiance in meta-analyses and randomised controlled trials of psychotherapy: a systematic appraisal. *BMJ Open, 5,* 6. doi: 10.1136/bmjopen-2014-007206

Duncan, B. L., & Miller, S. D. (2006). Treatment manuals do not improve outcomes. In L. E. B. J. C. Norcross & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC: American Psychological Association.

Duncan, B. L., Miller, S. D., & Sparks, J. (2004). *The Heroic Client. A revolutionary way to improve effectiveness through client-directed, outcome informed Therapy*. San Francisco: Jossey-Bass.

Ellis A. (1957). Rational psychotherapy and individual psychology. *J. Individ Psychol.* 13 38–44.

Ellis, A. (1962). *Reason and emotion in psychotherapy*. New York: Lyle Stuart.

Erikson, E. H. (1950). *Childhood and society*. New York: Norton.

Everyday Health (2013). Stats and Facts About depression in America. Downloaded 20/01-2014. http://www.everydayhealth.com/health-report/majordepression/depressionstatistics.aspx

Eysenck, H. J. (1960). *Behavior Therapy and the Neuroses*. Pergamon; Oxford

Fernandez, E., Salem, D., Swift., & Ramtahal, N. (2015) Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *J Consult Clin Psychol. 83*, 6, 1108-22. doi: 10.1037/ccp0000044.

Field, A. P (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics, 2*(2), 105-124.

Fox, K. R., Huang, X., Guzmán, E. M., Funsch, K. M., Cha, C. B., Ribeiro, J. D., & Franklin, J. C. (2020). Interventions for Suicide and Self-Injury: A Meta-Analysis of Randomized Controlled Trials Across Nearly 50 Years of Research. *Psychological Bulletin*. Advance online publication. http://dx.doi.org/10.1037/bul000030

Freud, S. (1917). Mourning and melancholia. *Standard edition, 14(19)*, 17.

Friborg, O., Martinsen, E. W., Martinussen, M., Kaiser, S., Øvergård, K. T., & Rosenvinge, J. H. (2014). Comorbidity of personality disorders in mood disorders: A meta-analytic review of 122 studies from 1988 to 2010. *Journal of Affective Disorders, 152*, 1-11. doi:10.1016/j.jad.2013.08.023

Galante, J., Iribarren, S. J., & Pearce, P. F. (2013). Effects of mindfulness-based cognitive therapy on mental disorders: A systematic review and meta-analysis of randomized controlled trials. *Journal of Research in Nursing, 18*, 133-155. doi: 10.1177/1744987112466087

Garfield, S. L. (1986). Research on client variables in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 213–256). New York: Wiley.

Goldberg, S. B., Tucker, R. P., Greene, P. A., Davidson, R. J., Kearney, D. J., & Simpson, T. L. (2019). Mindfulness-based cognitive therapy for the treatment of current depressive symptoms: A meta-analysis. *Cognitive Behavior Therapy, 48*, 445-462. doi:10.1080/16506073.2018.1556330

Goldberg, S. B., Tucker, R. P., Greene, P. A., Davidson, R. J., Wampold, B. E., Kearney, D. J., & Simpson, T. L. (2018). Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. *Clinical Psychology Review, 59*, 52-60. doi: 10.1016/j.cpr.2017.10.011

Goldstein, L., Adler Mandel, A. D., DeRubeis, R. J., & Strunk, D. E. (2020). Outcomes, skill acquisition, and the alliance: Similarities and differences between clinical trial and student therapists. *Behaviour Research and Therapy 129*:103608 DOI: 10.1016/j.brat.2020.103608

Gould, R. L., Coulson, M. C., & Howard, R. J. (2012). Cognitive behavioral therapy for depression in older people: A meta-analysis and meta-regression of randomized controlled trials. *Journal of the American Geriatrics Society, 60*(10), 1817-1830.

Goyal, M., Singh, S., Sibinga, E. S., & et al. (2014). Meditation programs for psychological Stress and well-being: A systematic review and meta-analysis. *JAMA Internal Medicine, 174*, 357-368. doi: 10.1001/jamainternmed.2013.13018

Gu, J., Strauss, C., Bond, R., & Cavanagh, K. (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. *Clinical Psychology Review, 37*, 1-12. doi: 10.1016/j.cpr.2015.01.006

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23,* 351-379. doi: 10.1136/jnnp.23.1.56

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81*, 75-88. doi: 10.1037/a0031080

Healthline. (2012). Unhappiness by the Numbers: 2012 Depression Statistics. downloaded 01-10/2013. *http://www.healthline.com/health/depression/statistics-infographic*

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Hedlund, J. L., & Vieweg, B. W. (1979). The Hamilton rating scale for depression: a comprehensive review. *Journal of Operat Psychiatry, 10*, 149-165.

Hedman-Lagerlöf, M., Hedman-Lagerlöf, E., & Öst, L.-G. (2018). The empirical support for mindfulness-based interventions for common psychiatric disorders: A systematic review and meta-analysis. *Psychological Medicine, 48,* 2116-2129. doi: 10.1017/S0033291718000259

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560. doi: 10.1136/bmj.327.7414.557

Hofmann, S.G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T.,& Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive Therapy and Research, 36*, 5, 427-440.

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-Based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78, 169-183*. doi: 10.1037/a0018555

Hollon, S. D., Shelton, R. C., & Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology, 59*(1), 88-99.

Hollon, S. D., DeRubeis, R. J., Shelton, R. C., et al. (2005). Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Arch Gen Psychiatry, 62*, 4, 417–422.

Honyashiki, M., Furukawa, T. A., Noma, H., Tanaka, S., Chen, P., Ichikawa, K. . . . Caldwell, D. M. (2014). Specificity of CBT for depression: A contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research, 38*(3), 249-260.

Horvath, A. O., & Greenberg, L. S. (1986). The development of the working alliance inventory. In L. S. Greenberg & W. M. Pinsoff (Eds.). *The Psychotherapeutic Process: A research handbook* (pp. 529-556). New York: Guilford Press.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986).The dose–effect relationship in psychotherapy. *American Psychologist*, *41*, 159–164.

Huberty, C. J. (2002). A History of Effect Size Indices. *Educational and Psychological Measurement*, *62*(2), 227–240. https://doi.org/10.1177/0013164402062002002

Huntley, A. L., Araya, R., & Salisbury, C. (2012). Group psychological therapies for depression in the community: Systematic review and meta-analysis. *British Journal of Psychiatry, 200*, 184–90. doi: 10.1192/bjp.bp.111.092049

Huppert, J. D., Bufka, L. F., Barlow, D. H.,Gorman, J. M., Shear, M. K., & Woods, S.W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*, 747-755.

Imel, Z. E., & Wampold, B. E. (2008). The importance of treatment and the science of common factors in psychotherapy. In S. D. Brown & R. W. Lent (Eds), *Handbook of counseling psychology* (pp. 249-262): John Wiley & Sons Inc.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials, 17*, 1-12. doi: 10.1016/0197-2456(95)00134-4

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*, 747-768. doi: 10.1037/bul0000015

Johnsen, T. J., & Thimm, J. C. (2018). A meta-analysis of group cognitive-behavioral therapy as an anti-depressive treatment: Are we getting better? *Canadian Psychology/Psychologie canadienne, 59*, 15-30. doi: 10.1037/cap0000132

Joutsenniemi, K., Laaksonen, M. A., Knekt, P., Haaramo, P., & Lindfors, O. (2012). Prediction of the outcome of short- and long-term psychotherapy based on socio-demographic factors. *Journal of Affective Disorders, 141,* 331–342. http://dx.doi.org/10.1016/j.jad.2012.03.027

Jung, C. G., Von Franz, M. L., Henderson, J. L., Jaffé, A., & Jacobi, J. (1964). *Man and his Symbols*, New York, N.Y.: Anchor Books, Doubleday.

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York: Dell Publishing

Kegel, A. F., & Fluckiger, C. (2014). Predicting Psychotherapy Dropouts: A Multilevel Approach. *Clinical Psychology & Psychotherapy 22(5). DOI: 10.1002/cpp.1899*

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11*, 289-299.

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.

Khoshbooii, R. (2012). Comparison group and individual cognitive behavioral therapy in treatment of depression among Iranian women around menopause. *International Journal of Psychological Studies, 4*, 175-181. doi: 10.5539/ijps.v4n1p174

Knapik, M. L. (2012). Transradial arterial access for cath and PCI, and the impact on hospital bottom lines. *Cath Lab Digest, 20*(9), 44-48.

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P, Barber, J., Thase, M. E. . . . Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprensive Psychiatry, 51*, 319–324.

Kohler, S., Unger, T., Hoffmann, S., Steinacher, B., Fydrich, T. (2013). Acute and long-term treatment outcome in depressed inpatients with vs. without anxious features: Results of a one-year follow-up study. *Journal of Affective Disorders, 150*(3), 1055-1061.

Kovacs, M., Rush, A. J., Beck, A. T., & Hollon, S. D. (1981). Depressed outpatients treated with cognitive therapy or pharmacotherapy. A one-year follow-up. *Arch Gen Psychiatry, 38,* 1, 33-39. doi: 10.1001/archpsyc.1981.01780260035003.

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE, 9*, e105825. doi: 10.1371/journal.pone.0105825

Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. (2002). *Meta-Analysis*: *A Guide to Calibrating and Combining Statistical Evidence.* Hoboken, NJ: Wiley.

Kuyken, W., Warren, F. C., Taylor, R. S., Whalley, B., Crane, C., Bondolfi, G. . . . Schweizer, S. (2016). Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: An individual patient data meta-analysis from randomized trials. *JAMA Psychiatry, 73*, 565-574. doi: 10.1001/jamapsychiatry.2016.0076

Lambert, M. J. (1992). Psychotherapy outcome research. In: Norcross, J. C., & Goldfried, M. R. (ed.). *Handbook of Psychotherapy Integration.* New York: Basic.

Lambert, M. J. (2001).The status of empirically supported therapies: Comment on Westen and Morrison's (2001) multidimensional meta-analysis. *Journal of Consulting and Clinical Psychology*, *69*, 910–913.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 143-198). New York: Wiley.

Lau, M. (2016). The theory underlying mindfulness-based cognitive therapy as a relapse prevention approach to depression. In A. Wells & P. L. Fisher (Eds.), *Treating depression* (pp. 194-200). New York: Wiley.

Lau, M. A., Segal, Z. V., & Williams, J. M. G. (2004). Teasdale's differential activation hypothesis: Implications for mechanisms of depressive relapse and suicidal behaviour. *Behaviour Research and Therapy, 42*, 1001-1017. doi: 10.1016/j.brat.2004.03.003

Leichsenring, F., Abbass, A., Hilsenroth, M. J., Luyten, P., Munder, T., Rabung, S., & Steinert, C. (2018). "Gold standards,"plurality and monocultures: The need for diversity in psychotherapy. *Frontiers in Psychiatry, 9*, 159.

Leichsenring, F., Steinert, C., & Ioannidis, J. (2019). Toward a paradigm shift in treatment and research of mental disorders. *Psychological Medicine, 49* (13), 2111-2117. doi:10.1017/S0033291719002265

Lejuez, C. W., Hopko, D. R., & Hopko, S. D. (2001). A brief behavioral activation treatment for depression. Treatment manual. *Behav Modif 25(2)*, 255-86. doi: 10.1177/0145445501252005

Lenz, A. S., Hall, J., & Bailey Smith, L. (2016). Meta-analysis of group mindfulness-based cognitive therapy for decreasing symptoms of acute depression. *Journal for Specialists in Group Work, 41*, 44-70. doi: 10.1080/01933922.2015.1111488

Lewinsohn, P. M., & Libet, J. (1972). Pleasant Events, Activity Schedules, and Depressions. *Journal of Abnormal Psychology, 79(3*), 291–5.

Ljótsson, B., Hedman, E., Mattsson, S., & Andersson, E. (2017). The effects of cognitive behavioral therapy for depression are not falling: A re-analysis of Johnsen & Friborg (2015). *Psychological Bulletin, 143,* 321-325. doi: 10.1037/bul0000055

Luborsky, L., Mark, D., Hole, A., Popp, C., Goldsmith, B., & Cacciola, J. (1995) Supportive-expressive dynamic psychotherapy of depression: A time-limited version. *In J. Barber & P. Crits-Christoph (Eds). Dynamic therapies for the psychiatric disorders: Axis I.* NY: Basic Books.

Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach,A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42,* 602-611.

Manicavasgar, V., Parker, G., & Perich, T. (2011). Mindfulness-based cognitive therapy vs cognitive behaviour therapy as a treatment for non-melancholic depression. *Journal of Affective Disorder*s, *130*(1-2), 138-44.

McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment, 90*, 105-109. doi: 10.1080/00223890701845104

McCormack, D. & Chadler, T. (2018). Horses for courses: choosing an evidence-based psychological therapy for your patient. *BJPsych Advances, 24,* (06):1-10

McDermut, W., Miller, I. W., & Brown, R. A. (2001). The efficacy of group psychotherapy For depression: A meta-analysis and review of the empirical research. *Clinical Psychology-Science and Practice, 8*, 98-116. doi: 10.1093/clipsy/8.1.98

McGlothlin, AE, Lewis, RJ (2014). Minimal clinically important difference: defining what really matters to patients. *Journal of the American Medical Association* 312, 1342–1343.

McGuire, J. F., Piacentini, J., Brennan, E. A., Lewin, A. B., Murphy, T. K., Small, B. J., & Storch, E. A. (2014). A meta-analysis of behavior therapy for tourette syndrome. *Journal of Psychiatric Research. 50*, 106-112.

Mental Health Foundation. (2010). Economic burden of mental illness cannot be tackled without research investment. Downloaded 21/11-2015. *http://www.mentalhealth.org.uk/content/assets/PDF/campaigns/MHF-Business-casefor-MH-research-Nov2010.pdf*

Montgomery, S.A., & Aasberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry, 134*, 382-389. http://dx.doi.org/10.1192/bjp.134.4.382

Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: opening the black box. *Lancet Psychiatry, 4* (12), 953-62.

National Collaborating Centre for Mental Health (2004). *Depression: Management of Depression in Primary and Secondary Care*. British Psychological Society and Royal College of Psychiatrists: Leicester and London

Newton-Howes, G., Tyrer, P., & Johnson,T. (2006). Personality disorder and the outcome of depression: Meta-analysis of published studies. *The British Journal of Psychiatry, 188*(1), 13-20.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62*(9), 1157-1172.

Okumura, Y., & Ichikura, K. (2014). Efficacy and acceptability of group cognitive behavioral therapy for depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 164*, 155-164. doi: 10.1016/j.jad.2014.04.023

Owen, J., & Hilsenroth, M. J. (2014). Treatment adherence: The importance of therapist flexibility in relation to therapy outcomes. *Journal of Counseling Psychology*, *61*, 280-288. doi: 10.1037/a0035753

Pallesen, S., Mitsem, M., Kvale, G., Johnsen, B-H., & Molde, H. (2005). Outcome of psychological treatments of pathological gambling: A review and meta-analysis. *Addiction, 100*(10), 1412-1422.

Patton, S. B. (2006). A major depression prognosis calculator based on episode duration. *Clinical Practice and Epidemiology in Mental Health,*2:13. doi: 10.1186/1745-0179-2-13.

Pearson, K. (1905). *Mathematical contributions to the theory of evolution, XIV: On the general theory of skew correlation and non-linear regression*. London: Dulau.

Penninx, B. W., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P. … Beekman, A. T. (2011). Two-year course of depressive and anxiety disorders, results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 133*(1-2), 76-85.

Piet, J., & Hougaard, E. (2011). The effect of mindfulness-based cognitive therapy for Prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 31*, 1032-1040. doi: 10.1016/j.cpr.2011.05.002

Rathbone, A, L., Clarry, L., & Prescott, J. (2017). Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. *J. Med. Internet Research, 19* (11). https://doi.org/10.2196/jmir.8598(2017)

Rector, N. A., Zuroff, D. C., & Segal, Z. V. (1999). Cognitive change and the therapeutic alliance: The role of technical and nontechnical factors in cognitive therapy. *Psychotherapy: Theory, Research, Practice, Training, 36*(4)*,* 320-328.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010) Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ ,340*, 221-228.

Rosenthal, R. (1993). *Meta-analytic procedures for social research*. Newbury Park: CA: Sage

Roshanaei-Moghaddam, B., Pauly, M. C., Atkins, D. C., Baldwin, S. A., Stein, M. B., Roy-Byrne, P. (2011). Relative effects of CBT and pharmacotherapy in depression versus anxiety: is medication somewhat better for depression, and CBT somewhat better for anxiety? *Depression & Anxiety, 28*(7), 560-567.

Rush, A. J., South, C., Jain, S., et al. (2021) Clinically Significant Changes in the 17- and 6-Item Hamilton Rating Scales for Depression: A STAR*D Report. *Neuropsychiatr Dis Treat. 14,* 17. 2333-2345. doi: 10.2147/NDT.S305331.

Saman, M., Diefenbacher, A., Schade, C., et al (2021). A clinical effectiveness trial comparing ACT and CBT for inpatients with depressive and mixed mental disorders, *Psychotherapy Research, 31*, 3. doi: 10.1080/10503307.2020.1802080

Schaub, A., Roth, E., & Goldmann, U. (2006). *Cognitive psycho-educational therapy for depression. Therapy manual*. Göttingen, Germany: Hogrefe.

Schuele, C. M., & Justice, L. M. (2006). The importance of effect sizes in the interpretation of research. *The ASHA Leader 11,* 14–27. https://doi.org/10.1044/leader.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse.* New York: Guilford Press.

Segal, Z. V., Williams, M., Teasdale, J. D., & Kabat-Zinn, J. (2013). *Mindfulness-based cognitive therapy for depression*. New York: Guilford.

Sekizawa, Y., Konishi, Y., & Kimura, M. (2018). Are the effects of blood pressure lowering treatment diminishing? Meta-regression analyses. *Clinical Hypertension 24* (16).

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*(12), 965-974.

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A. . . . Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy, 47*(11), 902-909.

Shaw, B. F. (1977). Comparison of cognitive therapy and behavior therapy in the treatment of depression. *Journal of Consulting and Clinical Csychology, 45*, 543-551. doi:10.1037/0022-006X.45.4.543

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K.,…Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology., 78* (5), 751-756.

Skinner, B. F. (1977) *Science and human behaviour*. Free Press; New York: 1953.

Sitko, K., Bewick, B. M., Owns, D., & Masterson, C. (2020). Meta-analysis and Meta-regression of Cognitive Behavioral Therapy for Psychosis (CBTp) Across Time: The Effectiveness of CBTp has Improved for Delusions. *Schizophrenia Bulletin Open 1*(1).

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760.

Sobocki, P., Jönsson, B. Angst, J., & Rehnberg, C. (2006). Cost of depression in Europe. *Journal of Mental Health Policy and Economics, 9*, 87-98.

Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., … Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry, 157*, 229-233.

Strauss, C., Cavanagh, K., Oliver, A., & Pettman, D. (2014). Mindfulness-based interventions For people diagnosed with a current episode of an anxiety or depressive disorder: A metaanalysis of randomised controlled trials. *PLOS one, 9*, e96110

Tucker, M., & Oei, T. P. S. (2007). Is group more effective than individual cognitive behavior therapy? The evidence is not solid yet. *Behavioural and Cognitive Psychotherapy, 35*, 77-91. doi: 10.1017/S1352465806003134

Twomey, C., O'Reilly, G., Bültmann, O., & Meyer, B. (2020). Effectiveness of a tailored, integrative Internet intervention (deprexis) for depression: Updated meta-analysis. *PLoS One 15*, 1. https://doi.org/10.1371/journal.pone.0228100

van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in Psychological Bulletin from 1990–2013. *Journal of Open Psychology Data*, *5*, 4. doi: 10.5334/jopd.33

Vos, T., Corry, J., Haby, M. M., Carter, R., & Andrews, G. (2005). Cost-effectiveness of CBT and drug interventions for major depression. *Australian and New Zealand Journal of Psychiatry, 39*, 683-692.

Waltman, S. H., Creed, T. A., & Beck, A. T. (2016). Are the effects of cognitive behavior Therapy for depression falling? Review and critique of the evidence. *Clinical Psychology: Science and Practice, 23*, 113-122. doi: 10.1111/cpsp.12152

Waltman, S. H., Sokol, L., & Beck, A. T. (2017). Cognitive Behavior Therapy treatment fidelity in clinical trials: Review of recommendations. *Current Psychiatry Reviews, 13*(4), 311–315. https://doi.org/10.2174/1573400514666180109150208

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*, 914–923.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate*. New York: Routledge

Wampold, B. E., Minami, T., Baskin, T. W., & Callen, S. T. (2002). A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of Affective Disorders, 68*(2-3), 159-165.

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.

Wang, Y. Y., Li, X. H., Zheng, W., Xu, Z. Y., Ng, C. H., Ungvari, G. S., . . . Xiang, Y. T. (2018). Mindfulness-based interventions for major depressive disorder: A comprehensive meta-analysis of randomized controlled trials. *Journal of Affective Disorders, 229*, 429-436. doi: 10.1016/j.jad.2017.12.093

Weissman, M. M., Klerman, G. L. & Prusoff, B. A, et al. (1981). Depressed outpatients: results one year after treatment with drugs and/or interpersonal psychotherapy. *Arch Gen Psychiatry. 1981: 38*; 52–55.

Wells, A. (2000). *Emotional disorders and metacognition: Innovative cognitive therapy.* Chichester, UK: Wiley.

Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. *Advances in Psychiatric Treatment, 16*, 219-227. doi: 10.1192/apt.bp.108.005744

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology, Research and Practice, 24,* 190–195. http://dx.doi.org/10.1037/0735-7028.24.2.190

Wittchen H. U., Jacobi F., Rehm J., Gustavsson A., Svensson M., Jonsson B., et al. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21*, 655-679.

Wolpe, J. (1958). *Psychotherapy by Reciprocal Inhibition*. Stanford University Press; Stanford.

World Health Organization (1955). *International Classification of Diseases: Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death, 7th revision*. Geneva: WHO

World Health Organization (1965). *International Classification of Diseases: Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death, 8th revision.* Geneva: WHO

World Health Organization (1975). *International Classification of Diseases, 9th Revision*. Geneva, WHO

World Health Organization (1978). *International Classification of Diseases, 9th Revision, Clinical Modification.* Ann Arbor, Michigan: Commission on Professional and Hospital Activities.

World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: WHO.

World Health Organization (2012). Depression fact sheet. Downloaded 03.10.2013. *http://www.who.int/mediacentre/factsheets/fs369/en/index.html*

World Health Organization (2017). *Depression and other common mental disorders: Global health estimates*. Geneve: World Health Organzation.

World Health Organization (2019). *International Statistical Classification of Diseases* (11th ed.). https://icd.who.int/

World Health Organization (2020). Mental health and substance abuse. Downloaded 08.12. 2020. *https://www.who.int/teams/mental-health-and-substance-use/gender-and-women-s-mental-health*

Young, J. E., & Beck, A. T. (1980). *Cognitive Therapy Scale: Rating manual.* Unpublished Manuscript, University of Pennsylvania, Philadelphia, PA.

Zoogman, S., Goldberg, S. B., Hoyt, W. T., Miller, Lisa. (2014). Mindfulness interventions with youth: A meta-analysis. *Mindfulness.* E-puh ahead of publication. http://dx.doi.org/10.1007/s12671-013-0260-4.

Zung, W. W. (1965). A self-rating depression scale. *Archives of Gen. Psychiatry*. 12: 63–70.

Psychological Bulletin

The Effects of Cognitive Behavioral Therapy as an Anti-Depressive Treatment is Falling: A Meta-Analysis

Tom J. Johnsen and Oddgeir Friborg

Online First Publication, May 11, 2015. http://dx.doi.org/10.1037/bul0000015

Psychological Bulletin

0033-2909/15/$12.00 http://dx.doi.org/10.1037/bul0000015

The Effects of Cognitive Behavioral Therapy as an Anti-Depressive Treatment is Falling: A Meta-Analysis

Tom J. Johnsen and Oddgeir Friborg
UiT the Arctic University of Norway, University of Tromsø

A meta-analysis examining temporal changes (time trends) in the effects of cognitive behavioral therapy (CBT) as a treatment for unipolar depression was conducted. A comprehensive search of psychotherapy trials yielded 70 eligible studies from 1977 to 2014. Effect sizes (ES) were quantified as Hedge's *g* based on the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). Rates of remission were also registered. The publication year of each study was examined as a linear metaregression predictor of ES, and as part of a 2-way interaction with other moderators (Year × Moderator). The average ES of the BDI was 1.58 (95% CI [1.43, 1.74]), and 1.69 for the HRSD (95% CI [1.48, 1.89]). Subgroup analyses revealed that women profited more from therapy than did men ($p < .05$). Experienced psychologists ($g = 1.55$) achieved better results ($p < .01$) than less experienced student therapists ($g = 0.98$). The metaregressions examining the temporal trends indicated that the effects of CBT have declined linearly and steadily since its introduction, as measured by patients' self-reports (the BDI, $p < .001$), clinicians' ratings (the HRSD, $p < .01$) and rates of remission ($p < .01$). Subgroup analyses confirmed that the declining trend was present in both within-group (pre/post) designs ($p < .01$) and controlled trial designs ($p = .02$). Thus, modern CBT clinical trials seemingly provided less relief from depressive symptoms as compared with the seminal trials. Potential causes and possible implications for future studies are discussed.

Keywords: cognitive–behavioral therapy, effectiveness, depressive disorders, meta-analysis

Depressive disorders (DDs) can be highly disabling and are ranked third in terms of disease burden as defined by the World Health Organization (WHO, 2014), and first among all psychiatric disorders in terms of disability adjusted life years (Wittchen et al., 2011). In addition, DDs seem to be rising globally (Everyday Health, 2013), and a 20% annual increase in its incidence has been predicted (Healthline, 2012). Improvements in treatment methods and prevention measures, and the availability of community psychiatric services are, therefore, as important as ever before. In response, the WHO has prioritized the combating of depression by launching an action plan called "The Mental Health Gap Action Program," aimed at improving mental health services globally (WHO, 2012).

Psychotherapy is a critical asset for dealing with the future challenges associated with DDs; hence, the optimization of existing therapeutic methods and the development of new ones are important clinical research tasks. Cognitive–behavioral therapy (CBT) has represented an innovative psychotherapy approach since its introduction more than 40 years ago; it has continuously developed and overall, it has been highly successful. The CBT method refers to a class of interventions sharing the basic premise that mental disorders and psychological distress are maintained by cognitive factors or cognitive processes (Hofmann et al., 2012). As posited by Beck (1970) and Ellis (1962), maladaptive thoughts maintain emotional distress and dysfunctional behavior, for which alleviation or cure is realized by changing them. The original theory has been refined continuously by introducing new cognitive concepts (e.g., automatic thoughts, intermediate and core beliefs, and schema theory), and adapted to treat new psychiatric diagnoses. Its potential success in alleviating symptoms of schizophrenia (Rector & Beck, 2012), which was considered impervious to treatment with psychotherapy (Tarrier, 2005), is one striking example. Later variations of the method, building on the foundations of CBT, such as CBT combined with mindfulness (Segal, Williams, & Teasdale, 2002), integrated cognitive therapy with elements of interpersonal therapy (Castonguay, 1996), and metacognitive therapy (Wells, 2000), represent further innovations in CBT. These newer forms of CBT have exhibited promising efficacy in clinical trials of treatments for illnesses, such as hypochondriasis (Lovas & Barsky, 2010) and generalized anxiety disorder (Wells & King, 2006). However, few studies have demonstrated these innovations to be significantly more effective in treating DDs than classical CBT (e.g., Ashouri et al., 2013; Manicavasgar, Parker, & Perich, 2011).

A large amount of research has confirmed the efficacy of classical CBT in treating depression. Meta-analyses published in the 1980s (Dobson, 1989), the 1990s (Hollon, Shelton, & Loosen, 1991; Gloaguen et al., 1998), and after 2000 (Cuijpers et al., 2008; Wampold et al., 2002), concluded that CBT had a high treatment efficacy. Despite the large number of clinical trials and reviews of

Tom J. Johnsen and Oddgeir Friborg, Faculty of Health Sciences, Department of Psychology, UiT the Arctic University of Norway, University of Tromsø.

Correspondence concerning this article should be addressed to Tom J. Johnsen, Faculty of Health Sciences, Department of Psychology, UiT the Arctic University of Norway, University of Tromsø, N-9037 Tromsø, Norway. E-mail: tjj@psykologtromso.no

CBT, to the best of our knowledge, no attempts have been made to evaluate how the efficacy of CBT has evolved over time. Thus, the aim of the present meta-analysis was to study temporal changes (time trends) in the treatment effects of CBT, by posing a simple question: How have the effects of CBT changed over time? Have they improved, stayed the same, or even waned?

A hallmark of our modern society has been the rapid development in many domains, particularly in science, technology, and health. Old procedures and methods have been replaced with safer and more effective solutions. For example, in somatic health care, cruciate ligament surgery currently takes considerably less time, requires fewer resources, and has a better long-term prognosis than it did 30 years ago (Cirstoiu et al., 2011). Another example is a percutaneous coronary intervention (PCI, formerly known as coronary angioplasty), which uses a catheterization technique to insert a stent in the groin or arm to improve blood flow in the heart's arteries. The technique is quick and presently requires minimal rehabilitation (an overnight hospital stay); hence, it represents a huge improvement compared with older techniques (Knapik, 2012). Although comparable improvements in psychiatric methods and techniques are much more difficult to achieve, the purpose of this meta-analysis was to examine whether improvements in CBT, in the treatment of DDs, have taken place since its introduction.

Factors Influencing Treatment Effects

When a treatment is efficacious, psychotherapy research trials point to four sources to explain the observed improvements: (a) client factors, (b) therapist factors, (c) the so-called common factors, and finally, (d) technique-specific factors. Client factors represent the characteristics of the patient, such as personality traits, temperament, motivation for treatment, or important life events experienced by the patient during the course of therapy. Therapist factors are the characteristics of the therapist, which can include anything from gender, age, and education, to personal style and appearance. Clinical training, competency, and skills in establishing a therapeutic alliance and using therapeutic techniques are of particular importance (Crits-Christoph et al., 1991). The two latter components may also be denoted as common and technique-specific therapy factors, which influence the outcome of CBT.

The common factors represent characteristics of the treatment setting that are important and common to all therapy models. These characteristics may include the context of therapy; the client, the therapist, and their relationship (usually coined as the therapeutic alliance); how expectancies for improvement develop; a plausible rationale explaining the patient's illness; or even therapeutic techniques that are not specific to a therapy model. The technique-specific therapy factors represent those elements that are specific to a particular therapy model, and typically are described thoroughly in therapy manuals, indicating specific topics to be addressed during therapy, how they should be conveyed, the implementation of structure, the number of therapy sessions, the degree of exposure, and/or the schedule of homework tasks.

The use of experimental designs has given insight regarding which of these four variance components contribute most to the treatment effect. The major part of the treatment effect seems to be caused by the client-related and common factors, which explain between 30% and 40% and 30%–50% of the total treatment effect, respectively (e.g., Horvath & Greenberg, 1986; Luborsky et al., 1988). The therapist-related factors have been found to explain 5%–15% of the treatment outcomes (Huppert et al., 2001; Wampold & Brown, 2005). That leaves approximately 10%–20% of the effect attributable to the specific therapy (Duncan, Miller, & Sparks, 2004; Lambert, 1992). Recent research has extended our insight into the role of the various components, as it seems that the role of specific versus nonspecific factors in CBT shift with the provision of an increasing number of therapy sessions (Honyashiki et al., 2014). This makes sense, as common factors (e.g., alliance) should be more important in the beginning of therapy, while efficient implementation of treatment-specific factors are increasingly important as therapy progresses. In addition, the role of common factors depends on the mental disorder of the patient. For example, patients with borderline personality disorder may respond much more favorably to the relationship and alliance-building skills of a therapist (Bienenfeld, 2007) compared with patients with bipolar disorders. Although the role of specific versus nonspecific factors may vary, the role of common factors in treating depression is more substantial, as one of the core issues in CBT treatment is to address distorted thoughts related to interpersonal consequences (Castonguay et al., 1996).

Because the common factors seem to be so important for attaining improvement following therapy, psychotherapy researchers have become concerned with them, and how to integrate them into the therapy (Imel & Wampold, 2008). An important line of support of the common factors model comes from meta-analyses showing that different treatment modalities produce relatively comparable treatment effects (e.g., Smith & Glass, 1977; Wampold et al., 1997); hence, the assumption that elements common to all therapies underlie the lack of marked differences among them (Lambert & Bergin, 1994; Seligman, 1995). As specific techniques dictated by a therapy model apparently represent a small part of the overall treatment effect, one would theoretically expect that refinements or improvements of CBT approaches over the past 30 years would have little impact on treatment efficacy, or reported effect sizes (ES). However, the implementation of specific treatment components is usually embedded within a common factors model approach to psychotherapy (Hoffart et al., 2009); otherwise, psychotherapy would stand out as highly decontextualized and mechanistically delivered and experienced by the patient. Therapists who use CBT are trained to establish rapport by, for example, socializing the patient to the cognitive therapy process (thus, being explicit about how the therapy will progress, which may reduce uncertainty), communicating to the patient how CBT might be helpful (instilling hope and positive expectations), and educating the patient about the disorder per se (helping patients to understand their problems). Moreover, CBT therapists set an agenda in collaboration with the patient in order to avoid spending the limited amount of time they have on irrelevant topics. They actively invite the patient to provide feedback (to ensure a mutual understanding and provide opportunities for quick adjustments). They construct and continuously refine their conceptualization of the case (further facilitating and deepening the understanding of the patient's problems). They collaborate actively with the patient in making plans for between-session tasks that may help eliminate negative personal beliefs and behaviors. The latter may help the patient to attribute positive changes to their own efforts, thereby increasing self-efficacy. For this reason, improvements in self-efficacy may be mediated by the use of specific techniques aimed at improving

self-efficacy, in addition to an effective integration of the common factors. The integration of the common factors is, thus, utterly important as they represent the chassis that enables the motor to move the vehicle forward. An important part in this context is the working alliance between the therapist and the patient, which is associated with quicker and larger treatment effects (Rector, Zuroff, & Segal, 1999), and a reduction of the number of early dropouts (Kegel & Fluckiger, 2014).

Although CBT treatments have focused less on the common factors, we believe that CBT therapists have become increasingly aware of the importance of integrating common and specific techniques to take full advantage of the therapy. Therefore, we expected that contemporary CBT treatments would show better treatment outcomes as compared with older clinical trials. If not, that would be a quite interesting and unexpected finding, which would warrant timely questions about the direction of CBT in the future, such as, "Should CBT researchers continue to improve current techniques of CBT?" and "Should they improve the integration of common factors, or should they enhance CBT via the inclusion of, for example, metacognitive (Wells, 2000) or transdiagnostic aspects (Fairburn, Cooper, & Shafran, 2003)"? In order to examine if the client, therapist, or common factors were related to treatment effects differently across time, we included all of the available data related to these components in the CBT studies in the meta-analysis.

An advantage of examining the temporal trends in treatment effects based on CBT trials is the high degree of standardization, a factor that has not changed appreciably over the years. Since the 1970s, almost all studies have utilized the Beck Depression Inventory (BDI; Beck et al., 1961). The BDI is a self-report rating inventory that measures different attitudes, symptoms, and behaviors that characterize depression. The internal consistency is generally good with high alpha coefficients (e.g., .86 and .81 in psychiatric and nonpsychiatric populations; Beck, Steer, & Carbin, 1988). The other depression measures have been more variable, with the exception of the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), which has been utilized more frequently, in conjunction with the BDI or by itself. The HRSD is a clinician administered rating scale, measuring similar characteristics of depression as the BDI. The interrater reliability is generally high with coefficients typically exceeding .84 (Hedlund & Vieweg, 1979). The correlation between the BDI and the HRSD is in the moderate to high range, $r = .5$ to .8 (Beck et al., 1988; Beck, Steer, & Brown, 1996). Moreover, most clinical researchers have followed a standardized CBT treatment manual that therapists have been trained to deliver. This methodological allegiance has allowed a more empirically valid and reliable comparison of effect sizes for CBT interventions across the decades.

Moderators of Temporal Treatment Effects

In addition to the temporal (i.e., "time") factor in the present study, we examined the role of selected moderator variables that are available in most clinical studies. The following client-specific factors were included in the analysis: gender, age, degree of psychiatric comorbidity, use of psychotropic medication, severity of depression, and provision of CBT to special patient samples (such as those with diabetes). The therapist-related factors were the type of therapist (e.g., psychologist or student) and ratings of the competence of the therapist. The treatment-specific and methodological factors included the publication year, number of therapy sessions, application of the original CBT manual (Beck et al., 1979) or not, checks of adherence to the treatment protocol (including subsequent feedback to the therapists), type of statistical analyses (intention to treat [ITT] or completers only), and ratings of the methodological quality of the study. The only available variable indicating common factors was the ratings of the therapeutic alliance; however; the number of studies reporting the alliance was disappointingly small.

Client-Related

Previous studies have typically not revealed any significant differences in treatment effects related to gender and age (Joutsenniemi et al., 2012; Wierzbicki & Pekarik, 1993). A higher degree of psychiatric comorbidity often implies a worse course of illness or treatment prognosis. The most common Axis I comorbidity is anxiety disorders (Kessler et al., 2003), which usually imply a higher degree of severity at intake (Kohler et al., 2013), as well as a poorer natural course (Penninx et al., 2011). The presence of comorbid Axis-II disorders, of which the Cluster C diagnoses, particularly, avoidant personality disorder, are the most prevalent (Friborg et al., 2014), heightens the risk of a worse outcome following treatment (Newton-Howes, Tyrer, & Johnson, 2006). The relative efficacy of psychotropic medication versus CBT has been subjected to many clinical trials; however, a meta-analysis of 21 studies found no differences between the two treatment modalities in alleviating depression (Roshanaei-Moghaddam et al., 2011). The addition of medication to CBT has been studied to a lesser degree; however, a meta-analysis consisting of seven studies found that CBT plus medication was slightly better ($d = 0.32$) than CBT alone (Cuijpers et al., 2009). The present meta-analysis is not entirely comparable with the study by Cuijpers et al., as we recorded the percentage of patients receiving simultaneous medication. With regard to the severity of depression, previous research has found that patients who were more severely depressed reported larger treatment effects than less severely depressed patients, a phenomenon also known as regression to the mean (Garfield, 1986; Lambert, 2001). Some of the CBT studies that were included in the present meta-analysis also recruited patients who had other somatic illnesses or difficulties in addition to depression, for instance, diabetes, alcoholism, or marital discord. Few previous studies (if any) have examined whether these patients respond differently to CBT treatment than purely depressed patients. However, as the effect size has tended to be lower for patients with psychiatric comorbidities, one also could expect a similar trend among patients having somatic or other ailments in addition to depression.

Therapist-Related

More therapeutic experience has been found to relate to a shorter time to remission (Okiishi et al., 2006), and hence, psychologists should do better than student therapists should. In the current study, three types of therapists were registered: psychiatrists, psychologists, and psychology students.

Treatment-Specific/Methodological

A dose-response relationship has been documented, in that additional sessions of therapy usually lead to a higher treatment efficacy (e.g., Howard et al., 1986). As adherence to a treatment manual ensures the implementation of CBT and improves the outcome (Shafran et al., 2009), we expected a similar relationship in the present meta-analysis. Studies integrating adherence or fidelity checks should demonstrate higher ESs than those without such checks. In the same vein, we expected that the use of the Beck treatment manual would yield stronger treatment effects than trials not using it, as fidelity checks are often involved with its use. As studies using stricter criteria with regard to methodological quality (Gould, Coulson, & Howard, 2012), and studies using between rather than within group designs (Pallesen et al., 2005) generally yield lower treatment effects, we expected the same results in the present analysis. A recent meta-analysis (Hans & Hiller, 2013) showed a slightly larger effect size in depression treatment trials using a statistical design requiring treatment completion ($d = 1.13$), as compared with an ITT design ($d = 1.06$); therefore, we expected the same trend in the present study.

Common Factors

Patients experiencing a stronger alliance with their therapist were expected to report better effects of their treatment (Rector et al., 1999).

Meta-Analytic Advantages and Objectives

The benefits of using meta-analytic methods to summarize clinical results are well-known (Borenstein et al., 2009). By accessing a large pool of studies and assigning the individual studies different weights according to their sample size, the potentially troublesome role of individual studies indicating weak or even contradictory results is minimized. A meta-analysis is also preferable in situations where the majority of studies are well-defined or similar in terms of patients, diagnoses, intervention procedures, and the measurement instrument used (e.g., the BDI), thus, simplifying the quantification of the effect size considerably. Moreover, metaregression approaches may be used to identify potential sources of covariation between study-related factors and treatment effects.

Objectives of the Present Study

The primary objective was to examine whether published clinical CBT trials (both uncontrolled and randomized controlled) aimed at treating unipolar DDs demonstrate a historical change in treatment effects, independent of study-related moderating variables. A more effective therapy should demonstrate larger positive changes in prepost scores, as rated by the patients (the BDI) and the therapists (the HRSD) over the years.

The secondary purpose was to examine the role of various moderators of the reported effect sizes. We predicted that diagnostic severity and type of therapist (psychologist better than student therapist), and therapist competency would be associated with better treatment effects, while the variables age and gender were not expected to covary with therapeutic outcome. Finally, we examined whether these moderators modified the regression slopes describing the time trends in the treatment effects.

Method

Data Collection, Studies, and Selection Criteria

We used the OvidSP Internet-based platform to identify relevant empirical English-language studies. All searches were conducted in January 2015 using the following databases (without publication year restrictions): PsycINFO, APA PsycNET, Embase, and Ovid Medline. In PsycINFO, the query "treatment effectiveness evaluation" returned 14,935 titles. In APA PsycNET, the search "depression *and* study" and "depression *and* treatment" returned 5,996 and 1,974 titles, respectively. A third query in all databases using "depression *and* efficacy or efficacious" returned 4,353 titles. A final query in all the databases using the phrase, "depression *and* trial *and* cognitive" returned an additional 1,793 titles. The total number of titles was 29,051. After examining the titles, 1,670 abstracts were considered relevant. Following a review of the abstracts, 489 articles were obtained via the university library. The following exclusion criteria were then applied: (a) the implemented therapy was not pure CBT. Thus, we did not include studies/study arms of CBT combined with other treatment forms, such as mindfulness based CBT. We did include one study arm consisting of integrative CT (Castonguay et al., 2004) as the published treatment protocol, in essence, indicated standard cognitive therapy, albeit, with an additional structured procedure for repairing any ruptures in the patient-therapist alliance. Among the studies comparing CBT with other treatment forms (interpersonal therapy, for instance), we included only the CBT treatment arm; (b) a unipolar DD (either mild, moderate, severe, or recurrent) was not the primary psychiatric diagnosis; (c) participants were not adults (mean age < 18); (d) therapy was not implemented by a therapist trained in CBT; (e) the psychotherapeutic intervention was not intended to treat depression; (f) the outcome was not measured with the BDI or the HRSD; (g) patients had acute physical illnesses or suffered from bipolar or psychotic disorders; (h) treatment was not implemented as individual face-to-face therapy; and (i) the patients had a BDI score lower than 13.5. The last criterion is in accordance with the manual of the revised BDI, and several depression treatment researchers (Beach & O'Leary, 1992; Emanuels-Zuurven & Emmelkamp, 1997; Kendall et al., 1987; Murphy et al., 1995; Wright et al., 2005).

If a study assigned patients to different subgroups based on diagnostic severity (usually based on the pretest BDI scores), only the most severe subgroup was included to avoid inflating the number of independent studies. This procedure was relevant for three studies. For the same reason, if a study assigned patients to treatment subgroups consisting of one group with CBT, and one group with CBT plus medication, we only included the pure CBT group in our analysis. The selection procedure yielded a final study pool of 70 studies (see Figure 1).

Coding of Study Information and Moderator Variables

The following data from the studies were coded: demographic information (gender and age), year of implementation of the in-

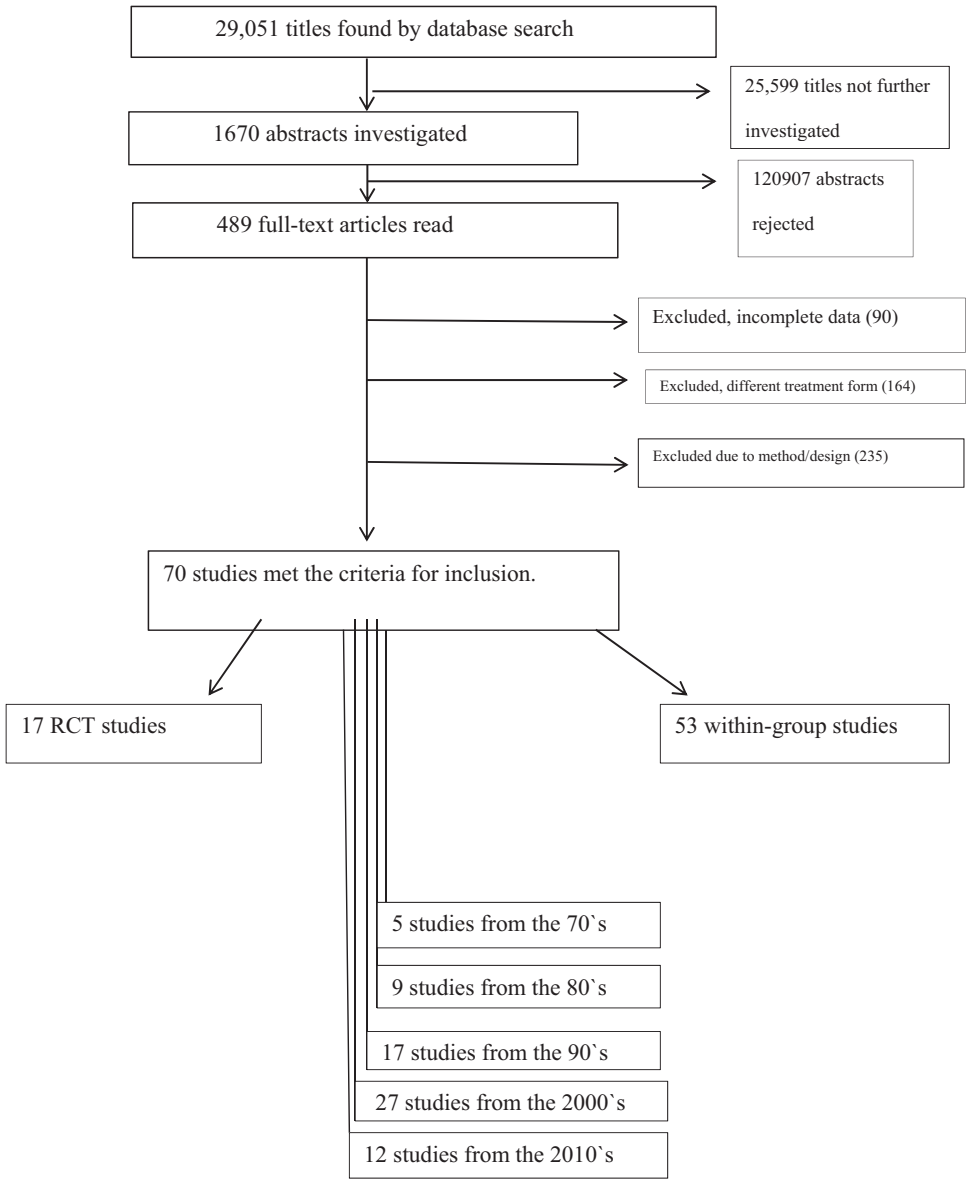


Figure 1. Flowchart of the search and selection procedure.

tervention, duration (number of sessions), type of therapist (psychologist, trained psychology-student, or other/unknown), therapist competence (as measured by the Cognitive Therapy Scale), information about the severity of the diagnosis (mild, moderate, severe, or recurrent depression) along with the proportion (%) of the sample having comorbid psychiatric diagnoses, whether the patient population had any special characteristics (marital discord, HIV, multiple sclerosis, diabetes, Parkinson's disease, alcohol abuse disorders or pregnancy), and the proportion (%) of patients using psychotropic medication. The DD diagnoses of the patients were coded according to the original authors' definitions. If unreported, we categorized the DD diagnoses based on the BDI pretest scores as mild (13–19.5), moderate (20–29.5), or severe ($>$ 30). We coded recurrent depression as the main diagnosis if at least half of the patients previously had two or more episodes of depression.

The Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS) was used to rate the methodological quality of the published studies (Kocsis et al., 2010). It is a comprehensive instrument consisting of 24 items measuring six study quality dimensions: (a) adequate descriptions of subjects; (b) the definition and delivery of treatment; (c) the quality of the outcome measures utilized; (d) the data analyses (e.g., description of dropouts, ITT, appropriate tests); (e) strong methods for assignment to treatment groups; and (f) an overall quality rating. Each item is assigned a score of 0 (*poor description, execution, or justification of a design element*), 1 (*brief description or either a good description or an appropriate method or criteria set, but not both*), or 2 (*well described, executed, and, where necessary, justified design element*). The scale yields a total score ranging from 0 to 48, which was used in a subsequent metaregression analysis.

Moderator Analyses

We investigated whether the effect sizes covaried with any of the following moderator variables: type of statistical analysis (ITT vs. completers analysis), gender (as % men), age, proportion of patients using medication, proportion of comorbidity, use of the Beck CBT treatment manual versus no manual, checks (and subsequent feedback) of therapist adherence to the treatment manual versus no adherence check, version of BDI (I or II), severity of the depressive disorder, diversity of the study populations (ordinary depressed patients vs. patients with co-occurring illnesses or other special characteristics, such as Parkinson's, HIV, diabetes, marital discord, alcoholism, or multiple sclerosis), number of therapy sessions, type of therapist, therapist competency, and the publication year of the CBT intervention (the moderator of most interest). We also examined whether the latter variable covaried with the effect sizes in the waiting list control groups. The competence of the therapist was, in a few studies, rated using the Cognitive Therapy Scale (CTS; Dobson et al., 1985), and it was included as a moderator. The CTS is an observer-based rating scale (usually rated by an expert in CBT) designed to measure how well the therapist applies CBT across several therapist skills dimensions, including adherence to the manual.

Effect Sizes

We used two procedures when calculating the effect sizes based on the BDI and the HRSD pre-/postintervention scores: a prepost within-study design, and a controlled trial (CT) design. For studies that did not include a no-intervention control group, a standardized mean difference (SMD, also denoted Cohen's *d*) was calculated for the intervention group ($M_{pre} - M_{post}$, divided by the standard deviation of the change score). A Hedges *g* correction was applied to the SMD, which reduced the SMD for studies having small sample sizes (Hedges & Olkin, 1985). The vast majority of these intervention studies were drawn from different randomized controlled trials, but because of methodological choices or study design issues, they could not be categorized as CTs in the present analysis. For example, some studies compared CBT with antidepressant medication or other forms of therapy, and hence, for these studies only, the intervention group receiving CBT was included as a within-group study.

For the controlled trials condition (which included 15 randomized studies with a waiting list condition as the control group, and two studies with a treatment as usual type of control group without specific interventions for depression), the effect sizes were calculated from the difference between the pre- and posttest scores on the BDI and the HRSD for the intervention group and the control group, respectively, and then standardized using the change scores. This method was preferred to standardization using post scores, because studies including a smaller number of participants might contain preintervention differences despite randomization. The change score variant is less sensitive to such differences compared to standardization using post scores. Another advantage of using the *SD* for change scores is that the effect sizes for CT studies are estimated similarly as studies without a control group (within-study designs). Standardization by change scores also is recommended when the objective is to assess change relative to preintervention scores (Kulinskaya et al., 2002), and it has frequently been used to quantify treatment effects in other meta-analytic reviews of psychotherapy (e.g., Abbass et al., 2013; Kishi et al., 2012; McGuire et al., 2014; Watts et al., 2013; Zoogman et al., 2014). However, one limitation is that change scores require knowledge of the prepost correlation, and consequently, we imputed a conservative value of $r = .7$ for studies that did not report one ($k = 65$), as recommended by Rosenthal (1993).

When available, we calculated the ES based on scores from completers of an intervention (51 studies). The remaining studies only provided data from ITT samples (19 studies), and were thus coded accordingly.

The effect sizes for the treatment recovery rates (the number of patients who ended treatment with a BDI score below a predefined clinical cut-off score, <10) were coded as an event rate (rate = number of events/sample size), which the Comprehensive Meta-Analysis (CMA) program linearized by calculating the logit before estimating the metaregression coefficients. This method also is known to yield standard errors that are more accurate.

Interrater Reliability

The second author (OF), coded a random sample of 20 studies. The level of agreement between the raters (the first and second author) was determined by Kappa (κ) coefficients for dichotomously scored variables, and intraclass correlation coefficients (two-way mixed, average models) for continuously scored variables. Kappa coefficients (range: 0-no agreement, 1-perfect agreement) within the range of .41–.60 and .61–.80 were interpreted as moderate versus substantial agreement, respectively (Rigby, 2000). The ICCs (range: 0-no agreement, 1-perfect agreement) were interpreted similarly as Cronbach's alpha, with ICCs > .70 and > .80, indicating moderate and high consistency, respectively, between the raters. The coefficients were: BDI effect size calculations (ICC = .95), publication year (ICC = 1.0), study design (κ = .77), diagnosis (κ = .63), gender % (ICC = .99), therapist (κ = .59), no. of session (ICC = .97), patient's age (ICC = 1.0), remission rate (ICC = .83), type of analysis (κ = .69), comorbidity % (ICC = .96), use of the Beck manual (κ = .62), BDI version (κ = 1.0), and study quality (ICC = .89). The interrater reliability analyses thus revealed substantial agreement. The studies with disparate ratings were followed-up by the two coders, and agreement was reached by consensus following a discussion. It turned out that the first author had coded almost all of the disparate cases correctly, which was reassuring as the first author coded all of the studies.

Quantitative Data Synthesis and Statistical Calculations

The CMA software, Version 2 (Borenstein et al., 2005) was used for all statistical analyses, except for the two-way interaction analyses between the moderator variables, which had to be analyzed in SPSS 21. The random weights from the CMA program were imported into SPSS and a weighted least-squares regression analysis was conducted.

The average weighted effect sizes were estimated according to a random-effects model (in preference to a fixed effects model), as we assumed the true effect sizes would vary between studies due to the study-related factors, for example, severity of diagnosis, age, or gender. Employing a random-effects model also increases the

generalizability of the results (Field, 2003). A Q-test statistic (chi-square distributed) was calculated to examine whether the variance between studies was larger than the variance within the studies, thus, indicating predictors (or moderators) of between-study variation. Metaregression analyses were used to analyze the role of the continuous moderator variables (e.g., publication year), and were based on the unrestricted maximum-likelihood method, as it assumes an underlying random distribution of effect sizes. The moderator analyses for the categorical variables were based on a similar Q-test statistic to examine whether the variability between categories (subgroups in the study) was larger than the variability within studies. The I^2 statistic also was reported to indicate the amount of heterogeneity that was related to the true differences in effect sizes between studies, relative to sampling error. The influence of the time variable on ES was examined based on both the BDI and the HRSD, in addition to the remission rates. The associations of the other moderator variables and ES were examined based on the BDI measure, which had the largest number of studies.

Publication Bias

To measure the potential biasing effect of including studies with few participants, we visually inspected the funnel plot and used Duval and Tweedie's trim-and-fill method.

Results

Studies and Participants

The search procedure resulted in 70 eligible studies with CBT implemented as individual therapy for depression. Fifty-two studies were randomized controlled trials, five studies were controlled trials without randomization, two studies were uncontrolled and nonrandomized (pilot studies), while 11 were clinical field studies. The studies were conducted from 1977 to 2014, with 1999 as the average year. Seventeen studies were categorized as CT in the present meta-analysis (those including a waiting list control group), while 53 were categorized as within-group studies (those lacking a waiting list control group, plus the 11 field studies). The average quality rating of the studies was 28.4 (SD = 7.5, range 7–42).

The total number of patients was 2,426. The number varied from seven to 217 patients in the studies, with an average of 34.6 patients per study (SD = 34.1). Males accounted for 30.9% of the patients, and the average age was 40.5 years (SD = 10.9). On average, 43% of the patients had a comorbid psychiatric diagnosis (SD = 23%, k = 26 number of studies). Thirteen studies included patient populations having other conditions or characteristics, for example, marital discord or diabetes.

The mean BDI preintervention score was 26.1 (SD = 4.1). Fifty-seven percent (SD = 18.9) of the patients had remissions from depression following treatment (k = 43). The patients received an average of 14.6 sessions of CBT (SD = 5.12, range = 6–34). Sixty-seven studies included prepost data from the BDI (48 used the BDI-I and 19 used the BDI-II), and 34 studies included depression endpoints based on the HRSD. See Table 1 for a descriptive overview of the included studies.

Effects of CBT

The average weighted effect size for the BDI (k = 67) was g = 1.58 (95% CI [1.43, 1.74]). The variance in the effect sizes and the associated confidence intervals are presented in a forest plot (see Figure 2). A Q-test indicated that the methodological design did not yield significantly different (p = .13) treatment effects (within-group, g = 1.65 vs. between-group CT, g = 1.37). The difference in the weighted ES between the ITT and completers was not significantly different (p = .34). For the HRSD (k = 34), the average ES was 1.69 (95% CI [1.48, 1.89]). The HRSD effect sizes and the associated CIs are presented in Figure 3. The methodological design again revealed no significant (p = .10) effect size differences (within-group, g = 1.81; between-Group CT, g = 1.44).

Are CBT Treatment Effects Contingent on the Year of Publication?

The CBT effect sizes (based on the BDI) had a significant negative relationship with time, that is, publication year ($p < .001$, see Table 2 for coefficients and Figure 4 for a scatterplot). According to a subgroup analysis, a similar negative relationship was evident among studies using within-group designs ($p < .001$), and CT designs ($p < .05$).

The effect sizes for the HRSD showed a comparable picture (Table 2 and Figure 5). The ES decreased with time (p = .01). The significant negative relationship was evident for the within-group design studies ($p < .01$). The ES in the CT studies also showed a declining trend, but it was not significant (p = .51).

The remission rates (percentage of patients recovering) also were negatively related with publication year ($p < .01$; see Figure 6).

Because Figure 4 indicated an apparent decline in the effect sizes for studies conducted in 1995–2002, these studies were excluded to determine if they had an undue influence on the results, but the results the same ($p < .001$, see upper part of Table 3). A similar inspection was done for the 11 clinical field studies, but excluding these studies did not change the results either (p = .001). We also examined how the slope of the regression line for time changed when we excluded studies consecutively, beginning with the first publication year in 1977. As seen in Figure 7, all of the coefficients for time were negative, except for those from the studies published between 1994 and 1997. However, the decline in treatment effects was again evident from 1998 and onward.

The waiting list control group condition exhibited no significant changes in effect sizes across time (p = .48).

Publication Bias

The funnel plots for all of the CBT studies suggested a certain degree of publication bias for ESs based on the BDI. A significant proportion of the effect sizes were plotted to the upper left of the inverted curve, which suggests that the studies with low numbers of participants had a higher ESs than the studies with more participants. This was not the case for the HRSD, which showed a more symmetrical plot (see Figures 8 and 9).

Duvall and Tweedie's trim-and-fill method also indicated a bias for the BDI, but not for the HRSD. Consequently, nine studies

Table 1
A Descriptive Overview of the 70 CBT Studies Included

Author (publication year)	Patient characteristic	Trial	*N*	*ES* BDI (HRSD)	Rec %	Sessions
Rush et al.(1977)	None	RCT	18	3.68 (4.17)	79	15
Taylor & Marshall (1977)	None	**RCT**	15	1.71	—	6
Carrington (1979)	Women	**RCT**	11	2.49	—	12
Dunn (1979)	None	RCT	10	2.16	—	8
Mclean & Hakstian (1979)	None	RCT	10	2.50	50	10
Liberman & Eckman (1981)	Suicide attempters	RCT	12	2.22	—	32
Gallagher & Thompson (1982)	Elderly	RCT	27	2.12 (1.85)	73	16
Wilson et al. (1983)	None	**RCT**	8	2.94 (2.19)	—	8
Beck et al. (1985)	None	RCT	18	2.64 (3.34)	71	14
McNamara & Horan (1986)	None	RCT	10	3.95	80	10
Thompson et al. (1987)	Elderly	**RCT**	27	1.97 (0.93)	52	16
Wierzbicki & Bartlett (1987)	None	**RCT**	11	1.97	—	12
Persons et al. (1988)	None	FS	35	1.53	80	25
Elkin et al. (1989)	None	RCT	37	1.90 (2.11)	70	12
Selmi et al. (1990)	None	**RCT**	12	1.07 (1.89)	42	6
Jacobson (1991)	Marital discord	CT	7	2.63 (2.24)	58	20
Thase et al. (1991)	None	NRT	38	2.42 (3.13)	61	20
Beach & O'Leary (1992)	Marital discord	**RCT**	15	1.36	—	16
Hollon et al. (1992)	None	RCT	16	2.44	—	15
Propst et al. (1992)	Religious	**RCT**	10	1.39 (0.82)	90	18
Gallagher-Thompson & Steffen (1994)	Elderly	RCT	36	1.32 (1.08)	77	20
Shapiro et al. (1994)	None	RCT	18	2.12	—	12
Murphy et al. (1995)	None	RCT	11	2.39 (2.85)	82	17
Teichman et al. (1995)	None	**RCT**	38	0.24	14	13
Emanuels-Zuurveen & Emmelkamp (1996)	Marital discord	RCT	14	0.38	—	16
Jacobson et al. (1996)	None	RCT	50	2.19 (2.28)	71	16
Blackburn & Moore (1997)	None	RCT	24	0.70 (1.12)	33	16
Brown et al. (1997)	Alcoholism	RCT	19	1.68 (1.47)	—	8
Emanuels-Zuurveen & Emmelkamp (1997)	None	RCT	10	0.97	—	16
Markowitz et al. (1998)	HIV	**RCT**	17	0.67 (0.95)	40	12
Persons et al. (1999)	None	CT	27	1.32	80	34
King et al. (2000)	None	RCT	63	1.63	74	12
Thase et al. (2000)	None	CT	52	1.35 (1.80)	38	16
Thompson et al. (2001)	Elder	RCT	31	0.39 (1.61)	—	16
Cahill et al. (2003)	None	FS	30	1.55	—	16
Merrill et al. (2003)	None	FS	100	1.81	55	8
Watson et al. (2003)	None	RCT	33	1.64	—	16
Castonguay et al. (2004)	None	**RCT**	11	2.44 (1.16)	100	17
Misri et al. (2004)	Postpartum	RCT	19	−(1.92)	63	12
Hardy et al. (2005)	None	FS	76	1.33	61	12
Westbrook & Kirk (2005)	None	FS	95	1.42	36	13
Wright et al. (2005)	None	**RCT**	15	1.82 (1.06)	—	9
Dimidjian et al. (2006)	None	RCT	18	1.00 (1.51)	48	16
Jarrett et al. (2007)	None	FS	126	2.21 (1.82)	63	20
McBride et al. (2006)	None	RCT	29	2.73	—	17
Persons et al. (2006)	None	FS	38	1.17	—	18
Strauman et al. (2006)	None	RCT	7	1.97 (2.04)	33	18
Dobkin et al. (2007)	Parkinson	Pilot	13	1.12	—	12
Forman et al. (2007)	None	RCT	44	0.65	61	16
Luty et al. (2007)	None	RCT	86	1.18 (1.44)	43	13
Cho et al. (2008)	Pregnant	**RCT**	12	1.74	—	18
Constantino et al. (2008)	None	RCT	11	1.53	82	16
David et al. (2008)	None	RCT	56	2.20 (2.11)	50	20
Laidlaw et al. (2008)	Elder	RCT	20	1.23 (1.35)	—	8
Quilty et al. (2008)	None	RCT	45	1.73	—	18
Craigie & Nathan (2009)	None	FS	77	1.87	53	11
Gibbons et al. (2010)	None	FS	217	0.7	36	16
Dobkin et al. (2011)	Parkinson	**RCT**	41	1.37 (2.03)	—	10
Mohr et al. (2001)	Mul. Scler.	RCT	20	1.42 (1.31)	40	16
Rieu et al. (2011)	None	RCT	11	1.07 (1.09)	—	6
Estupina & Encinas (2012)	None	FS	30	2.09	80	18
Power & Freeman (2012)	None	**RCT**	46	0.38	50	16
Ammerman et al. (2013)	Mothers	**CT**	47	1.18 (1.05)	—	11
Gibbons et al. (2013)	None	CT	18	2.21	—	19

(*table continues*)

Table 1 (*continued*)

Author (publication year)	Patient characteristic	Trial	*N*	*ES* BDI (HRSD)	Rec %	Sessions
Kohler et al. (2013)	None	FS	105	1.50 (3.02)	58	14
Parker et al. (2013)	None	RCT	11	−(0.74)	45.4	10
Wagner et al. (2014)	None	RCT	28	1.30	50	8
Lopes et al. (2014)	None	RCT	29	1.16	40	14
Tovote et al. (2014)	Diabetes	**RCT**	32	0.92	29	8
Kalapatapu et al. (2014)	**Alcoholism**	**RCT**	**53**	**−(1.84)**	**41**	**18**

Note. CT = controlled trial, not randomized; NRT = nonrandomized, noncontrolled trial; FS = field study; RCT = randomized controlled trial; *ES* = Hedge's *g*; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale of Depression; Rec % = percentage of patients remitting; Sessions = number of treatment sessions. Studies in bold (**RCT**) are the ones with a nonintervention comparison group, hence the ones subsequently included in the CT condition in this meta-analysis.

were trimmed, which adjusted the *g* from 1.58 to 1.46. However, removal of all studies (30 in total) with small sample sizes ($n < 20$) did not change the above findings; the slope was still negative ($p < .05$, see Table 3). The removal of these studies also excluded the two potential outliers with the highest ESs observed in Figure 2, without having a substantial influence on the outcome.

Moderators Related to Client, Therapist, Treatment-Specific/Methodologies, and Common Factors

A separate analysis for each moderator variable was conducted.

Client-related. Age was not significantly related to variation in treatment effects; however, the gender variable was ($p < .05$). Studies that included a higher percentage of women demonstrated a better treatment effect than studies consisting of more men. The proportion of comorbid psychiatric diagnoses in the studies did not significantly moderate the reported weighted ESs, nor did the proportion of psychotropic medication use or the severity of diagnosis (see Table 3). Milder depression, although not statistically significant, tended to yield lower treatment effects compared with more severe or recurrent depression. The low number of available studies in some of the subgroups (particularly the recurrent group), speaks to the need for exercising caution in interpreting these results. Diagnostic diversity in the patient group was not significantly related to ES. The 13 studies, including those with patients with special characteristics, (e.g., comorbid somatic diseases or marital discord problems), did not significantly differ from patients with depression only (see Table 4). Excluding studies with special characteristics did not change the negative temporal trend in the treatment effects. A marginally significant negative trend in ESs with time was also observed among the 13 studies with special characteristics (see Tables 3 and 4).

Therapist-related. Therapist competency did not have a significant relationship with treatment effects. However, the number of available studies was low ($k = 5$), implying low statistical power and a high vulnerability to bias of the results from single studies. Yet, the regression line was positive as expected, indicating higher ESs with higher levels of competence. The effect size differences between types of therapists were significant ($p < .01$), indicating that trained psychologists achieved better treatment effects ($g = 1.59$) than psychology students ($g = 0.98$).

Treatment-specific/methodological factors. The number of therapy sessions was not related to a better treatment effect; neither was the use of the Beck CBT manual, adherence checks, the data analysis method (ITT vs. completers), or the study quality ratings (see Tables 3 and 4). A nonlinear weighted regression model, which examined whether shorter or longer therapy trials yielded poorer treatment results compared to a moderate amount, was not significant ($p = .99$).

Common-factors. Seven studies contained information about the patient–therapist alliance. However, five of the studies used qualitative or customized measures that were not suitable for quantification and statistical analysis. Only two studies provided quantitative data based on standardized measures of alliance. Thus, the role of common factors was not possible to analyze.

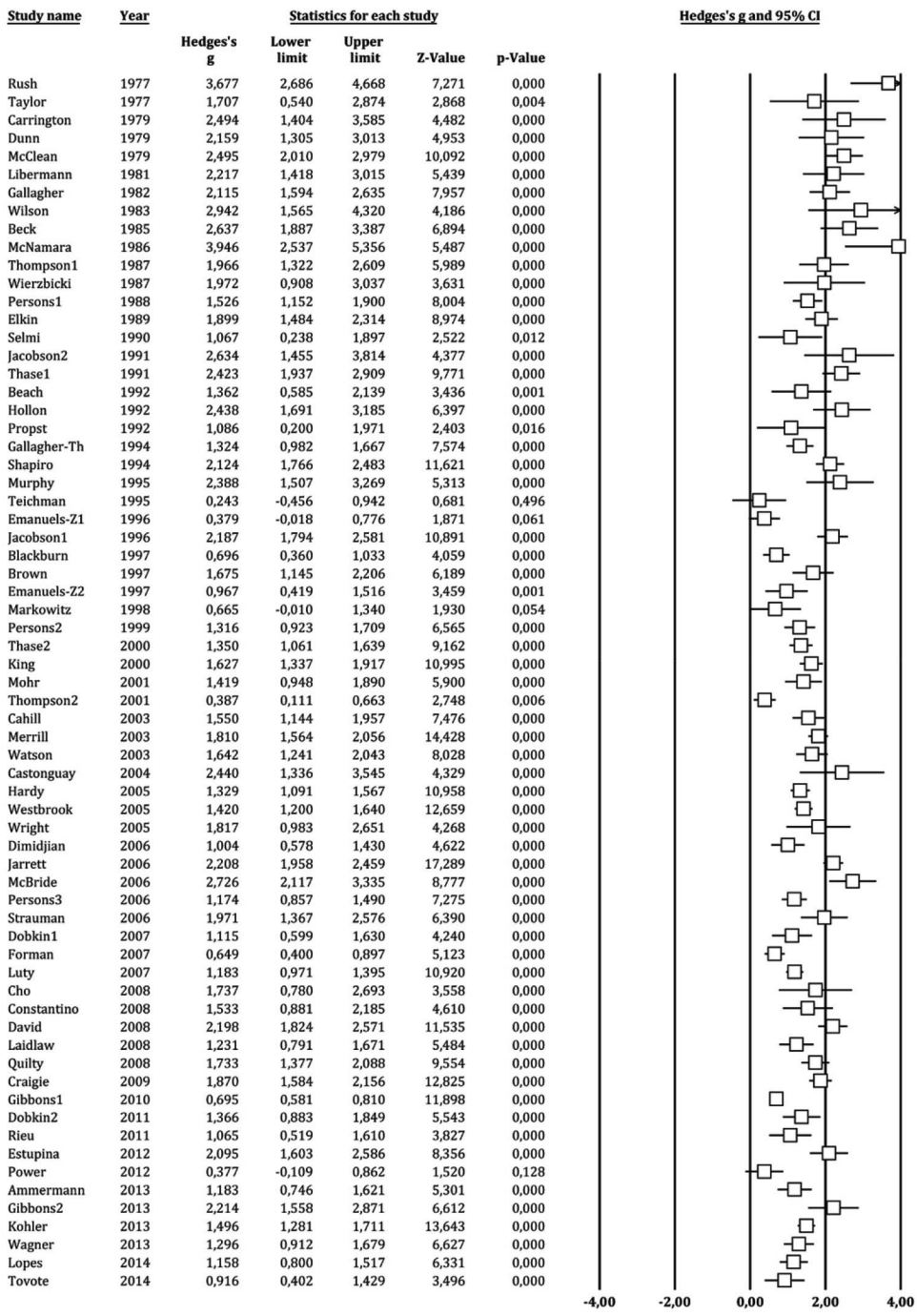
Correlations Between Time and Moderator Variables and Two-Way Interaction Tests (Time × Moderator)

The weighted correlation coefficients between *time* (publication year) and the moderator variables were as follows: (a) *client-related*: gender (male %, $r = .09$, $p = .48$), age, $r = .08$, $p = .53$, preintervention score BDI, $r = .26$, $p = .04$, comorbidity %, $r = -.14$, $p = .52$, medication %, $r = .25$, $p = .10$, patient (psychiatric vs. special) type, $r = .05$, $p = .69$, and severity (mild-moderate-severe) of depression, $r = -.04$, $p = .78$); (b) *therapist-related*: type of (student vs. psychologist) therapist, $r = .17$, $p = .26$); and (c) *study-related*: number of therapy sessions, $r = -.08$, $p = .52$, methodological quality, $r = .43$, $p < .001$, type of statistical (ITT vs. completers) analysis, $r = -.17$, $p = .17$, use of the Beck manual (no vs. yes) manual, $r = -.13$, $p = .29$, and BDI (I vs. II) version, $r = .59$, $p < .001$.

These analyses indicate that the methodological quality has improved significantly over the years. Newer studies also include more patients with higher initial BDI scores than the older studies, and employ the BDI-II rather than the original BDI-I version. Patients on medication are also more frequently included, but this coefficient was not significant.

Two-Way Interaction Tests

Finally, we examined whether the observed decline in the treatment effects depended on any of the above moderators by conducting two-way interaction tests (Time × Moderator). If the interaction coefficient was significant, or its unstandardized weight ($beta_{int}$) was positive and higher than the unstandardized time coefficient ($beta_{time}$), that would indicate the slope depended on the moderator and qualitatively changed its direction following the inclusion of the moderator. Conversely, a negative interaction



Study name	Year	Hedges's g	Lower limit	Upper limit	Z-Value	p-Value
Rush	1977	3,677	2,686	4,668	7,271	0,000
Taylor	1977	1,707	0,540	2,874	2,868	0,004
Carrington	1979	2,494	1,404	3,585	4,482	0,000
Dunn	1979	2,159	1,305	3,013	4,953	0,000
McClean	1979	2,495	2,010	2,979	10,092	0,000
Libermann	1981	2,217	1,418	3,015	5,439	0,000
Gallagher	1982	2,115	1,594	2,635	7,957	0,000
Wilson	1983	2,942	1,565	4,320	4,186	0,000
Beck	1985	2,637	1,887	3,387	6,894	0,000
McNamara	1986	3,946	2,537	5,356	5,487	0,000
Thompson1	1987	1,966	1,322	2,609	5,989	0,000
Wierzbicki	1987	1,972	0,908	3,037	3,631	0,000
Persons1	1988	1,526	1,152	1,900	8,004	0,000
Elkin	1989	1,899	1,484	2,314	8,974	0,000
Selmi	1990	1,067	0,238	1,897	2,522	0,012
Jacobson2	1991	2,634	1,455	3,814	4,377	0,000
Thase1	1991	2,423	1,937	2,909	9,771	0,000
Beach	1992	1,362	0,585	2,139	3,436	0,001
Hollon	1992	2,438	1,691	3,185	6,397	0,000
Propst	1992	1,086	0,200	1,971	2,403	0,016
Gallagher-Th	1994	1,324	0,982	1,667	7,574	0,000
Shapiro	1994	2,124	1,766	2,483	11,621	0,000
Murphy	1995	2,388	1,507	3,269	5,313	0,000
Teichman	1995	0,243	-0,456	0,942	0,681	0,496
Emanuels-Z1	1996	0,379	-0,018	0,776	1,871	0,061
Jacobson1	1996	2,187	1,794	2,581	10,891	0,000
Blackburn	1997	0,696	0,360	1,033	4,059	0,000
Brown	1997	1,675	1,145	2,206	6,189	0,000
Emanuels-Z2	1997	0,967	0,419	1,516	3,459	0,001
Markowitz	1998	0,665	-0,010	1,340	1,930	0,054
Persons2	1999	1,316	0,923	1,709	6,565	0,000
Thase2	2000	1,350	1,061	1,639	9,162	0,000
King	2000	1,627	1,337	1,917	10,995	0,000
Mohr	2001	1,419	0,948	1,890	5,900	0,000
Thompson2	2001	0,387	0,111	0,663	2,748	0,006
Cahill	2003	1,550	1,144	1,957	7,476	0,000
Merrill	2003	1,810	1,564	2,056	14,428	0,000
Watson	2003	1,642	1,241	2,043	8,028	0,000
Castonguay	2004	2,440	1,336	3,545	4,329	0,000
Hardy	2005	1,329	1,091	1,567	10,958	0,000
Westbrook	2005	1,420	1,200	1,640	12,659	0,000
Wright	2005	1,817	0,983	2,651	4,268	0,000
Dimidjian	2006	1,004	0,578	1,430	4,622	0,000
Jarrett	2006	2,208	1,958	2,459	17,289	0,000
McBride	2006	2,726	2,117	3,335	8,777	0,000
Persons3	2006	1,174	0,857	1,490	7,275	0,000
Strauman	2006	1,971	1,367	2,576	6,390	0,000
Dobkin1	2007	1,115	0,599	1,630	4,240	0,000
Forman	2007	0,649	0,400	0,897	5,123	0,000
Luty	2007	1,183	0,971	1,395	10,920	0,000
Cho	2008	1,737	0,780	2,693	3,558	0,000
Constantino	2008	1,533	0,881	2,185	4,610	0,000
David	2008	2,198	1,824	2,571	11,535	0,000
Laidlaw	2008	1,231	0,791	1,671	5,484	0,000
Quilty	2008	1,733	1,377	2,088	9,554	0,000
Craigie	2009	1,870	1,584	2,156	12,825	0,000
Gibbons1	2010	0,695	0,581	0,810	11,898	0,000
Dobkin2	2011	1,366	0,883	1,849	5,543	0,000
Rieu	2011	1,065	0,519	1,610	3,827	0,000
Estupina	2012	2,095	1,603	2,586	8,356	0,000
Power	2012	0,377	-0,109	0,862	1,520	0,128
Ammermann	2013	1,183	0,746	1,621	5,301	0,000
Gibbons2	2013	2,214	1,558	2,871	6,612	0,000
Kohler	2013	1,496	1,281	1,711	13,643	0,000
Wagner	2013	1,296	0,912	1,679	6,627	0,000
Lopes	2014	1,158	0,800	1,517	6,331	0,000
Tovote	2014	0,916	0,402	1,429	3,496	0,000

Figure 2. Forest plot for the Beck Depression Inventory effect sizes.

effect indicated an even steeper decline. The size of the $beta_{time}$ coefficients varied in these analyses due to different sample sizes and correlations with the moderators.

Client-related. None of these interaction coefficients were significant: male % ($beta_{time} = -.027$; $beta_{int} = .001$, $p = .33$), age ($beta_{time} = -.031$; $beta_{int} = .0004$, $p = .58$); preintervention score BDI ($beta_{time} = -.081$; $beta_{int} = -.003$, $p = .09$), comorbidity % ($beta_{time} = -.021$; $beta_{int} = -.0005$, $p = .54$); medication % ($beta_{time} = -.019$; $beta_{int} = -.00002$, $p = .94$); patient (normal vs. special) type ($beta_{time} = -.030$; $beta_{int} = -.007$, $p = .72$); and severity (mild-moderate-severe) of depression ($beta_{time} = -.031$; $beta_{int} = .003$, $p = .87$).

Therapist-related. The single available variable, therapist (student vs. psychologist) type ($beta_{time} = -.021$; $beta_{int} = -.008$, $p = .79$), did not show a significant interaction with time.

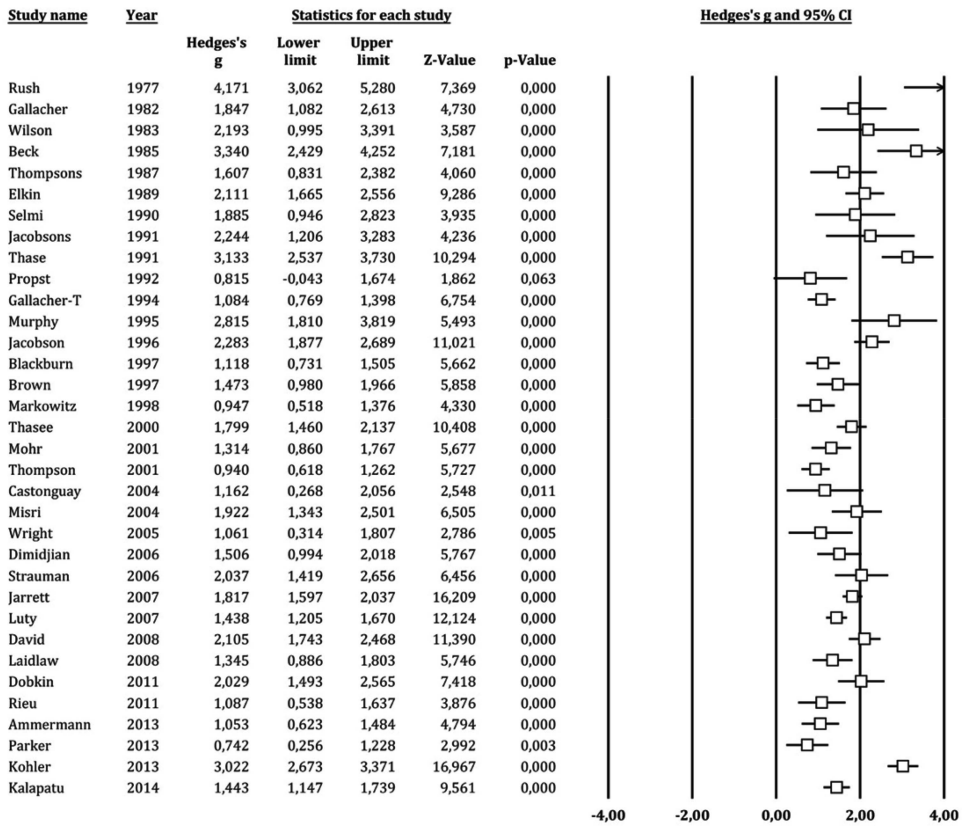


Figure 3. Forest plot for the Hamilton Rating Scale of Depression effect sizes.

Study-related. The following interaction effects were not significant: The number of sessions ($beta_{time} = -.028$; $beta_{int} = .001$, $p = .37$), methodological quality ($beta_{time} = -.032$; $beta_{int} = .001$, $p = .60$), type of statistical (ITT vs. completers) analysis ($beta_{time} = -.029$; $beta_{int} = .009$, $p = .64$), and use of the Beck manual (no vs. yes; $beta_{time} = -.033$; $beta_{int} = -.023$, $p = .14$). Although the moderator, manual use, was not significant, it is interesting to note that studies using the Beck manual showed an even steeper decline than studies that did not use it. The difference in the predicted decline of ES across a 30-year period was $g = -.023 \times 30 = -0.69$.

The final moderator, BDI-I versus BDI-II, was not significant ($beta_{time} = -.024$; $beta_{int} = .034$, $p = .33$). However, as the interaction coefficient was higher than, and inversely related to the time coefficient, this relationship was examined closer. A plot of the interaction (see Figure 10) indicated a significant decline in studies using the BDI-I measure, but not in studies using the BDI-II. The predicted treatment effect was equal for studies using the BDI-I and the BDI-II at about year 2006. Hence, the treatment effects that were observed when studies began employing the BDI-II started at about the same point in time as the effects of the BDI-I studies ended. The narrow range of publications for the studies using the BDI-II, however, restricted this comparison considerably. When the analyses were restricted to the years 1998–2014 (when the first study using the BDI-II was published), the interaction coefficient was not significant and slightly negative

Table 2
A Metaregression Analysis With Publication Year (or Time) as a Continuous Predictor of Effect Size

Studies (outcome)	*K*	b_0	b_1	95% CI	Z (b_1)	p Value
All studies (BDI)	67	60.17	−0.0293	[−0.044, −0.015]	−4.00	<.001
Within design	50	65.75	−0.0320	[−0.049, −0.015]	−3.70	<.001
CT design	17	51.96	−0.0253	[−0.050, −0.001]	−2.05	<.05
All studies (HRSD)	34	62.82	−0.0305	[−0.054, −0.007]	−2.53	.01
Within design	26	78.31	−0.0382	[−0.067, −0.009]	−2.61	<.01
CT design	8	22.37	−0.0105	[−0.042, 0.021]	−0.56	.51
Remission rate (%)	**42**	**54.43**	**−0.0271**	**[−0.047, −0.008]**	**−2.72**	**<.01**

Note. b_0 = intercept (year 0 A.D); b_1 = time slope (change coefficient); CI = confidence interval; BDI = Beck Depression Inventory; Within = the within-group condition; CT = controlled trials; HRSD = Hamilton Rating Scale of Depression.

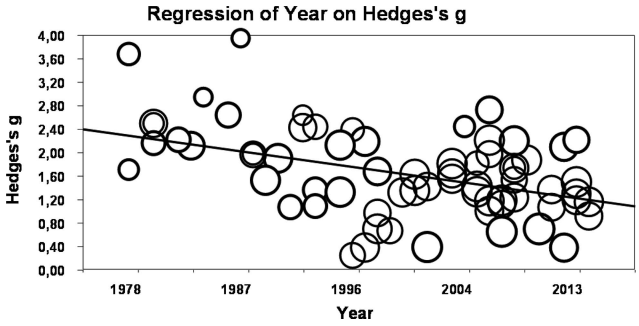


Figure 4. The plot portrays the negative change ($p < .001$) in Beck Depression Inventory effect sizes across time ($k = 61$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

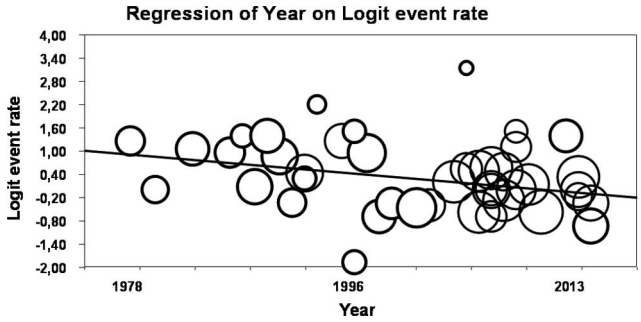


Figure 6. The plot portrays the negative change ($p = .03$) in the remission rates across time ($k = 42$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

($\beta = -.015$), indicating a negative time trend that was more pronounced for the BDI-II compared with the BDI-I studies. When the analysis was restricted to 2006–2014, when use of the BDI-II really started, the negative interaction coefficient was stronger ($\beta = -.045$).

The competence of the therapists was rated in only five studies. Moreover, one of the studies (Jacobson et al., 1996) had a CTS score that deviated considerably from the remaining four with respect to time; hence, this precluded a moderator analysis due to the low number of studies and an outlier case.

Discussion

The Temporal Trends in Treatment Effects

The main objective of the present meta-analysis was to examine the temporal changes in the effects of CBT in treating unipolar DDs. Almost all of the studies utilized the BDI to quantify the treatment's effect, whereas a smaller number of studies used the HRSD by itself or in conjunction with the BDI. The main finding was that the treatment effect of CBT showed a declining trend across time and across both measures of depression (the BDI and the HRSD). Contemporary clinical treatment trials therefore, seem to be less effective than the therapies conducted decades ago.

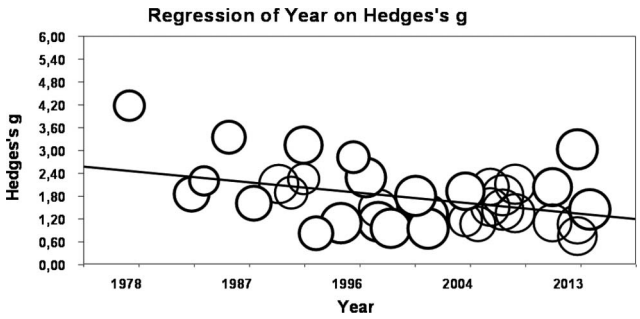


Figure 5. The plot portrays the negative Change ($p < .01$) in Hamilton Rating Scale of Depression effect sizes across time ($k = 34$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

Moreover, most of the subgroup analyses supported this conclusion. The CBT studies employing different research designs (controlled trials vs. pre-posttest within-group study designs) showed similar declining trends. Studies based on the HRSD employing a CT design also showed a similar trend, however, the decline was not significant. The small number of studies in this group ($k = 8$) reduced the statistical power considerably. Additional subgroup analyses separating the study samples (clinical trials vs. field studies), or the number of patients in the studies (low vs. high), revealed a similar downward trend. Studies consisting of potential outliers (according to the plot diagram) were also taken into account; however, the outcome was the same.

Studies employing the BDI II did not reveal a comparable decline in treatment effects. However, as these studies were almost exclusively published after 2006, restricting the time range for the BDI-II studies considerably, this comparison is of limited value. Moreover, the treatment effects of the BDI-II studies started at approximately the same time as the BDI-I effects ended. Keeping in mind that the time trend was negative for all studies from 1998 onward (see Figure 7), and that the time trend differences from 1998 were minor for these two instruments ($beta_{int} = -.015$), these findings raise no significant precautions. The timeframe of the studies using the BDI-I ranged from 1977 to 2010; hence, providing a more accurate picture of the timeframe in question. Moreover, the results of the HRSD and the remission rates for depression confirmed a significant decline in treatment effects.

The discovery of a weaker treatment effect over time cannot be explained based on a general temporal decline in patients' ability to recover from DDs, as patients on a waiting list improved in equal degrees across the entire time span. Nor can the effect be explained by lower preintervention BDI scores in the more recent studies. The correlation between the year of publication and BDI prescores was small, but positive. Two-way regression models between time (publication year) and the remaining moderators did not reveal any significant interactions, either, which would have indicated a different time trend, depending on the moderator. In summary, the declining effect of treatment over time seems robust.

Moderators of Treatment Effects

Client specific factors. The age of the patients was not related to the treatment effects, nor did it moderate the decline in treatment effects. The role of age in treatment response has yielded mixed

Table 3
A Metaregression Analysis Examining the Association Between Continuous Moderators and Effect Sizes (BDI as the Outcome)

Moderator variable	*K*	b_0	b_1	95% CI	Z (b_1)	*P* Value
Time, years '95–'02 excl.	54	65.84	−0.0320	[−0.045, −0.020]	−5.00	<.001
Time, field studies excl.	56	69.56	−0.0340	[−0.050, −0.018]	−4.10	<.001
Time, low *N* studies excl.	37	47.76	−0.0231	[−0.043, −0.003]	−2.23	.02
Time, special patients excl.	54	64.34	−0.0314	[−0.048, −0.015]	−3.76	<.001
Time, special patients	13	47.38	−0.0230	[−0.049, 0.003]	−1.76	.08
Time, waiting list	16	−9.53	0.0050	[−0.009, 0.018]	0.71	.48
Sessions	67	1.46	0.0093	[−0.021, 0.040]	0.59	.56
Age	64	2.00	−0.0103	[−0.025, 0.004]	−1.39	.17
Gender (male %)	65	1.93	−0.0104	[−0.019, −0.001]	−2.32	.03
Medication (%)	41	1.53	−0.0070	[−0.006, 0.005]	−0.25	.81
Comorbidity (%)	25	1.69	−0.0027	[−0.012, 0.006]	−0.57	.56
Study quality (0–48)	67	1.84	−0.0085	[−0.031, 0.014]	−0.75	.45
Therapist competency (0–72)	**5**	**0.25**	**0.0253**	**[−0.026, 0.076]**	**0.97**	**.33**

Note. BDI = Beck Depression Inventory; b_0 = intercept (year 0 A.D.); b_1 = time slope (change coefficient); CI = confidence interval; Time = publication year; excl. = excluded.

findings in the clinical literature (e.g., Ammerman, Peugh, Putnam, & Van Ginkel, 2012; Lewis, Simons, & Kim, 2012), which the present analysis confirmed.

A significant gender difference was evident, indicating that women profited more from CBT for depression than did men. This was somewhat surprising, given that previous studies (Joutsenniemi et al., 2012; Wierzbicki & Pekarik, 1993) have indicated no sex differences with regard to who benefits the most from psychotherapy. We have no interpretation for this finding, but as women represent the majority of those being treated for depression, this difference means that overall, more patients improve following CBT. However, if the *p* value had been adjusted due to multiple significance testing, this difference would not have been significant.

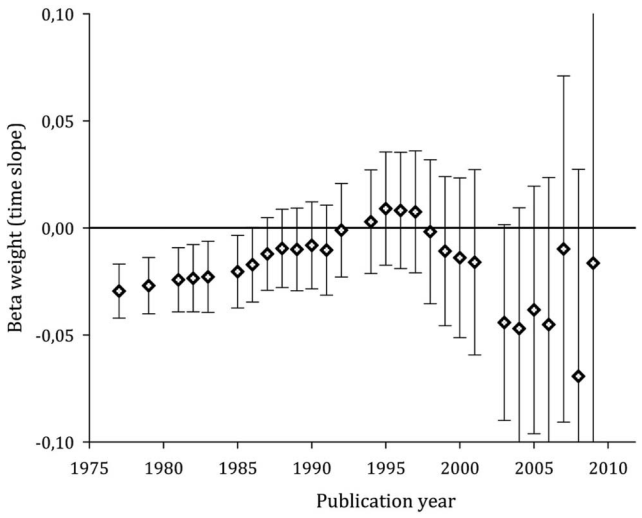


Figure 7. Temporal changes depending on the publication year start. Coefficients below 0 indicate a declining effect if estimated from the publication year as indicated on the *x* axis. The 95% error bars are increasing due to a lower number of available studies when advancing the publication year start.

The degree of comorbidity did not moderate the reported ESs, nor did it interact with time. One may thus, exclude the possibility that the declining effect of CBT is because recent studies have included patients with a higher degree of psychiatric comorbidity. An often-used strategy in clinical research is to implement new treatments on highly selected samples (comorbid conditions are excluded), that use highly trained or competent therapists who implement therapy according to a treatment manual. Such clinical trials are referred to as efficacy trials, whereas trials that are not as strict in these requirements are known as effectiveness trials. The latter include patients with varying degrees of comorbidity and/or therapist competence, which better reflect the reality of how mental health services are delivered. Therefore, one could expect that the more recent CBT trials had an overrepresentation of effectiveness trials than the previous ones. However, the situation seems to be going in the opposite direction, as the more recent studies included fewer patients with comorbidity. The declining trend in treatment effects over time was not moderated by therapist experience either. Hence, any strong objections against the present meta-analysis for not controlling for different types of implementations, efficacy versus effectiveness, seem less relevant.

The percentage of patients on stable dosages of psychotropic medication, including antidepressants, did not covary with ES. This finding is somewhat surprising, given that several studies and meta-analyses have indicated a higher treatment effect when psychotherapy was combined with antidepressants (e.g., de Maat et al., 2008; Keller et al., 2000; Pampallona et al., 2004). The meta-analysis of Cuijpers et al. (2009), comparing psychotherapy in general, with psychotherapy plus medication, and with seven trials of CBT and CBT plus medication, indicated a similar trend. The advantage of CBT plus medication was, however, small. One explanation for the lack of confirmatory findings here, may be that our study recorded a continuous percentage score of the number of patients on medication, and hence, it did not compare two distinctively defined patient groups (i.e., 100% pure CBT compared with 100% CBT+ medication), which other studies have done. This particular moderator analysis, therefore, may have been statistically underpowered. Another explanation may be related to the characteristics of the clinical samples, as most of the studies

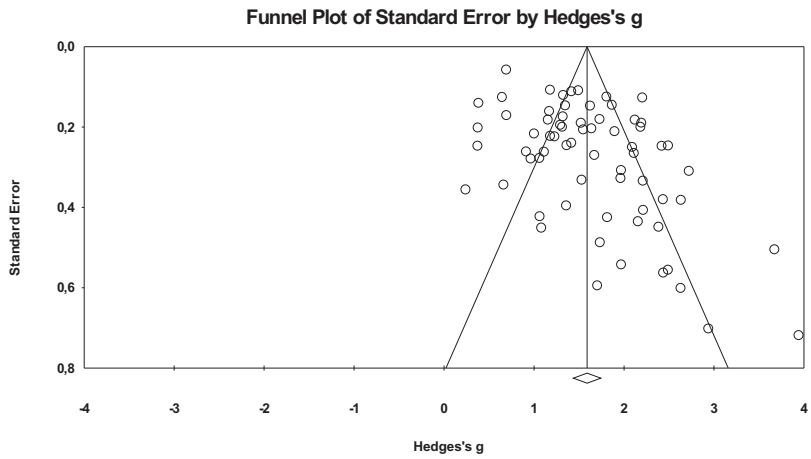


Figure 8. Funnel plot of the 67 included studies based on the Beck Depression Inventory.

sampled patients with a moderate degree of depression. It is conceivable that psychotherapy combined with medication has a higher treatment effect mainly for the severely depressed patients, as indicated by the American Psychiatric Association's guidelines for the treatment of depression (APA, 2010).

Although different diagnostic classifications of depression as mild, moderate, severe, or recurrent did not yield statistically significant effect differences (potentially due to the small number of studies), the differences were nevertheless meaningful. The highest treatment effects were seen in patients with recurrent depression. This result seems reasonable given that the diagnostic criteria for recurrent depression imply that remission is achieved between depressive episodes. These patients have a longer treatment history than those depressed for the first time; they know the rationale for CBT, and what to expect from therapy. They also are more acquainted with the methodological approaches, such as the importance of constructing a case conceptualization that the homework tasks are designed to test. These patients may also have more knowledge about how to find a skilled therapist, and thus, experience a stronger or quicker effect.

Although our study did not reveal any significant differences in ES related to samples with special characteristics, a tendency for a higher ES was found in ordinary patient populations ($g = 1.64$ vs. 1.35). This tendency is not surprising, given the fact that comorbidity, in general, is connected with poorer outcomes of therapy. However, the negative time trend was not affected by the inclusion of special patient samples. Rather, the trend was negative irrespective of the sample's patient characteristics (ordinary vs. special patient subpopulations). Restricting the time-trend analysis to the special patient group revealed a similar decline in treatment effects, albeit, not significant, probably due to the small number of studies.

Therapist-related factors. The competence of the therapist probably exerts more influence on how treatment works (Simons et al., 2010), which the present meta-analysis partly suggests: patients receiving CBT from experienced psychologists had a more pronounced reduction in depressive symptoms compared with patients receiving CBT from psychology students, with less experience doing therapy. The difference represented half of a standard deviation, which is considered a moderate effect size

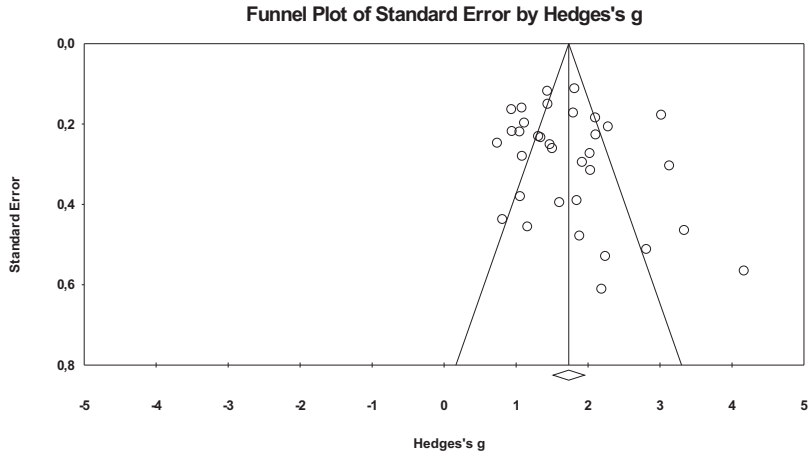


Figure 9. Funnel plot of the 34 included studies based on the Hamilton Rating Scale of Depression.

Table 4
A Subgroup Analysis of Dichotomous Variables and Effect Size Based on the BDI

Moderator	*k*	*g*	95% CI	Q_{df}	*p* Value	I^2
Diagnostic severity				3.10_3	.38	.89
Mild	9	1.28	[0.84, 1.71]			.90
Moderate	40	1.62	[1.42, 1.82]			.88
Severe	10	1.56	[1.18, 1.97]			.88
Recurrent	8	1.86	[1.33, 2.39]			.92
Data analysis				1.89_1	.17	.89
ITT	18	1.43	[1.18, 1.69]			.89
Completers	49	1.66	[1.46, 1.85]			.89
Beck manual				0.02_1	.89	.89
Yes	38	1.60	[1.39, 1.81]			.89
No	29	1.58	[1.36, 1.80]			.89
Adherence check				0.02_1	.89	.90
Yes	32	1.56	[1.35, 1.78]			.87
No	30	1.54	[1.32, 1.77]			.91
Patient type				2.54_1	.11	.89
Ordinary	54	1.64	[1.47, 1.81]			.90
Special	13	1.35	[1.03, 1.67]			.72
Therapist*				7.14_1	<.01	.85
Trained student	7	0.98	[0.59, 1.36]			.65
Psychologist	**37**	**1.55**	**[1.38, 1.72]**			**.83**

Note. BDI = Beck Depression Inventory; CI = confidence interval; Q_{df} = Q value for the between group difference(s); *df* = associated degrees of freedom; I^2 = I-squared indicates the degree of between study variance relative to total variance; ITT = intention to treat.
* Psychiatrist was not included due to few studies ($k = 2$). The remaining studies ($k = 21$) used a combination of therapists, or type of therapist was not reported. These studies were excluded from this analysis.

difference in statistical terms. Such differences may be of clinical concern as half a standard deviation on the BDI instrument typically represents a 5-point decrease in the raw score (Dworkin et al., 2008). As most CBT studies have been conducted with patients with moderate degrees of depression (BDI scores ranging between 20 and 29, and an expected mean of 25), about one third of the patients would thus, be expected to shift from the moderate to the mild diagnostic category. This represents a non-negligible difference that needs to be taken into account when assigning patients to available therapists in a clinic. The most competent therapist should be assigned to the most depressed patients. It is, of course, important that students are trained to conduct CBT, but student therapy should be offered to patients with primarily mild, or at the maximum, a moderate degree of depression.

In addition, there was a tendency (albeit a tentative one) indicating that therapist competence, as measured by the CTS, implied better treatment effects. However, this relationship was not significant. As the present meta-analysis only identified five studies reporting sufficient data, the statistical power and the possibility of generalization from these studies were low. Nevertheless, the direction of the effect concurred with the common finding that therapists who are more competent help their patients achieve remission more quickly (Stein & Lambert, 1995; Strunk, Brotman, DeRubeis, & Hollon, 2010). Yet again, variation in competence was unrelated to the reported time trend.

Specific treatment or study quality related factors. The number of therapy sessions did not reveal different treatment effects following CBT. A caveat should be noted as most of the studies consisted of interventions consisting of between 10 and 20 sessions of psychotherapy, hence, precluding any conclusions regarding further improvement (or deterioration) beyond 20 sessions of therapy. This hardly represents a limitation of the analyses, as CBT for depression is designed as a short-term therapy. The weighted mean number of therapy sessions was 14.8 ($SD = 5.2$). Because we did not find support for an inverse U-shaped relationship between treatment effects and number of sessions, length of therapy seems to be less important for efficacy.

We did not find evidence of significant differences in the treatment effects resulting from the use of the Beck manual (Beck et al., 1979). Contrary to expectations, the interaction analyses showed a slightly steeper decline for the CBT trials that used the manual compared to those that did not. This finding was rather surprising given that the original manual had a reputation among clinical researchers as one of the best ways to implement CBT. We cannot conceive of any sensible explanation for why clinical studies using the Beck manual fare relatively worse than those not using it. To the best of our knowledge, there have been no thorough investigations of how different ways of conducting CBT for depression may influence the outcome. Our findings indicate that further investigations regarding this matter are warranted.

This study revealed no differences in ES related to the utilization of adherence checks. This finding is at odds with the perceived importance of adhering to a treatment manual (Crits-Christoph et al., 1991; Shafran et al., 2009). One explanation may be that most therapists in the included studies were well-trained or experienced psychologists, and thus, likely to conduct CBT in a proper fashion even without checks or feedback regarding adherence to the manual. Another possibility is that adherence checks were not reported consistently.

The methodological quality of the studies was rated with the RCT-PQRS published by Kocsis et al. (2010). It is a comprehensive measure of the methodological quality of clinical trials (Gerber et al., 2011). Many of the items are derived from preexisting measures of the quality of randomized controlled trials. An advantage of the PCT-PQRS is that it was developed to fit different

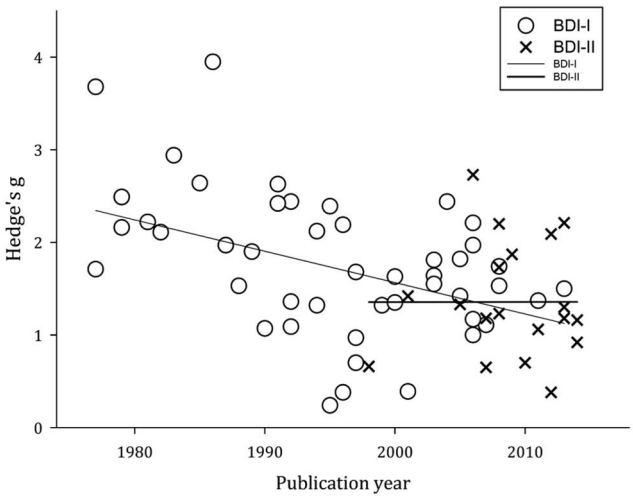


Figure 10. A plot of the interaction between publication year and type of Beck Depression Inventory (BDI) measure used.

therapy traditions (e.g., CBT, psychodynamic therapy, or pharmacology). The quality ratings have improved considerably over the years; newer studies have received much higher quality ratings than the older ones. Although the quality ratings were not significantly related with the ESs, the relationship was, nevertheless, in the expected direction, as higher quality studies yielded slightly lower therapy effects than the lower quality studies. We also observed lower effect sizes in CBT studies using CT versus within-group research designs, although they were, yet again, nonsignificant. As both methodological quality indicators pointed in the same direction, the present findings are in line with previous meta-analyses (e.g., Gould, Coulson, & Howard, 2012; Pallesen et al., 2005).

The present analysis did not reveal a significant difference in the ES between the statistical designs completers versus ITT. We did, however, replicate the tendency observed in Hans and Hiller's (2013) meta-analysis, and found a slightly larger ES for completers ($g = 1.66$) versus ITT ($g = 1.43$). This rather modest difference probably is due to the larger ratio of the early drop-outs from the ITT design, thus, preventing these patients from benefitting from all of the components of the CBT intervention.

As the number of studies reporting data related to common factors, such as the patient–therapist alliance, was extremely low, no conclusions about common factors could be drawn.

Potential Reasons for a Decline in Therapy Effect

The original manual for how to deliver and implement CBT was developed in the 1970s, and subsequently, served as the gold standard for many practitioners of psychotherapy. The reason for the declining effect is hard to explain beyond the fact that CBT for depression has not led to systematic improvements.

It is possible that the ostensibly simple treatment objective of CBT (i.e., changing maladaptive cognitions to alleviate emotional disorders), has made it particularly attractive and has created a misconception of being easy to learn. However, proper training, considerable practice, and competent supervision are very important to provide CBT in an efficacious manner. Thus, clinical researchers have warned against deviating from the evidence-based therapeutic interventions (Shafran et al., 2009), as therapists who frequently depart from the manual demonstrate poorer treatment effects than therapists who follow the manual (Luborsky et al., 1997, 1985). The lack of a stronger treatment effect among studies employing the Beck manual in the present meta-analysis does not invalidate this recommendation, as the studies that did not explicitly state that the manual was used may still have used skilled therapists that properly implemented CBT.

Another possibility is that the degree of experience or therapeutic competence may affect treatment outcomes differently, depending on whether a CBT manual is followed or not (Crits-Christoph et al., 1991). This interaction was not possible to address in our analysis. From a CBT point of view, it may be realistic to expect that the original founders of the therapy may have been more concerned with therapy fidelity (strong adherence to the manual) and with acquiring a large amount of experience with the method before examining it in a randomized clinical trial. There has been a tendency to publish clinical trials based on CBT without properly describing the contents of the treatment given, which may indicate less concern with adherence to the manual. Although this is a possibility, the interaction effect would need to be quite strong for the declining slope to be nonsignificant, and even stronger to shift the slope to a positive direction, which is highly unlikely.

Standardization of the data collected from clinical trials may be helpful for future reviews of CBT, in order to avoid missing important moderator data, and be able to conduct more nuanced analyses in the future. Future trials should include measures of the therapeutic alliance and therapist competence, as well as an adequate description of what was done during the therapy sessions, and how it was done and when it was done. A minimum set of data related to client factors, therapist factors, as well as common and specific factors should be collected.

An interesting confounder related to the common factors should be mentioned: the placebo effect. The placebo effect is typically stronger for newer treatments, however, as time passes and experience with therapy is gained, the strong initial expectations wane. One may question whether this is the case with CBT. In the initial phase of the cognitive era, CBT was frequently portrayed as the gold standard for the treatment of many disorders. In recent times, however, an increasing number of studies (e.g., Baardseth et al., 2013; Wampold et al., 2002, 1997) have not found this method to be superior to other techniques. Coupled with the increasing availability of such information to the public, including the Internet, it is not inconceivable that patients' hope and faith in the efficacy of CBT has decreased somewhat, in recent decades. Moreover, whether widespread knowledge of the present meta-analysis results might worsen the situation, remains an open question.

If technical factors represent 10%–20% of the total treatment effect, it seems reasonable to suggest that newer psychotherapy approaches should diligently address improvements in the common factors to realize larger treatment effects. In this respect, it seems strange that CBT apparently reached a ceiling effect during its first few years.

Limitations

The present meta-analysis is not without limitations. First, this study only included depression, thus, excluding CBT trials aimed at treating other diagnosis, such as anxiety, posttraumatic stress, eating, schizophrenia, and sleep disorders. There is no reason to expect the present findings to generalize to these disorders. In particular, anxiety disorders, which include a heterogeneous group of disorders that probably yield different time trends, have been subjected to the CBT approach. The clinical presentations of, for example, panic, obsessive–compulsive, and posttraumatic stress disorders are very different, as are the CBT approaches used. A meta-analysis of five trials comparing cognitive therapy with exposure therapy to treat obsessive–compulsive disorder (Ougrin, 2011) did not indicate a decline for the newer trials. Another review examining the efficacy of 12 trials examining transdiagnostic CBT in treating common anxiety disorders, such as obsessive–compulsive, generalized, and social anxiety disorder (Reinholt & Krogh, 2014), indicated no temporal changes either. A study by Hofmann and Smits (2008), that we will finally mention, examined the efficacy of 25 clinical trials on the use of CBT for the treatment of anxiety disorders even showed a minor positive

temporal change. These examples indicate that a comprehensive meta-analysis covering other mental health disorders may yield quite different results.

The BDI has undergone some modifications during its 40-year existence. The original BDI was revised and made more user friendly in 1988, and given the acronym, BDI-Ia (Beck et al., 1988). The latest version, the BDI-II, has incorporated an item measuring hypochondriasis, changed the timeframe of symptoms from 1 week to 2 weeks, and put more emphasis on measuring all diagnostic criteria related to depression. Still, the forms are very similar to each other (Beck et al., 1996). Despite these differences, the treatment and control groups responded to the equivalent forms at any point in time. Thus, these considerations should not pose major threats to the validity of the current conclusions.

Very few studies ($k = 5$) included correlations between the BDI pre- and postintervention scores, requiring us to impute this value for the remaining 65 studies. However, the potential for this value to exert undue influence on the results does seem small for two reasons. First, the variations in correlations need to be quite high in order to change the ESs substantially. Second, and most importantly, we have no reason to expect that the prepost BDI correlations should change considerably over time. Although a shift in therapy effect over the years changed the mean of the post intervention BDI scores, the relative position between the pre- and postscores should not have changed by much.

Recovery rates were calculated according to somewhat varying criteria across the studies included in this analysis. The most stringent criterion was a cut-off score for clinical depression of 7 on the BDI, while the most liberal was 10. Although this difference might not seem substantial, it could have a confounding effect on the calculated total percentage of recovered patients, and the correlation between recovery rates and year of intervention.

A minor possible caveat relates to the time moderator. As all of the studies' years were coded based on their publication dates, it is conceivable that this date could vary somewhat from the actual year of the intervention. However, it is reasonable to assume that this discrepancy is similar to contemporary and older studies, and that the difference between the publication and actual year of intervention is not very large.

Implications

The practical significance of this study is to heighten the awareness among practitioners and clinical researchers of the trends in modern psychotherapy. If the psychotherapy of today has a lower efficacy than that conducted 30 to 40 years ago, this threatens the validity of current comparative studies. If we compare the efficacy of a new psychotherapeutic approach with the current best standard, which, for example, may be CBT, we risk concluding that the newer approach is preferable even though it may have a weaker effect than the seminal CBT trials of the 1970s. Researchers conducting randomized placebo-controlled trials today, thus, risk keeping newer treatment approaches that are relatively better than the current best CBT. Yet, what is the benefit of doing so if the absolute change is minor or even negative compared to the seminal studies?

The fact that individual cognitive therapy demonstrates a declining temporal trend implies, however, that the possibility of significant improvement exists. Treatment outcomes may be improved, not only through technical variations or new additions, but also by considering better ways of integrating common, therapist, and patient-related factors. Further research and randomized trials that include measures of the four major variance components underpinning the therapy's effects are recommended to determine the formula behind the optimal practice of CBT. All future clinical trials should be conducted according to a common standard that prescribes which information should be collected, at a minimum, in all psychotherapy studies.

References

References marked with an asterisk indicate studies that are included in the meta-analysis.

Abbass, A. A., Rabung, S., Leichsenring, F., Refseth, J. S., & Midgley, N. (2013). Psychodynamic psychotherapy for children and adolescents: A meta-analysis of short-term psychodynamic models. *Journal of the American Academy of Child and Adolescent Psychiatry, 52,* 863–875. http://dx.doi.org/10.1016/j.jaac.2013.05.014

American Psychiatric Association. (2010). *Practice guidelines for the treatment of patients with major depressive disorder* (3rd ed.). Arlington, VA: APA.

Ammerman, R. T., Peugh, J. L., Putnam, F. W., & Van Ginkel, J. B. (2012). Predictors of treatment response in depressed mothers receiving in-home cognitive-behavioral therapy and concurrent home visiting. *Behavior Modification, 36,* 462–481. http://dx.doi.org/10.1177/0145445512447120

*Ammerman, R. T., Putnam, F. W., Altaye, M., Stevens, J., Teeters, A. R., & Van Ginkel, J. B. (2013). A clinical trial of in-home CBT for depressed mothers in home visitation. *Behavior Therapy, 44,* 359–372. http://dx.doi.org/10.1016/j.beth.2013.01.002

Ashouri, A., Atef Vahid, M. K., Gharaee, B., & Rasoulian, M. (2013). Effectiveness of meta-cognitive and cognitive-behavioral therapy in patients with major depressive disorder. *Iranian Journal of Psychiatry and Behavioral Sciences, 7,* 24–34.

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., . . . Wampold, B. E. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review, 33,* 395–405. http://dx.doi.org/10.1016/j.cpr.2013.01.004

*Beach, S. R. H., & O'Leary, D. K. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23,* 507–528. http://dx.doi.org/10.1016/S0005-7894(05)80219-9

Beck, A. T. (1970). Cognitive therapy: Nature and relation to behavior therapy. *Behavior Therapy, 1,* 184–200. http://dx.doi.org/10.1016/S0005-7894(70)80030-2

*Beck, A. T., Hollon, S. D., Young, J. E., Bedrosian, R. C., & Budenz, D. (1985). Treatment of depression with cognitive therapy and amitriptyline. *Archives of General Psychiatry, 42,* 142–148. http://dx.doi.org/10.1001/archpsyc.1985.01790250036005

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. (1996). Comparison of beck depression inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588–597. http://dx.doi.org/10.1207/s15327752jpa6703_13

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II, Beck depression Inventory: Manual.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R., & Carbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100. http://dx.doi.org/10.1016/0272-7358(88)90050-5

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571. http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

Bienenfeld, D. (2007). Cognitive therapy of patients with personality disorders. *Psychiatric Annals, 37,* 133–139.

*Blackburn, I. M., & Moore, R. G. (1997). Controlled acute and follow-up trial of cognitive therapy and pharmacotherapy in out-patients with recurrent depression. *The British Journal of Psychiatry, 171,* 328–334. http://dx.doi.org/10.1192/bjp.171.4.328

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. New York, NY: Wiley. http://dx.doi.org/10.1002/9780470743386

*Brown, R. A., Evans, D. M., Miller, I. W., Burgess, E. S., & Mueller, T. I. (1997). Cognitive-behavioral treatment for depression in alcoholism. *Journal of Consulting and Clinical Psychology, 65,* 715–726. http://dx.doi.org/10.1037/0022-006X.65.5.715

*Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcomes of patients completing and not completing cognitive therapy for depression. *British Journal of Clinical Psychology, 42,* 133–143. http://dx.doi.org/10.1348/014466503321903553

*Carrington, C. H. (1979). A comparison of cognitive and analytically oriented brief treatment approaches to depression in Black women. *Dissertation Abstracts International, 40,* 28–29.

Castonguay, L. G. (1996). *Integrative cognitive therapy for depression treatment manual*. Unpublished manuscript, The Pennsylvania State University.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology, 64,* 497–504. http://dx.doi.org/10.1037/0022-006X.64.3.497

*Castonguay, L. G., Schut, A. J., Aikens, D. E., Constantino, M. J., Laurenceau, J-P., Bologh, L., & Burns, D. D. (2004). Integrative cognitive therapy for depression: A preliminary investigation. *Journal of Psychotherapy Integration, 14,* 4–20.

*Cho, H. J., Kwon, J. H., & Lee, J. J. (2008). Antenatal cognitive-behavioral therapy for prevention of postpartum depression: A pilot study. *Yonsei Medical Journal, 49,* 553–562.

Cirstoiu, C., Circota, G., Panaitescu, C., & Niculaita, R. (2011). The advantage of arthroscopic anterior cruciate ligament reconstruction with autograft from the tendons of the semitendinosus-gracilis muscles for the recovery of the stability of the knee. *Maedica, 6,* 109–113.

*Constantino, M. J., Marnell, M. E., Haile, A. J., Kanther-Sista, S. N., Wolman, K., Zappert, L., & Arnow, B. A. (2008). Integrative cognitive therapy for depression: A randomized pilot comparison. *Psychotherapy, 45,* 122–134. http://dx.doi.org/10.1037/0033-3204.45.2.122

*Craigie, M. A., & Nathan, P. (2009). A nonrandomized effectiveness comparison of broad-spectrum group CBT to individual CBT for depressed outpatients in a community mental health setting. *Behavior Therapy, 40,* 302–314. http://dx.doi.org/10.1016/j.beth.2008.08.002

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., . . . Ztrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91. http://dx.doi.org/10.1080/10503309112331335511

Cuijpers, P., Dekker, J., Hollon, S. D., & Andersson, G. (2009). Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: A meta-analysis. *Journal of Clinical Psychiatry, 70,* 1219–1229. http://dx.doi.org/10.4088/JCP.09r05021

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. http://dx.doi.org/10.1037/a0013075

*David, D., Szentagotai, A., Lupu, V., & Cosman, D. (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: A randomized clinical trial, posttreatment outcomes, and six-month follow-up. *Journal of Clinical Psychology, 64,* 728–746. http://dx.doi.org/10.1002/jclp.20487

de Maat, S., Dekker, J., Schoevers, R., van Aalst, G., Gijsbers-van Wijk, C., Hendriksen, M., . . . de Jonghe, F. (2008). Short psychodynamic supportive psychotherapy, antidepressants, and their combination in the treatment of major depression: A mega-analysis based on three randomized clinical trials. *Depression and Anxiety, 25,* 565–574. http://dx.doi.org/10.1002/da.20305

*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,* 658–670. http://dx.doi.org/10.1037/0022-006X.74.4.658

*Dobkin, R. D., Allen, L. A., & Menza, M. (2007). Cognitive-behavioral therapy for depression in Parkinson's disease: A pilot study. *Movement Disorders, 22,* 946–952. http://dx.doi.org/10.1002/mds.21455

*Dobkin, R. D., Menza, M., Allen, L. A., Gara, M. A., Mark, M. H., Tiu, J., . . . Friedman, J. (2011). Cognitive-behavioral therapy for depression in Parkinson's disease: A randomized, controlled trial. *The American Journal of Psychiatry, 168,* 1066–1074. http://dx.doi.org/10.1176/appi.ajp.2011.10111669

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57,* 414–419. http://dx.doi.org/10.1037/0022-006X.57.3.414

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,* 295–300. http://dx.doi.org/10.1111/j.2044-8260.1985.tb00662.x

Duncan, B. L., Miller, S. D., & Sparks, J. (2004). *The heroic client. A revolutionary way to improve effectiveness through client-directed, outcome informed Therapy*. San Francisco, CA: Jossey-Bass.

*Dunn, R. J. (1979). Cognitive modification with depression-prone psychiatric patients. *Cognitive Therapy and Research, 3,* 307–317. http://dx.doi.org/10.1007/BF01185971

Dworkin, R. H., Turk, D. C., Wyrwich, K. W., Beaton, D., Cleeland, C. S., Farrar, J. T., . . . Zavisic, S. (2008). Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The Journal of Pain, 9,* 105–121. http://dx.doi.org/10.1016/j.jpain.2007.09.005

*Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982. http://dx.doi.org/10.1001/archpsyc.1989.01810110013002

Ellis, A. (1962). *Reason and emotion in psychotherapy*. New York, NY: Stuart.

*Emanuels-Zuurveen, L., & Emmelkamp, P. M. (1996). Individual behavioural-cognitive therapy v. marital therapy for depression in maritally distressed couples. *The British Journal of Psychiatry, 169,* 181–188. http://dx.doi.org/10.1192/bjp.169.2.181

*Emanuels-Zuurveen, L., & Emmelkamp, P. M. (1997). Spouse-aided therapy with depressed patients. *Behavior Modification, 21,* 62–77. http://dx.doi.org/10.1177/01454455970211003

*Estupina Puig, F. J., & Labrador Encinas, F. J. (2012). Effectiveness of cognitive-behavioral treatment for major depressive disorder in a university psychology clinic. *The Spanish Journal of Psychology, 15,* 1388–1399. http://dx.doi.org/10.5209/rev_SJOP.2012.v15.n3.39423

Everyday Health. (2013). *Stats and facts about depression in America*. Retrieved from http://www.everydayhealth.com/health-report/major-depression/depression-statistics.aspx

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41,* 509–528. http://dx.doi.org/10.1016/S0005-7967(02)00088-8

Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics, 2,* 105–124. http://dx.doi.org/10.1207/S15328031US0202_02

*Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31,* 772–799. http://dx.doi.org/10.1177/0145445507302202

Friborg, O., Martinsen, E. W., Martinussen, M., Kaiser, S., Overgård, K. T., & Rosenvinge, J. H. (2014). Comorbidity of personality disorders in mood disorders: A meta-analytic review of 122 studies from 1988 to 2010. *Journal of Affective Disorders,* 152–154, 1–11. http://dx.doi.org/10.1016/j.jad.2013.08.023

*Gallagher, D. E., & Thompson, L. W. (1982). Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. *Psychotherapy: Theory, Research, & Practice, 19,* 482–490. http://dx.doi.org/10.1037/h0088461

*Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62,* 543–549. http://dx.doi.org/10.1037/0022-006X.62.3.543

Garfield, S. L. (1986). Research on client variables in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 213–256). New York, NY: Wiley.

Gerber, A. J., Kocsis, J. H., Milrod, B. L., Roose, S. P., Barber, J. P., Thase, M. E., . . . Leon, A. C. (2011). A quality-based review of randomized controlled trials of psychodynamic psychotherapy. *The American Journal of Psychiatry, 168,* 19–28. http://dx.doi.org/10.1176/appi.ajp.2010.08060843

*Gibbons, C. J., Fournier, J. C., Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Beck, A. T. (2010). The clinical effectiveness of cognitive therapy for depression in an outpatient clinic. *Journal of Affective Disorders, 125,* 169–176. http://dx.doi.org/10.1016/j.jad.2009.12.030

*Gibbons, C. R., Stirman, S. W., Derubeis, R. J., Newman, C. F., & Beck, A. T. (2013). Research setting versus clinic setting: Which produces better outcomes in cognitive therapy for depression? *Cognitive Therapy and Research, 37,* 605–612. http://dx.doi.org/10.1007/s10608-012-9499-7

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49,* 59–72. http://dx.doi.org/10.1016/S0165-0327(97)00199-7

Gould, R. L., Coulson, M. C., & Howard, R. J. (2012). Cognitive behavioral therapy for depression in older people: A meta-analysis and meta-regression of randomized controlled trials. *Journal of the American Geriatrics Society, 60,* 1817–1830. http://dx.doi.org/10.1111/j.1532-5415.2012.04166.x

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–62. http://dx.doi.org/10.1136/jnnp.23.1.56

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81,* 75–88. http://dx.doi.org/10.1037/a0031080

*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,* 59–67. http://dx.doi.org/10.1037/0022-006X.73.1.59

Healthline. (2012). *Unhappiness by the numbers: 2012 depression statistics*. Retrieved from http://www.healthline.com/health/depression/statistics-infographic

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Hedlund, J. L., & Vieweg, B. W. (1979). The Hamilton rating scale for depression: A comprehensive review. *Journal of Operational Psychiatry, 10,* 149–165.

Hoffart, A., Borge, F.-M., Sexton, H., & Clark, D. M. (2009). The role of common factors in residential cognitive and interpersonal therapy for social phobia: A process-outcome study. *Psychotherapy Research, 19,* 54–67.

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36,* 427–440. http://dx.doi.org/10.1007/s10608-012-9476-1

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

*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,* 774–781. http://dx.doi.org/10.1001/archpsyc.1992.01820100018004

Hollon, S. D., Shelton, R. C., & Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology, 59,* 88–99. http://dx.doi.org/10.1037/0022-006X.59.1.88

Honyashiki, M., Furukawa, T. A., Noma, H., Tanaka, S., Chen, P., Ichikawa, K., . . . Caldwell, D. M. (2014). Specificity of CBT for depression: A contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research, 38,* 249–260. http://dx.doi.org/10.1007/s10608-014-9599-7

Horvath, A. O., & Greenberg, L. S. (1986). The development of the working alliance inventory. In L. S. Greenberg & W. M. Pinsoff (Eds.), *The Psychotherapeutic Process: A research handbook* (pp. 529–556). New York, NY: Guilford Press.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41,* 159–164. http://dx.doi.org/10.1037/0003-066X.41.2.159

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69,* 747–755. http://dx.doi.org/10.1037/0022-006X.69.5.747

Imel, Z. E., & Wampold, B. E. (2008). The importance of treatment and the science of common factors in psychotherapy. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 249–262). New York, NY: Wiley

*Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmaling, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. *Journal of Consulting and Clinical Psychology, 59,* 547–557. http://dx.doi.org/10.1037/0022-006X.59.4.547

*Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64,* 295–304.

*Jarrett, R. B., Vittengl, J. R., Doyle, K., & Clark, L. A. (2007). Changes in cognitive content during and following cognitive therapy for recurrent depression: Substantial and enduring, but not predictive of change in

depressive symptoms. *Journal of Consulting and Clinical Psychology, 75,* 432–446.

Joutsenniemi, K., Laaksonen, M. A., Knekt, P., Haaramo, P., & Lindfors, O. (2012). Prediction of the outcome of short- and long-term psychotherapy based on socio-demographic factors. *Journal of Affective Disorders, 141,* 331–342. http://dx.doi.org/10.1016/j.jad.2012.03.027

*Kalapatapu, R. K., Ho, J., Cai, X., Vinogradov, S., Batki, S. L., & Mohr, D. C. (2014). Cognitive-behavioral therapy in depressed primary care patients with co-occurring problematic alcohol use: Effect of telephone-administered vs. face-to-face treatment-a secondary analysis. *Journal of Psychoactive Drugs, 46,* 85–92. http://dx.doi.org/10.1080/02791072.2013.876521

Kegel, A. F., & Flückiger, C. (2014). Predicting psychotherapy dropouts: A multilevel approach. *Clinical Psychology & Psychotherapy*. Advance online publication. http://dx.doi.org/10.1002/cpp.1899

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., . . . Zajecka, J. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *The New England Journal of Medicine, 342,* 1462–1470. http://dx.doi.org/10.1056/NEJM200005183422001

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11,* 289–299. http://dx.doi.org/10.1007/BF01186280

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). *Journal of the American Medical Association, 289,* 3095–3105. http://dx.doi.org/10.1001/jama.289.23.3095

*King, M., Sibbald, B., Ward, E., Bower, P., Lloyd, M., Gabbay, M., & Byford, S. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment, 4,* 1–83.

Kishi, T., Kafantaris, V., Sunday, S., Sheridan, E. M., & Correll, C. U. (2012). Are antipsychotics effective for the treatment of anorexia nervosa? Results from a systematic review and meta-analysis. *Journal of Clinical Psychiatry, 73,* e757–e766. http://dx.doi.org/10.4088/JCP.12r07691

Knapik, M. L. (2012). Transradial arterial access for cath and PCI, and the impact on hospital bottom lines. *Cath Lab Digest, 20,* 44–48.

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., . . . Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry, 51,* 319–324. http://dx.doi.org/10.1016/j.comppsych.2009.07.001

*Kohler, S., Hoffmann, S., Unger, T., Steinacher, B., Dierstein, N., & Fydrich, T. (2013). Effectiveness of cognitive-behavioural therapy plus pharmacotherapy in inpatient treatment of depressive disorders. *Clinical Psychology & Psychotherapy, 20,* 97–106. http://dx.doi.org/10.1002/cpp.795

Kohler, S., Unger, T., Hoffmann, S., Steinacher, B., & Fydrich, T. (2013). Acute and long-term treatment outcome in depressed inpatients with vs. without anxious features: Results of a one-year follow-up study. *Journal of Affective Disorders, 150,* 1055–1061. http://dx.doi.org/10.1016/j.jad.2013.05.043

Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. (2002). *Meta-analysis: A guide to calibrating and combining statistical evidence*. Hoboken, NJ: Wiley.

*Laidlaw, K., Davidson, K., Toner, H., Jackson, G., Clark, S., Law, J., . . . Cross, S. (2008). A randomised controlled trial of cognitive behaviour therapy vs. treatment as usual in the treatment of mild to moderate late life depression. *International Journal of Geriatric Psychiatry, 23,* 843–850. http://dx.doi.org/10.1002/gps.1993

Lambert, M. J. (1992). Psychotherapy outcome research. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of Psychotherapy Integration*. New York, NY: Basic.

Lambert, M. J. (2001). The status of empirically supported therapies: Comment on Westen and Morrison's (2001). multidimensional meta-analysis. *Journal of Consulting and Clinical Psychology, 69,* 910–913. http://dx.doi.org/10.1037/0022-006X.69.6.910

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 143–198). New York, NY: Wiley.

Lewis, C. C., Simons, A. D., & Kim, H. K. (2012). The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. *Journal of Consulting and Clinical Psychology, 80,* 525–534. http://dx.doi.org/10.1037/a0029131

*Liberman, R. P., & Eckman, T. (1981). Behavior therapy vs. insight-oriented therapy for repeated suicide attempters. *Archives of General Psychiatry, 38,* 1126–1130. http://dx.doi.org/10.1001/archpsyc.1981.01780350060007

*Lopes, R. T., Gonçalves, M. M., Fassnacht, D. B., Machado, P. P., & Sousa, I. (2014). Long-term effects of psychotherapy on moderate depression: A comparative study of narrative therapy and cognitive-behavioral therapy. *Journal of Affective Disorders, 167,* 64–73. http://dx.doi.org/10.1016/j.jad.2014.05.042

Lovas, D. A., & Barsky, A. J. (2010). Mindfulness-based cognitive therapy for hypochondriasis, or severe health anxiety: A pilot study. *Journal of Anxiety Disorders, 24,* 931–935. http://dx.doi.org/10.1016/j.janxdis.2010.06.019

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice, 4,* 53–63. http://dx.doi.org/10.1111/j.1468-2850.1997.tb00099.x

Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42,* 602–611. http://dx.doi.org/10.1001/archpsyc.1985.01790290084010

Luborsky, R. P., Crits-Christoph, P., Mintz, J., & Auerbach, A. (1988). *Who will benefit from psychotherapy? Predicting therapeutic outcomes*. New York, NY: Basic Books.

*Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *The British Journal of Psychiatry, 190,* 496–502. http://dx.doi.org/10.1192/bjp.bp.106.024729

Manicavasgar, V., Parker, G., & Perich, T. (2011). Mindfulness-based cognitive therapy vs cognitive behaviour therapy as a treatment for non-melancholic depression. *Journal of Affective Disorders, 130,* 138–144. http://dx.doi.org/10.1016/j.jad.2010.09.027

*Markowitz, J. C., Kocsis, J. H., Fishman, B., Spielman, L. A., Jacobsberg, L. B., Frances, A. J., . . . Perry, S. W. (1998). Treatment of depressive symptoms in human immunodeficiency virus-positive patients. *Archives of General Psychiatry, 55,* 452–457. http://dx.doi.org/10.1001/archpsyc.55.5.452

*McBride, C., Atkinson, L., Quilty, L. C., & Bagby, R. M. (2006). Attachment as moderator of treatment outcome in major depression: A randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology, 74,* 1041–1054. http://dx.doi.org/10.1037/0022-006X.74.6.1041

McGuire, J. F., Piacentini, J., Brennan, E. A., Lewin, A. B., Murphy, T. K., Small, B. J., & Storch, E. A. (2014). A meta-analysis of behavior therapy for Tourette Syndrome. *Journal of Psychiatric Research, 50,* 106–112. http://dx.doi.org/10.1016/j.jpsychires.2013.12.009

*McLean, P. D., & Hakstian, A. R. (1979). Clinical depression: Comparative efficacy of outpatient treatments. *Journal of Consulting and Clin-*

ical Psychology, 47, 818–836. http://dx.doi.org/10.1037/0022-006X.47.5.818

*McNamara, K., & Horan, J. J. (1986). Experimental construct validity in the evaluation of cognitive and behavioral treatments for depression. *Journal of Counseling Psychology, 33,* 23–30. http://dx.doi.org/10.1037/0022-0167.33.1.23

*Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. *Journal of Consulting and Clinical Psychology, 71,* 404–409.

*Misri, S., Reebye, P., Corral, M., & Milis, L. (2004). The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: A randomized controlled trial. *Journal of Clinical Psychiatry, 65,* 1236–1241. http://dx.doi.org/10.4088/JCP.v65n0913

*Mohr, D. C., Boudewyn, A. C., Goodkin, D. E., Bostrom, A., & Epstein, L. (2001). Comparative outcomes for individual cognitive-behavior therapy, supportive-expressive group psychotherapy, and sertraline for the treatment of depression in multiple sclerosis. *Journal of Consulting and Clinical Psychology, 69,* 942–949. http://dx.doi.org/10.1037/0022-006X.69.6.942

*Murphy, G. E., Carney, R. M., Knesevich, M. A., Wetzel, R. D., & Whitworth, P. (1995). Cognitive behavior therapy, relaxation training, and tricyclic antidepressant medication in the treatment of depression. *Psychological Reports, 77,* 403–420. http://dx.doi.org/10.2466/pr0.1995.77.2.403

Newton-Howes, G., Tyrer, P., & Johnson, T. (2006). Personality disorder and the outcome of depression: Meta-analysis of published studies. *The British Journal of Psychiatry, 188,* 13–20. http://dx.doi.org/10.1192/bjp.188.1.13

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62,* 1157–1172. http://dx.doi.org/10.1002/jclp.20272

Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: Systematic review and meta-analysis. *BMC Psychiatry, 11,* 200. http://dx.doi.org/10.1186/1471-244X-11-200

Pallesen, S., Mitsem, M., Kvale, G., Johnsen, B-H., & Molde, H. (2005). Outcome of psychological treatments of pathological gambling: A review and meta-analysis. *Addiction, 100,* 1412–1422. http://dx.doi.org/10.1111/j.1360-0443.2005.01204.x

Pampallona, S., Bollini, P., Tibaldi, G., Kupelnick, B., & Munizza, C. (2004). Combined pharmacotherapy and psychological treatment for depression: A systematic review. *Archives of General Psychiatry, 61,* 714–719. http://dx.doi.org/10.1001/archpsyc.61.7.714

*Parker, G., Blanch, B., Paterson, A., Hadzi-Pavlovic, D., Sheppard, E., Manicavasagar, V., . . . Perich, T. (2013). The superiority of antidepressant medication to cognitive behavior therapy in melancholic depressed patients: A 12-week single-blind randomized study. *Acta Psychiatrica Scandinavica, 128,* 271–281.

Penninx, B. W., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., . . . Beekman, A. T. (2011). Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 133,* 76–85. http://dx.doi.org/10.1016/j.jad.2011.03.027

*Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research, 23,* 535–548. http://dx.doi.org/10.1023/A:1018724505659

*Persons, J. B., Burns, D. D., & Perloff, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research, 12,* 557–575. http://dx.doi.org/10.1007/BF01205010

*Persons, J. B., Roberts, N. A., Zalecki, C. A., & Brechwald, W. A. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behaviour Research and Therapy, 44,* 1041–1051. http://dx.doi.org/10.1016/j.brat.2005.08.005

*Power, M. J., & Freeman, C. (2012). A randomized controlled trial of IPT versus CBT in primary care: With some cautionary notes about handling missing values in clinical trials. *Clinical Psychology & Psychotherapy, 19,* 159–169. http://dx.doi.org/10.1002/cpp.1781

*Propst, L. R., Ostrom, R., Watkins, P., Dean, T., & Mashburn, D. (1992). Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. *Journal of Consulting and Clinical Psychology, 60,* 94–103. http://dx.doi.org/10.1037/0022-006X.60.1.94

*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. http://dx.doi.org/10.1017/S0033291708003772

Rector, N. A., & Beck, A. T. (2012). Cognitive behavioral therapy for schizophrenia: An empirical review. *Journal of Nervous and Mental Disease, 200,* 832–839. http://dx.doi.org/10.1097/NMD.0b013e31826dd9af

Rector, N. A., Zuroff, D. C., & Segal, Z. V. (1999). Cognitive change and the therapeutic alliance: The role of technical and nontechnical factors in cognitive therapy. *Psychotherapy: Theory, Research, Practice, Training, 36,* 320–328. http://dx.doi.org/10.1037/h0087739

Reinholt, N., & Krogh, J. (2014). Efficacy of transdiagnostic cognitive behaviour therapy for anxiety disorders: A systematic review and meta-analysis of published outcome studies. *Cognitive Behaviour Therapy, 43,* 171–184. http://dx.doi.org/10.1080/16506073.2014.897367

*Rieu, J., Bui, E., Rouch, V., Faure, K., Birmes, P., & Schmitt, L. (2011). Efficacy of ultrabrief cognitive and behavioural therapy performed by psychiatric residents on depressed inpatients. *Psychotherapy and Psychosomatics, 80,* 374–376. http://dx.doi.org/10.1159/000323406

Rigby, A. S. (2000). Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. *Disability and Rehabilitation, 22,* 339–344. http://dx.doi.org/10.1080/096382800296575

Rosenthal, R. (1993). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Roshanaei-Moghaddam, B., Pauly, M. C., Atkins, D. C., Baldwin, S. A., Stein, M. B., & Roy-Byrne, P. (2011). Relative effects of CBT and pharmacotherapy in depression versus anxiety: Is medication somewhat better for depression, and CBT somewhat better for anxiety? *Depression and Anxiety, 28,* 560–567. http://dx.doi.org/10.1002/da.20829

*Rush, A. J., Beck, A. T., Kovacs, M., & Hollon, S. D. (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research, 1,* 17–37. http://dx.doi.org/10.1007/BF01173502

Segal, Z. V., Williams, J. M. G., & Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy. The Consumer Reports study. *American Psychologist, 50,* 965–974. http://dx.doi.org/10.1037/0003-066X.50.12.965

*Selmi, P. M., Klein, M. H., Greist, J. H., Sorrell, S. P., & Erdman, H. P. (1990). Computer-administered cognitive-behavioral therapy for depression. *The American Journal of Psychiatry, 147,* 51–56. http://dx.doi.org/10.1176/ajp.147.1.51

Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., . . . Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy, 47,* 902–909. http://dx.doi.org/10.1016/j.brat.2009.07.003

*Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-

interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522–534. http://dx.doi.org/10.1037/0022-006X.62.3.522

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., . . . Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology, 78,* 751–756. http://dx.doi.org/10.1037/a0020569

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760. http://dx.doi.org/10.1037/0003-066X.32.9.752

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 182–196. http://dx.doi.org/10.1037/0022-006X.63.2.182

*Strauman, T. J., Vieth, A. Z., Merrill, K. A., Kolden, G. G., Woods, T. E., Klein, M. H., . . . Kwapil, L. (2006). Self-system therapy as an intervention for self-regulatory dysfunction in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology, 74,* 367–376.

Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology, 78,* 429–437. http://dx.doi.org/10.1037/a0019631

Tarrier, N. (2005). Cognitive behaviour therapy for schizophrenia. A review of development, evidence and implementation. *Psychotherapy and Psychosomatics, 74,* 136–144. http://dx.doi.org/10.1159/000083998

*Taylor, F., & Marshall, W. (1977). Experimental analysis of a cognitive behavioral therapy for depression. *Cognitive Therapy and Research, 1,* 59–72. http://dx.doi.org/10.1007/BF01173505

*Teichman, Y., Bar-el, Z., Shor, H., Sirota, P., & Elizur, A. (1995). A comparison of two modalities of cognitive therapy (individual and marital) in treating depression. *Psychiatry, 58,* 136–148.

*Thase, M. E., Friedman, E. S., Fasiczka, A. L., Berman, S. R., Frank, E., Nofzinger, E. A., & Reynolds, C. F., III. (2000). Treatment of men with major depression: A comparison of sequential cohorts treated with either cognitive-behavioral therapy or newer generation antidepressants. *Journal of Clinical Psychiatry, 61,* 466–472. http://dx.doi.org/10.4088/JCP.v61n0702

*Thase, M. E., Simons, A. D., Cahalane, J., McGeary, J., & Harden, T. (1991). Severity of depression and response to cognitive behavior therapy. *The American Journal of Psychiatry, 148,* 784–789. http://dx.doi.org/10.1176/ajp.148.6.784

*Thompson, L. W., Coon, D. W., Gallagher-Thompson, D., Sommer, B. R., & Koin, D. (2001). Comparison of desipramine and cognitive/behavioral therapy in the treatment of elderly outpatients with mild-to-moderate depression. *The American Journal of Geriatric Psychiatry, 9,* 225–240. http://dx.doi.org/10.1097/00019442-200108000-00006

*Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology, 55,* 385–390. http://dx.doi.org/10.1037/0022-006X.55.3.385

*Tovote, K. A., Fleer, J., Snippe, E., Peeters, A. C., Emmelkamp, P. M., Sanderman, R., . . . Schroevers, M. J. (2014). Individual mindfulness-based cognitive therapy and cognitive behavior therapy for treating depressive symptoms in patients with diabetes: Results of a randomized controlled trial. *Diabetes Care, 37,* 2427–2434. http://dx.doi.org/10.2337/dc13-2918

*Wagner, B., Horn, A. B., & Maercker, A. (2014). Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. *Journal of Affective Disorders, 152–154,* 113–121. http://dx.doi.org/10.1016/j.jad.2013.06.032

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923. http://dx.doi.org/10.1037/0022-006X.73.5.914

Wampold, B. E., Minami, T., Baskin, T. W., & Callen Tierney, S. (2002). A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of Affective Disorders, 68,* 159–165. http://dx.doi.org/10.1016/S0165-0327(00)00287-1

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,* 203–215. http://dx.doi.org/10.1037/0033-2909.122.3.203

*Watson, J. C., Gordon, L. B., Stermac, L., Kalogerakos, F., & Steckley, P. (2003). Comparing the effectiveness of process-experiential with cognitive-behavioral psychotherapy in the treatment of depression. *Journal of Consulting and Clinical Psychology, 71,* 773–781. http://dx.doi.org/10.1037/0022-006X.71.4.773

Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y., Weeks, W. B., & Friedman, M. J. (2013). Meta-analysis of the efficacy of treatments for posttraumatic stress disorder. *Journal of Clinical Psychiatry, 74,* e541–e550. http://dx.doi.org/10.4088/JCP.12r08225

Wells, A. (2000). *Emotional disorders and metacognition: Innovative cognitive therapy*. Chichester, UK: Wiley.

Wells, A., & King, P. (2006). Metacognitive therapy for generalized anxiety disorder: An open trial. *Journal of Behavior Therapy and Experimental Psychiatry, 37,* 206–212. http://dx.doi.org/10.1016/j.jbtep.2005.07.002

*Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy, 43,* 1243–1261. http://dx.doi.org/10.1016/j.brat.2004.09.006

WHO. (2012). *Depression fact sheet*. Retrieved from http://www.who.int/mediacentre/factsheets/fs369/en/index.html

WHO. (2014). *Global burden of disease (GBD)*. Retrieved from http://www.who.int/healthinfo/global_burden_disease/gbd/en/

*Wierzbicki, M., & Bartlett, T. S. (1987). The efficacy of group and individual cognitive therapy for mild depression. *Cognitive Therapy and Research, 11,* 337–342. http://dx.doi.org/10.1007/BF01186284

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology, Research and Practice, 24,* 190–195. http://dx.doi.org/10.1037/0735-7028.24.2.190

*Wilson, P. H., Goldin, J. C., & Charbonneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. *Cognitive Therapy and Research, 7,* 111–124. http://dx.doi.org/10.1007/BF01190064

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., . . . Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21,* 655–679. http://dx.doi.org/10.1016/j.euroneuro.2011.07.018

*Wright, J. H., Wright, A. S., Albano, A. M., Basco, M. R., Goldsmith, L. J., Raffield, T., & Otto, M. W. (2005). Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. *The American Journal of Psychiatry, 162,* 1158–1164. http://dx.doi.org/10.1176/appi.ajp.162.6.1158

Zoogman, S., Goldberg, S. B., Hoyt, W. T., Miller, Lisa. (2014, January). Mindfulness interventions with youth: A meta-analysis. *Mindfulness*. Advance online publication. http://dx.doi.org/10.1007/s12671-013-0260-4

Received May 28, 2014
Revision received February 26, 2015
Accepted March 3, 2015 ■

TITLE

The Effect of Cognitive Behavioral Therapy as an Anti-Depressive Treatment is Falling: Reply to Ljòtsson et al. (2016) and Cristea et al. (2016).

RUNNING HEAD

Effect of Cognitive Behavioral Therapy is Falling

AUTHORS

Oddgeir Friborg

Tom J. Johnsen

AFFILIATION

Faculty of Health Sciences, Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway.

CORRESPONDING AUTHOR

Tom J. Johnsen, Faculty of Health Sciences, Department of Psychology, UiT The Arctic University of Norway, University of Tromsø, N-9037 Tromsø, Norway. E-mail: tjj@psykologtromso.no

ABSTRACT

This paper critically reassess Ljótsson et al.'s (2016) nonlinear reanalysis and review Cristea et al.'s (2016) extension of our original meta-analysis (Johnsen & Friborg, 2015) reporting a decline in the effects of cognitive behavioral therapy (CBT) for treating unipolar depression. Ljótsson fitted a piecewise meta-regression model to the data indicating a halt in the decline from the year 1995 onwards, hence concluding that CBT is not gradually losing its efficacy. We reanalyzed the data for nonlinear time trends and replicated their findings for the 34 studies using Hamilton Rating Scale for Depression as outcome, but not for the 67 studies using Beck's Depression Inventory as outcome. The best nonlinear model was quadratic rather than flat (or linear) from 2001 onwards; which opposes Ljótsson's conclusion of stability in effects. Cristea et al.'s identified additional studies, but their new analyses provided mixed support for a linear decline in CBT effects. They could not dismiss a decline except only in the most stringent analytic condition—namely, when analyzing only 29 randomized controlled trials based on between-group effect sizes solely. Their study includes several questionable methodological choices, so we expand on the discussion of these disparate meta-analytic findings. Of particular concern is the tendency to downplay the fact that when looking at all of the studies together—there is a clear decline in the effects of CBT, which should concern therapy researchers within the field rather than being explained away.

The Effects of Cognitive Behavioral Therapy as an Anti-Depressive Treatment is Falling: Reply to Ljòtsson et al. (2016) and Cristea et al. (2016).

As the authors of a meta-analysis examining the time trends in the effectiveness of cognitive behavioral therapy (CBT) as an anti-depressive treatment (Johnsen & Friborg, 2015), we are pleased to read critical follow-up papers. The increased focus on historical trends in the effects of psychotherapy is likely to contribute positively to the development and implementation of more efficacious psychotherapies in the future. Some steps towards this end have been taken with the recent publication of two re-reviews of our original findings: the first being a statistical re-analysis (Ljótsson, Hedman, Mattsson, & Andersson, 2016) and the second being a meta-analytic extension (Cristea et al., 2016). We do however note considerable methodological or statistical issues in both papers, which has prompted the current reply.

Reply to Ljótsson et al. (2016)

Ljótsson et al. (2016) concluded that the decline in CBT effects had stopped falling. They arrived at this conclusion following a meta-regression analysis that had curvature or segmented parameters added to the model. This model ostensibly revealed a leveling off in the decline from 1995 onwards. Hence, they concluded that the CBT treatment effects had not declined during the last 20 years, and that the effects of the current CBT protocols vary around their "true" clinical effect. Given the authors' statement in their final paragraph of the paper, "we did not find any support in their data for their conclusion that the effects of CBT are in decline", one could get the impression that the conclusions by Johnsen

and Friborg (2015) were ill-informed. Inspired by these new analyses, we also reanalyzed the dataset to examine whether we would arrive at similar conclusions.

We first wish to put into perspective the basic message regarding time trends in CBT treatment effects. The predicted decline from 1977 to 2014 in the Beck's Depression Inventory (BDI) effect sizes (ESs), based on Ljótsson's analyses, is a Hedge's g = 1.33 (falling from 2.76 to 1.43). The corresponding decline, according to the linear model, is g = 1.09 (falling from 2.27 to 1.18). The piecewise model predicts a steeper initial fall; differences in the 2014 effects between the two methods is g = .25, which is rather negligible. Comparable reduction in the Hamilton Rating Scale for Depression (HRSD) effects based on the piecewise model is g = 1.69 (falling from 3.29 to 1.60) and g = 1.13 for the linear model (falling from 2.52 to 1.39). The difference in 2014 effects is g = .21, also quite unimportant. What seems to be at stake here is more than just the difference between the two methods in predicting the treatment effects at present. Rather it seems to be the conceptual underpinnings of the linear model, which suggests continual decline. Conversely, the piecewise model provides more solace by indicating a halt in the fall. The current treatment effects are still to be considered good and perhaps even strong; yet, it is puzzling why this happens.

We did not explicitly analyze curvature or segmented time trends in the original paper as did Ljótsson et al. (2016); however, we did not miss this entirely, as is evident in Figure 7 from Johnsen and Friborg (2015). Figure 7 portrays how the

time coefficients change depending on the starting year for study inclusion, even turning positive only when including studies after the year 1995. The choice of 1995 as a breakpoint in Ljótsson's analyses was motivated by this figure, and hence Ljótsson and colleagues did not use an empirical criterion for deciding their breakpoint. Here, we thoroughly examined nonlinear time trends and used a statistical criterion for deciding a regression slope segmentation breakpoint.

A nonlinear reanalysis

BDI: We first visually inspected the scattering of the weighted ESs in Figure 1 and noticed that the decline was fairly stable until the year 2001. Moreover, the reported ESs between the years 2001 and 2014 seemed slightly inverse u-curved rather than completely flat, as the piecewise model suggests. In order to examine this possibility, we specified a segmented model consisting of two parts: a linear part describing the whole time period (1977-2014) and a quadratic part describing the time trend following the breakpoint. The fitted weighted least square regression model using random effects model weights from the Johnsen and Friborg (2015) paper was:

$$f(ES) = b_0 + b_1 Y + b_2 Y_{pos} + b_3 Y_{pos}^2$$

Coding of $Y_{pos} = \begin{Bmatrix} Y \; if \; Y \geq 0 \\ 0 \; if \; Y < 0 \end{Bmatrix}$, 0 representing the centered year (breakpoint).

Where Y = year.

The breakpoint was empirically chosen by searching for the publication year that could render both parts of the model to be statistically significant. This only

happened if publication year was centered at the year 2001. Statistical significance for the nonlinear regression parameters is presented in Table 1. The nonlinear model also yielded the highest model fit in terms of the R-square index. Figure 1 illustrates the linear and nonlinear trends visually, including the explanation of the respective amounts of between-study variance.

--- Insert Table 1 and Figure 1 about here ---

In addition to presenting normal standard errors, we also produced bootstrapped error bands based on 5,000 resamplings. Residuals of both models were normally distributed ($Z_{skewness}$ = 1.07 and .86, $Z_{kurtosis}$ = .06 and .11, and both Kolmogorov–Smirnov tests were non-significant). Hence, the bootstrapped confidence intervals overlapped strongly with the model-based intervals.

HRSD: These effects were best described by a piecewise model. The nonlinear model (similar as above), which fit best when centered at the year 2001 (R^2 =.298) was not better than the best piecewise model centered at 1998 (R^2 = .290). The first part of the segmented nonlinear regression was significant (b_1Y = -.080, p = .002); however, this was not the case with the second quadratic part (b_2Y_{pos} = .245, p = .07; $b_3Y^2_{pos}$ = -.011, p = .27) although the coefficients were surprisingly comparable with the nonlinear BDI coefficients.

Summary and discussion

These additional analyses indicate that CBT effects, as measured by the BDI, have fallen linearly from 1977 until 2001, and not until 1995 as proposed by Ljótsson

et al. (2016). The fall has been going on for about 24 years, which encompasses half of all studies (33 of 67). From 2001 onward, the treatment effects have not declined further, but stability in the effects cannot be claimed due to the significant segmented quadratic model. This model shows a temporary rise followed by another fall, which may or may not be ongoing. Whatever is true, the major point is that a flattening in the treatment effects of CBT or that the CBT effects now vary around their "true" value, as Ljótsson et al. (2016) conclude, is not well supported by the available data. The segmented nonlinear model with the publication year 2001 as the breakpoint also explained 2.4% more of the variation in the treatment effects than the piecewise model with year 1995 as the breakpoint. We acknowledge Ljótsson and colleagues' effort in addressing nonlinear time trends as it helped gain additional insight into temporal trends. But since they overlooked a significant quadratic curvature in the second part of the segmented model, their conclusions are overstated.

Regarding the HRSD effects, the piecewise model fit the data best; hence, we are left with a mixed picture. There are however good reasons for weighting the BDI outcome data more heavily since the statistical power for detecting nonlinear HRSD trends, with only 34 studies available, is considerably smaller compared to the 67 available BDI studies. The HRSD measure also compares less favorably with the BDI measure in terms of poorer sensitivity to the psychological symptoms of depression related to nonendogenous, atypical depression or personality dysfunctions (Enns, Larsen, & Cox, 2000). HRSD seems, on the other hand, to be more sensitive to somatic symptoms related to endogenous depression. This makes sense since the BDI was specifically designed by the

founder of CBT, Aaron Beck, to identify improvements in attitudinal and cognitive components following therapy (e.g., hopelessness, self-worthlessness, self-dislike, or guilt). We thus consider the BDI to be more valid in evaluating the effects of his therapy than HRSD, which also the large number of clinical trials using the BDI is a testimony of.

What additional points can be made of this reanalysis? First, the present reanalysis do not change the basic message stating that CBT effects have fallen considerably across two and half decades. In fact, the predicted ES for the year 2014 even comes out slightly worse for the segmented nonlinear (g = 1.12), as compared to the linear, model (g = 1.18). Nevertheless, the current ESs are strong, hence CBT is still to be considered as an effective anti-depressive treatment.

Second, it may be wise to include nonlinear time trends in future meta-analyses of therapy studies in order to obtain more accurate information about psychotherapy effects. Since the current reanalysis shows that the nonlinear time trend explains a considerable portion of the between-study treatment variance (almost 30%), future meta-analytic summaries of treatment effects should not dismiss potential time trends.

Third, the psychotherapy research field may profit hugely by establishing a common minimum of variables/measures that is to be included as moderator variables in all future therapy trials. That would not only benefit the individual researcher attempting to analyze reasons for better or worse treatment

outcomes in the study at hand, but also any future meta-analytic attempts at analyzing reasons for time trends in psychotherapy effects more exactly.

Last, since the BDI effects during the last 13 years do not follow a flat trend but rather are in decline again, we believe a weather-climate analogy is an apt comparison: although weather varies across decades, the long-term climate changes (as projected by a linear model) may be regarded as the most reliable indicator.

Reply to Cristea et al. (2016)

The study by Cristea et al. (2016) offers a comprehensive extension of the original meta-analysis as they identified a number of additional CBT studies. They also introduce a number of methodological changes that are poorly justified and even incorrect in our opinion. Their analysis offers rather mixed results concerning whether the effects of CBT are in decline or not, which we would like to critically review.

A clear strength of their paper is the identification of 30 additional studies, including 12 randomized controlled trials (RCTs), as compared to the original meta-analysis. This increases the reliability of the time trend coefficient. Unfortunately, they seem to suggest that Johnsen and Friborg (2015) had somehow missed these studies, when in fact they had simply revised the inclusion criteria by including papers published in all languages; in contrast, the original paper included only papers exclusively published in English for the purposes of interpretation. Cristea et al. (2016) also found some inconsistencies in Johnsen and Friborg's (2015) selection of papers—they pointed out that four of the papers should not have been included and questionable calculation of the effect sizes (ESs) for two of the included papers; however, these revisions did not change the original findings.

Cristea et al. (2016) introduce a number of methodological changes to the meta-analysis. For instance, they link the combination of non-RCT and RCT studies to the high degree of heterogeneity noted between studies, which we also believe may be the case. This is why we conducted sub-group analyses for within and

controlled designed studies separately. More troublingly, they argue for excluding non-RCTs from their main analysis because such trials may yield biased findings owing to a plethora of selection biases, which may cause the participating groups to differ at pretest. We fully acknowledge this important objection. However, they also argue that non-RCTs studies may be more correlated with the passage of time than RCT studies may be. They provide no justification for this claim, nor can we conceive of a sensible reason for making it. Why would a potential selection bias (e.g., more motivated patients or more depressed patients) be systematically present solely during the 70s or 80s and not later on? As we regard this possibility as tiny at best, we consider Cristea et al.'s (2016) exclusion of a large array of clinically relevant studies, instead of including them despite the risk of minor time trend biases, to be a major error. Their choice therefore seems to serve a confirmatory purpose.

Second, they object to the use of within-group ESs calculated from pre-post data and to the combination of within- and between group ESs when analysing all studies together. Their argument is that within-group ESs cannot be disentangled from the context in which the study was conducted, which in practise means having a comparison group. While we acknowledge this point, it is important to note that our study (Johnsen & Friborg, 2015) did not rely solely on within-group ESs. As mentioned above, we conducted sub-group analyses, which revealed that the decline extended to the between-group condition. It is important to keep in mind that the vast majority of the calculated within-group ESs was based on randomized clinical trials; however, these studies could not be calculated as between-group ESs as they were compared to other treatment

arms (e.g., medication) or did not include a no-intervention group. Many of these studies thus had a "context" that was not defined by a control group arm. The only method for quantifying the ESs from these relevant treatment arms was to use the within-group formula. A known problem is overestimation of the ES, which may be adjusted for with the correlation between the pre- and post-test measure. The higher this correlation, the lower the within-group ES. In our case, we imputed a large correlation (r = 0.7) for studies not reporting it, thus reducing overestimation risks. But even if these ESs were overestimated it is difficult to conceive of a sensible explanation for why within-group calculated ESs might favor earlier CBT trials compared to later ones, whereas between-group calculated ESs do not, which Cristea et al. (2016) assume. Again, they provide no justification for why this should be the case. Another problem is that Cristea et al. (2016) base their "new" analysis on post-test scores only, whereas we used the standard deviation of the difference score in both the within- and between-group estimations in order to use a comparable denominator. The use of difference scores also corrects between-group ESs for any pre-treatment differences that may occur despite randomization in small sample studies. The meta-analysis of all studies combined was thus more correct in our original approach, whereas Cristea et al. (2016) mix the use of post- and difference scores when analysing all studies together. A final argument for including all available studies is to ensure a substantially larger study pool, which is important for avoiding an underpowered statistical analysis and enabling weaker, yet still clinically important, statistical effects stretching across decades to appear. Studies of clinical effectiveness should, in our opinion, record whether any clinical improvement (or decay) is apparent in both lesser or better defined

contexts. This is well reflected in the long-standing discussion of the use of RCT designs in studies of clinical effectiveness (Persons & Silberschatz, 1998), quote: "*RCT advocates have sacrificed clinical validity in the effort to maximize experimental control*". If ESs do change with time, independently or within a particular context (control group or not), then time trends would still be clinically relevant and thus would need to be addressed. Omitting these studies, as Cristea et al. (2016) do in their "new" analysis, thus represents a larger mistake than including them.

Third, Cristea et al. (2016) consider the use of univariate regression analysis as misleading by increasing the risk of type I error in hypothesis testing. It is important to note that our primary hypothesis exclusively concerned publication year—namely, to what extent an increase in treatment effects was evident across time, as is evident within most other branches of medicine (e.g., publication series of *Advances in Medicine and Biology*). Our approach was not to examine a set of moderators and then select the one(s) that were statistically significant; hence, the univariate regression approach seemed optimal. We did conduct multiple two-way interaction tests between publication year and the moderators (i.e., *time* x *moderator*), hence these tests were prone to type 1 error. But since support of these tests would weaken the temporal (or time) hypothesis, any appropriate statistical adjustments would only make the rejection of the temporal hypothesis less likely. Cristea et al. (2016) conducted a so-called "full model" meta-regression analysis that included all moderators in a multivariate fashion. Moreover, they retained all variables in the model even though 10 of the 11 moderators were statistically non-significant, which reduces the degrees of

freedom substantially, which is quite negative for the statistical power in small samples. A defendable reason for conducting multivariable testing, as Cristea et al. (2016) did, would be if: a) theory or previous empirical evidence substantiate the inclusion of such a large array of predictors, b) omitting a moderator would significantly bias the estimation of the time coefficient, and c) the moderator contributes significantly to the explanation of ES. Since studies of temporal development of psychotherapy effects are a completely new endeavour, neither theory nor relevant empirical evidence exist and support such a-priori multivariable models. Estimation biases may nevertheless occur if an omitted moderator correlates positively with time. This was potentially the case for two moderators (i.e., study quality ratings, and type of BDI measure), but none of these contributed significantly to the explanation of between-study ESs. Inclusion of such non-significant variables (and Cristea et al. included 10 variables) would thus introduce a "spurious" adjustment of the regression model. Had our study context been one that embraces multiple hypothetical explanations for the decline, Cristea et al.'s approach had made sense. Since publication year was our sole hypothesis, their objection is irrelevant to the original statistical analysis.

Fourth, they claimed that time trend analyses should be based on "intention-to-treat" (ITT) rather than "completer" data. In our case, we had no *a priori* reason to consider ITT as any better than an analysis based on completers. Although ITT analyses do retain all patients and thus reduce systematic attrition, they can be biased (Lane, 2008) due to undue assumptions of no change among patients that drop out. Cristea et al. (2016) further argue that the ITT procedure may be less

susceptible to time trend effects than completer data, but again provide no justification or evidence for this point. In contrast, our choice of using data from completers was well-informed because this was the only information available in early CBT trials. Since completer data were uniformly reported and the dropout rate from CBT studies is low in general, we consider analyses based on such data as equally (if not more) correct than analyses based on ITT data.

Finally, Cristea et al. combined treatment effects from trials including several subgroups rather than coding selected subgroups according to an *a priori* criterion. This strategy yielded results supporting weaker time trends, which they argued as superior to basing the calculations on a particular group. Our argument for selecting the most severely depressed patient group was to achieve a uniform comparison group rather than merely collapsing a variety of groups to serve as a comparison. This strategy, if anything, should reduce rather than increase study heterogeneity, which was one of their prime concerns. Since baseline severity does not moderate the outcome of CBT for depression, as reported in a meta-analysis by one of the authors (Driessen, Cuijpers, Hollon, & Dekker, 2010), this is another example of poorly justified selection of studies.

The exclusion of CBT studies regarded as outliers is inherently problematic because such studies may represent less frequent but still true observations in the population. Indeed, we examined the unstandardized residuals for the segmented nonlinear time trend model in our reply to Ljótsson et al. (2016), which showed an almost perfect normal distribution (skewness $Z = 0.86$, kurtosis $Z = 0.11$) with no extreme observations. Hence, removal of outliers is

unjustified, particularly Cristea et al.'s (2016) choice to consider one-third of the within studies in their meta-reanalysis as outliers. They justify their choice by branding it "the winners curse", meaning that there is no way for ESs to go but down. Hence, we should not expect anything other than a decline—even after 40 years of time to improve psychotherapy. This is an extremely pessimistic view on psychotherapy as a field, which is highly speculative and is backed by no evidence, to our knowledge, from research on time trends in psychotherapy.

Cristea et al.'s (2016) choice of splitting the study pool according to whether studies were conducted in the US (k = 25) or the rest of the world (k = 20) is also poorly justified with regard to the time trend hypothesis, although it offers some interesting findings in itself.

We agree with Cristea et al. (2016) that meta-analyses are inherently tricky to design and perform because of the considerable heterogeneity of the studies being aggregated. Another challenging but important aspect relates to communication of the findings in an objective, unbiased, and prudent fashion. In this regard, Cristea et al.'s (2016) study appears to deliberately downplay the fact that many (if not most) of their adjusted analytic conditions support our original findings. In fact, it seems that only the most stringent condition—wherein only 29 RCTs based on between-group ESs were included—reliably contradicted our original findings. However, the largest analytic conditions, based on 45 and up to 75 RCTs, mainly supported the original findings. As readers, we are thus left with an obscure picture; the authors seem to selectively favor results that do not confirm a decline and reject those indicating such a

decline. Indeed, they exclude studies they construe as outliers, and studies including inpatients, and prioritize ITT over completers' analyses. Even after doing all this, the original findings were still evident, which led Cristea et al. to disregard within-group studies altogether to achieve the desired non-significance. Remarkably, to achieve this, they had to reduce the largest analytic condition from 75 to 29 studies! Even at this point, a negative time trend was still present (*beta* = -.01, *p* = .22); however, the low statistical power precludes any strong conclusions. In sum, their paper lacks, in our view, a balanced portrayal of the results, which is of major concern because at least four of the authors are, to our knowledge, adherers to or advocates of CBT. Cristea et al.'s (2016) characterization of their own meta-analysis as the "gold standard" analysis is thus not credible given the current criticism.

Despite Cristea et al.'s (2016) removal of within-group ES calculations to achieve a non-significant decline, the fact remains that, whatever the causes and contextual underpinnings, CBT as a treatment has overall suffered a systematic decline in its ability to treat depressive symptoms. In other words, today, fewer patients recover to the same extent as they did in the past. To brush off this important discovery as a spurious observation reminds us of the idiom, "burying one's head in the sand." To illustrate: if a chemotherapy drug exhibited a significant negative time trend in its ability to treat cancerous tumor cells when considering all studies in a meta-analysis, would any right-minded person still consider this drug as efficacious and safe as originally thought? Or, would it be wise to start addressing the problem and discuss ideas about how to improve this trend? This perspective is unfortunately lacking in both of the recent meta-

analytic re-analyses (Cristea et al., 2016; Ljótsson et al., 2016), which is of concern for future improvements.

REFERENCES

Cristea, I. A., Stefan, S., Karyotaki, E., David, D., Hollon, S. D., & Cuijpers, P. (2016). The effects of cognitive behavioral therapy are not systematically falling: a revision of Johnsen & Friborg (2015). *Psychological Bulletin, accepted for publication*.

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 metaanalysis. *Journal of Consulting and Clinical Psychology, 78*(5), 668–680. doi:10.1037/a0020570

Enns, M. W., Larsen, D. K., & Cox, B. J. (2000). Discrepancies between self and observer ratings of depression: The relationship to demographic, clinical and personality variables. *Journal of Affective Disorders, 60*, 33-41.

Johnsen, T. J., & Friborg, O. (2015). The Effects of Cognitive Behavioral Therapy as an Anti-Depressive Treatment is Falling: A Meta-Analysis. *Psychological Bulletin, 141*(4), 747-768.

Lane, P. (2008). Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical statistics, 7*, 93-106. doi:10.1002/pst.267

Ljótsson, B., Hedman, E., Mattsson, S., & Andersson, E. (2016). The Effects of Cognitive Behavioral Therapy for Depression are not Falling: A Re-Analysis of Johnsen & Friborg (2015). *Psychological Bulletin, accepted for publication.*

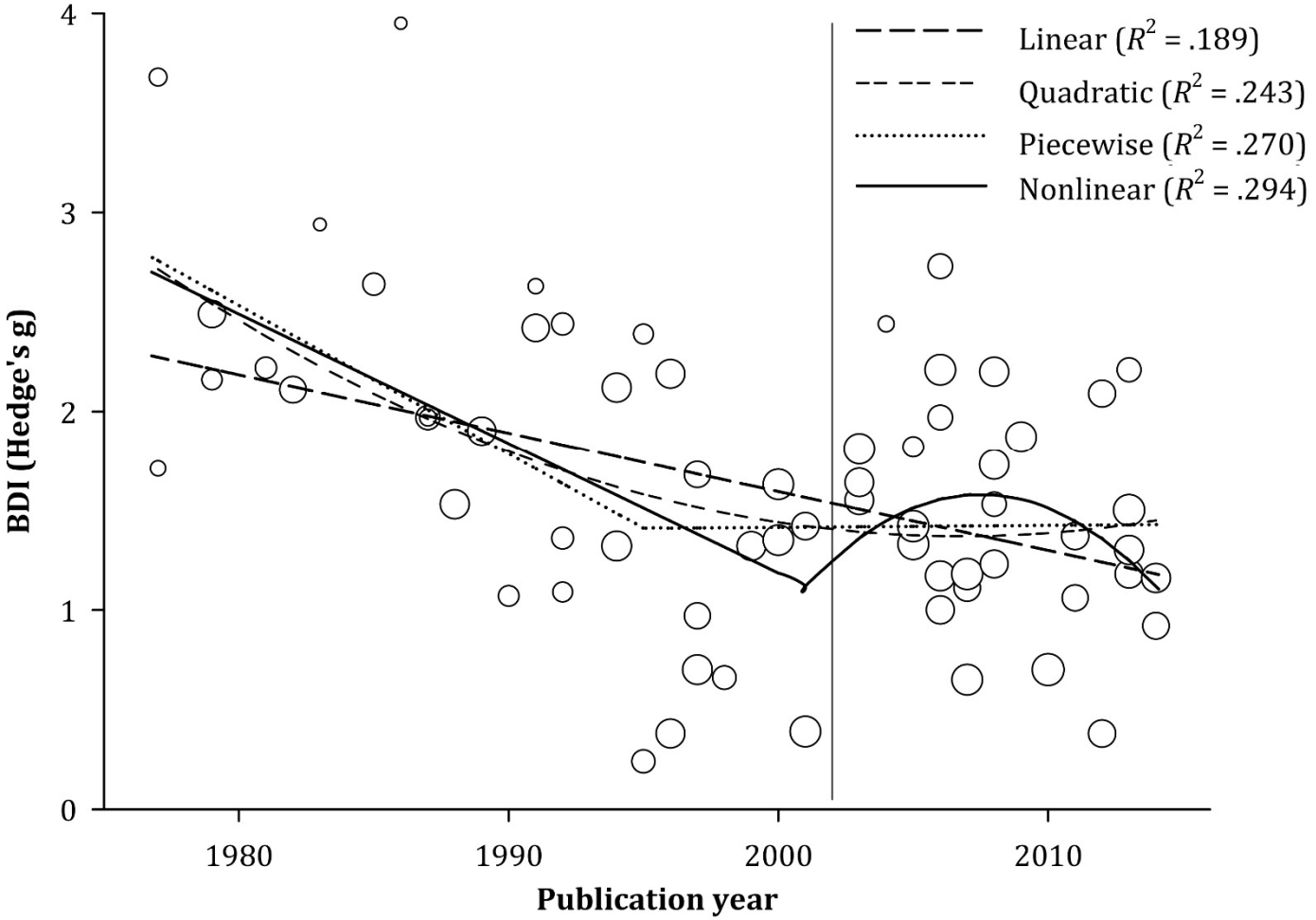
Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapist? *Journal of Consulting and Clinical Psychology, 66*(1), 126-135.

Table 1. *Comparison of Two Nonlinear Models for Predicting Change in BDI Effect Sizes Across Time.*

		beta	*p*	*CI* .95	Bootstrapped *CI* .95
1 Piecewise	R^2=.270				
b_0		1.4191			
Y		-0,0744	< .001	-.1112 -.0376	-.1130 -.0410
$Y_{pos} \geq 1995$		0,0741	.010	.0183 .1300	.0190 .1340
2 Nonlinear	R^2=.294[a]				
b_0		1.120			
Y		-.0653	< .001	-.0934 -.0371	-.0907 -.0408
$Y_{pos} \geq 2001$		.2070	.007	.0586 .3554	.0684 .3494
$Y^2_{pos} \geq 2001$		-.0109	.043	-.0215 -.0004	-.0210 -.0016

Notes. b_0 = intercept, *beta* = unstandardized coefficient, *p* = p-value, *CI* .95 = 95% confidence interval.

Figure 1. *Time Trends for the Different Meta-Regression Prediction Models*



Canadian Psychology / Psychologie canadienne
2018, Vol. 59, No. 1, 15–30

0708-5591/18/$12.00 http://dx.doi.org/10.1037/cap0000132

A Meta-Analysis of Group Cognitive–Behavioral Therapy as an Antidepressive Treatment: Are We Getting Better?

Tom J. Johnsen and Jens C. Thimm
UiT The Arctic University of Norway

This meta-analysis examines temporal changes (time trends) in the effects of group cognitive–behavioral therapy (GCBT) as a treatment for unipolar depression. In this exploratory study, 37 studies (comprising 1,016 patients) conducted between 1980 and 2015 were included, and their effect sizes (ESs) were quantified as Hedge's *g* based on the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). In the main analyses, within-group (prepost) comparisons ($k = 35$ for the BDI, and $k = 14$ for the HRSD) and controlled (between group) comparisons ($k = 16$ for the BDI) were conducted. The average within-group ES was 1.33 (95% confidence interval [CI] = 1.16–1.19) for the BDI and 1.56 for the HRSD (95% CI = 1.20–1.90). The publication year of each study was examined as a linear metaregression predictor of ES. The results showed a significant increase in effect sizes for GCBT over time, with the BDI as outcome measure. However, this was not found for the HRSD as outcome measure. Subgroup analyses were performed on selected moderator variables to determine whether there was any covariation with effect sizes. These analyses revealed that in trials conducted without following a set manual, effect sizes increased with time—an association not found for trials using a manual. Potential causes and implications are discussed.

Keywords: group cognitive–behavioral therapy, effectiveness, depressive disorders, meta-analysis, treatment manual

Depression is among the most prevalent psychiatric disorders in the western world (National Institute of Mental Health, 2015). Fourteen percent of patients with major depression have the illness for over 5 years (Patten, 2006), with an average duration of almost 10 years (Friborg et al., 2014). A crippling consequence of depression is its recurrent nature, as approximately 6 out of 10 patients experience a relapse (Solomon et al., 2000). Suffering is thus reinstated for the majority, many of whom do not return for more treatment (Andrews, 2001). Moreover, for those experiencing two or more episodes, depression may develop into a chronic condition (Blanco et al., 2010). The costs of depression are hence substantial for society (Sobocki, Jönsson, Angst, & Rehnberg 2006; Mental Health Foundation, 2010). Regular assessments of treatment effects, as well as efforts to identify the most efficient treatments, are therefore important.

In a recent comprehensive systematic review of the treatment effects of individual cognitive–behavioral therapy (CBT) for depression (Johnsen & Friborg, 2015), a significant decline in the CBT effect was observed across all outcome measures. Two prepost uncontrolled within-group analyses and two controlled between-groups analyses were conducted, using the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) as outcome measures, respectively. Seventy studies published between 1977 and 2014, with a total of 2,426 participants, were analyzed using a metaregression design with time as the main predictor. Strong effect sizes (ESs), in the range of $g = 1.37$ to 1.89, were observed for the different analytic conditions. However, a decline in ES was also observed with the passage of time, meaning that the most recent studies showed smaller effect sizes than did the older trials. The potential reasons for the observed decline in effect sizes were hypothesized to be related to the competency of the therapists, a possible loss of the placebo effect, and/or a possible lack of adherence to the treatment manual (Johnsen & Friborg, 2015).

Although the treatment effect of individual CBT is considered good, with an effect size of approximately $g = 1.4$ for within-group studies (Cristea et al., 2017; Johnsen & Friborg, 2015), the temporal decline in posttreatment effect is considerable. This finding motivated the present analysis, which examines whether a similar negative historical trend in treatment effects, may also be present for CBT delivered as group therapy.

The previously mentioned meta-analysis of individual CBT was criticized by some authors, who objected that some of the statistical choices were inappropriate (Cristea et al., 2017; Ljótsson, Hedman, Mattsson, & Andersson, 2017). These objections, which mainly focused on the analytic mixture of randomized controlled trials (RCTs) and clinical trials and the perceived lack of attention to potential issues of heterogeneity, were subsequently addressed by the original authors (Friborg & Johnsen, 2017), who acknowledged some uncertainties while countering irrelevant critiques and reaffirming their original conclusions. Other researchers have also recently contributed with theoretical insight and potential explan-

This article was published Online First December 28, 2017.

Tom J. Johnsen and Jens C. Thimm, Faculty of Health Sciences, Department of Psychology, UiT The Arctic University of Norway.

Correspondence concerning this article should be addressed to Tom J. Johnsen, Faculty of Health Sciences, Department of Psychology, UiT The Arctic University of Norway, North-9037 Tromsø, Norway. E-mail: tjj@psykologtromso.no

atory factors related to the original findings (Dobson, 2016; Waltman, Creed, & Beck, 2016).

The finding that the effect of individual CBT for depression has been falling in recent decades raises the question of whether this decline is also occurring for other CBT treatment modalities for depression. Because we are not aware of any previous attempts to measure time-trends for group CBT (GCBT), the purpose of the present meta-analysis was to investigate the temporal development of the effectiveness of CBT for depression delivered in a group format. The statistical design and procedures are similar to the Johnsen and Friborg (2015) study concerning individual CBT. However, attempts have been made to alleviate some of the previous concerns with that study and to make the findings of the current investigation more robust. The present article takes an adjusted approach regarding selecting studies and calculating effect sizes. For example, this analysis exclusively included RCTs, and it measures and discusses heterogeneity scores for all analytic conditions. The aim was to achieve less variability across trials and better comparability across time.

GCBT Versus Individual CBT

Group psychotherapy is a common treatment modality for many disorders, including depression. Group therapy may be defined as a meeting of two or more people (usually six to eight persons) who work toward a common therapeutic goal. GCBT is based on traditional cognitive therapy (Beck et al., 1979) and typically includes elements such as case-formulation, Socratic dialogues, and ABC-analyses of antecedent events or situations and associated behaviors and cognitions with fellow group members, all of whom contribute with their personal experiences and points of view. In addition, homework, conceptualization of problems, and group attention to self-defeating beliefs are common elements within this format. GCBT gained popularity as a depression treatment beginning in the late 1970s and showed a promising degree of efficacy (e.g., Beck et al., 1979; Shaw, 1977). Although the focus in modern psychotherapy, by and large, has been the individual format, the increasing number of studies during the last decade on group therapy as a viable antidepressive treatment (e.g., Hans & Hiller, 2013; Huntley, Araya, & Salisbury, 2012) indicates a shift in interest toward the group format.

Cognitive treatment methods and principles are clearly defined and well understood by researchers and clinicians, hence facilitating homogeneity regarding the use of treatment techniques, ingredients, and assessment measures across studies. This methodological allegiance allows valid and reliable comparisons of effect sizes for GCBT interventions.

A large amount of research has confirmed that CBT is efficacious in treating depression. Meta-analyses published in the past three decades (e.g., Cuijpers et al., 2008; Dobson, 1989; Gloaguen et al., 1998; Hollon, Shelton, & Loosen, 1991; Wampold et al., 2002) conclude that individual CBT has a high degree of treatment efficacy. The effectiveness of GCBT for depression has also been demonstrated in meta-analyses (e.g., Huntley et al., 2012; McDermut, Miller, & Brown, 2001; Oei & Dingle, 2008; Okumura & Ichikura, 2014). Studies comparing individual versus GCBT for depression have typically reported the latter as less expensive (Tucker & Oei, 2007; Vos et al., 2005) but not necessarily less effective (Hans & Hiller, 2013; Khoshbooii, 2012; Shaffer et al., 1981) than the former. A recent meta-analysis (Burlingame et al., 2016) showed no differences in outcome between individual CBT and GCBT for depression in studies in which the treatment, the patients, and the doses (number of sessions) were identical. Findings like these highlight the potential importance of implementing group psychotherapy formats in addition to the individual format. However, although individual CBT and GCBT share the same treatment philosophy and therapeutic techniques, differences between the two formats can influence the effect of these treatment modalities, necessitating the need for separate analyses to optimize statistical reliability and validity. For example, group cohesion and normalization are potential therapeutic factors that are specific to GCBT (Whitfield, 2010). Because there do not seem to have been any attempts to evaluate how the efficacy of GCBT has evolved over decades, the purpose of the present article is to summarize the treatment effects across publication years to inform the clinical field about potentially important temporal trends.

Outcome Measures

An advantage of reviewing the historical development of treatment effects based on GCBT trials is the high degree of standardization regarding the choice of outcome measures. Since the late 1970s, the vast majority of studies have utilized the BDI. Thus, our primary outcome measure was the BDI, a self-report checklist measuring 21 symptoms related to depression. The scale was revised in 1996 (Beck, Steer, & Brown, 1996) and subsequently named the BDI-II. Although its latest version puts more emphasis on measuring a wider range of depressive symptoms, the two formats are very similar: They have the same number of items and utilize similar cut-off values. Therefore, the two versions of the BDI have been treated as interchangeable in the current investigation, although we have also performed a subanalysis based on scores exclusively from the first version of the BDI.

Another outcome measure that has been frequently used with regard to GCBT for depression is the HRSD. The HRSD is a widely used clinician-administered 17-item checklist; it measures similar symptoms of depression as the BDI. The correlations between the HRSD and the BDI are usually moderate to high, $r = .5$ to $.8$ (Beck et al., 1988, 1996).

Despite the substantial overlap between the BDI and the HRSD, the nonperfect correlation between the two instruments indicates differences, which may be because of the source of information (self-report vs. clinician rating). In addition, the BDI tends to focus on cognitive aspects of depression, whereas the HRSD tends to emphasize physical symptoms related to depression (Wampold & Imel, 2015). Thus, because of the differences between the two instruments, effect sizes were calculated using both the BDI and the HRSD and were analyzed separately. The research question posed in this meta-analysis was whether similar time-trend results would appear on both outcome measures.

Moderators of Temporal Treatment Effects

A selection of possible moderator variables was examined in addition to the temporal *time* factor in the present study. Given the exploratory nature of the study, it was decided to examine a broad range of potential moderator variables: gender, age, type of therapist, number of therapy sessions, country of origin (for conduc-

tion of studies), severity of depression, use of a treatment manual, and study quality. These variables are identical to the ones investigated in previous meta-analyses regarding the temporal development of ES (Cristea et al., 2017; Johnsen & Friborg, 2015). The objective was to identify any possible moderators in the main analysis regarding temporal development of effect sizes. If any moderator stood out as significant, this would possibly indicate the need for follow-up analyses.

Significant differences in treatment efficacy related to gender and age are typically nonexistent (Joutsenniemi et al., 2012; Wierzbicki & Pekarik, 1993), while severity of depression tends to covary positively with treatment efficacy in most trials (Garfield, 1986; Lambert, 2001). This finding is not universal, however, as a more recent study found no such correlation (Driessen et al., 2010). In addition, the number of therapy sessions and treatment effects largely seem to follow a dose-response relationship (e.g., Howard et al., 1986), although a newer study did not support this finding (Cuijpers et al., 2013). Recent research has also shown that experienced psychologists achieve better treatment results than trained students regarding individual CBT (Johnsen & Friborg, 2015). Although there have been few attempts to examine this variable thoroughly within a GCBT context, it is conceivable that the same relationship holds for the group therapy format as well, particularly because it requires additional competence and training in group therapy in general.

Because a recent meta-analysis showed a different rate of historical decline for CBT-studies conducted inside versus outside the United States, with effects falling with time only in the United States (Cristea et al., 2017), this study investigated whether there were any systematic differences in ES related to country among the current GCBT studies.

The final variable investigated in this study was the application of a treatment manual. For the individual format of CBT, trials have largely used manuals of some form, with most studies preferring Beck's original treatment manual (Beck et al., 1979). However, in two large, recent meta-analyses, no differences in ESs were found between CBT trials using a Beck manual versus studies using another treatment manual or no manual (Cristea et al., 2017; Johnsen & Friborg, 2015). For GCBT, there are large variations regarding application of a treatment manual. A substantial number of the included RCTs did not use a manual (or did not report it), while a significant number of studies designed and utilized their own unique manuals. Approximately one third of the studies used treatment manuals that had previously been tested and validated in empirical trials. Our research question asked whether the use of a treatment manual is related to effect sizes for GCBT for depression.

Objectives of the Present Study

The aim of the present study was to examine whether there is a decline in effect size of GCBT for depression with advancing publication year, as was observed for individual CBT (Johnsen & Friborg, 2015). A second purpose was to investigate moderators of the reported effect sizes of GCBT. Meta-analyses are considered the best available method for such evaluations, as well as for examining time-trends. The rationale for the current study is thus to provide information and updates with regards to the development of effect sizes for GCBT over time.

Method

Data Collection, Studies, and Selection Criteria

The OvidSP Internet-based platform was used to locate empirical English language randomized controlled trials. The searches were conducted in October 2015 using the following databases: PsycINFO, APA PsycNET, Embase, and Ovid Medline. The first query "Group *AND* depression *AND* cognitive" resulted in 10,560 hits. The second query "depression *AND* study," followed by "depression *AND* treatment," and "depression *AND* efficacy *OR* efficacious," yielded 5,987, 1,965 and 4,353 hits, respectively. A third query "depression *AND* trial *AND* cognitive," and "group *AND* cognitive therapy" yielded 1,793 and 1,821 hits, respectively. In total, all queries returned 26,479 studies. By examining their titles, the abstracts of 934 papers were read by the first author to judge their relevance. Following that review, 181 papers were obtained from the university library. The following exclusion criteria were then applied (see Figure 1).

(1) The delivered therapy was not cognitive therapy, (2) a unipolar depressive disorder was not the primary Axis I psychiatric diagnosis, (3) participants were not adults (mean age <18), (4) therapy was not delivered by a therapist trained in CBT, (5) the psychotherapeutic intervention was not intended to treat depression, (6) the outcome was not measured with the BDI or the HRSD, (7) patients had acute, or life-threatening, comorbid physical (e.g., terminal cancer) or mental disorders (e.g., active psychosis), (8) treatment was not delivered as group therapy, (9) the trials were not randomized and controlled, and (10) the patients had a BDI/HRSD prescore below 13. The last criterion complies with the manual of the BDI-II and with the opinions of researchers in the field of depression treatment research (Beach & O'Leary, 1992; Kendall et al., 1987; Murphy et al., 1995; Wright et al., 2005). The selection procedure was conducted by the first author and yielded a final study pool of 37 papers.

Coding of Study Information, Moderator Data, and Moderator Analysis

The following data were coded from the papers: demographic information (gender and age), year of implementation of the intervention, country in which the intervention took place (United States vs. the world), duration of treatment (number of sessions), type of therapist (psychologist or trained student), and information about the severity of the diagnosis (mild, moderate, severe, or recurrent depression). The depression diagnoses of the patients were set according to the original authors' definitions. If unreported, we categorized the diagnoses based on the BDI prescores as mild (13–19.5), moderate (20–29.5), or severe (>30). The moderator "manual" was coded dichotomously, according to whether a set manual was followed. In some studies, the manual had been previously tested or validated, either by pilot studies or by its use in other trials investigating treatment efficacy. Many studies used manuals designed exclusively for their trial (Table 1). Effect sizes were examined as to whether they covaried with any of the moderator variables listed previously. Because the number of trials using HRSD scores was low, no subgroup analyses were performed for this outcome measure.

The Randomized Controlled Trial Psychotherapy Quality Rating Scale (RCT-PQRS; Kocsis et al., 2010) was used to rate the

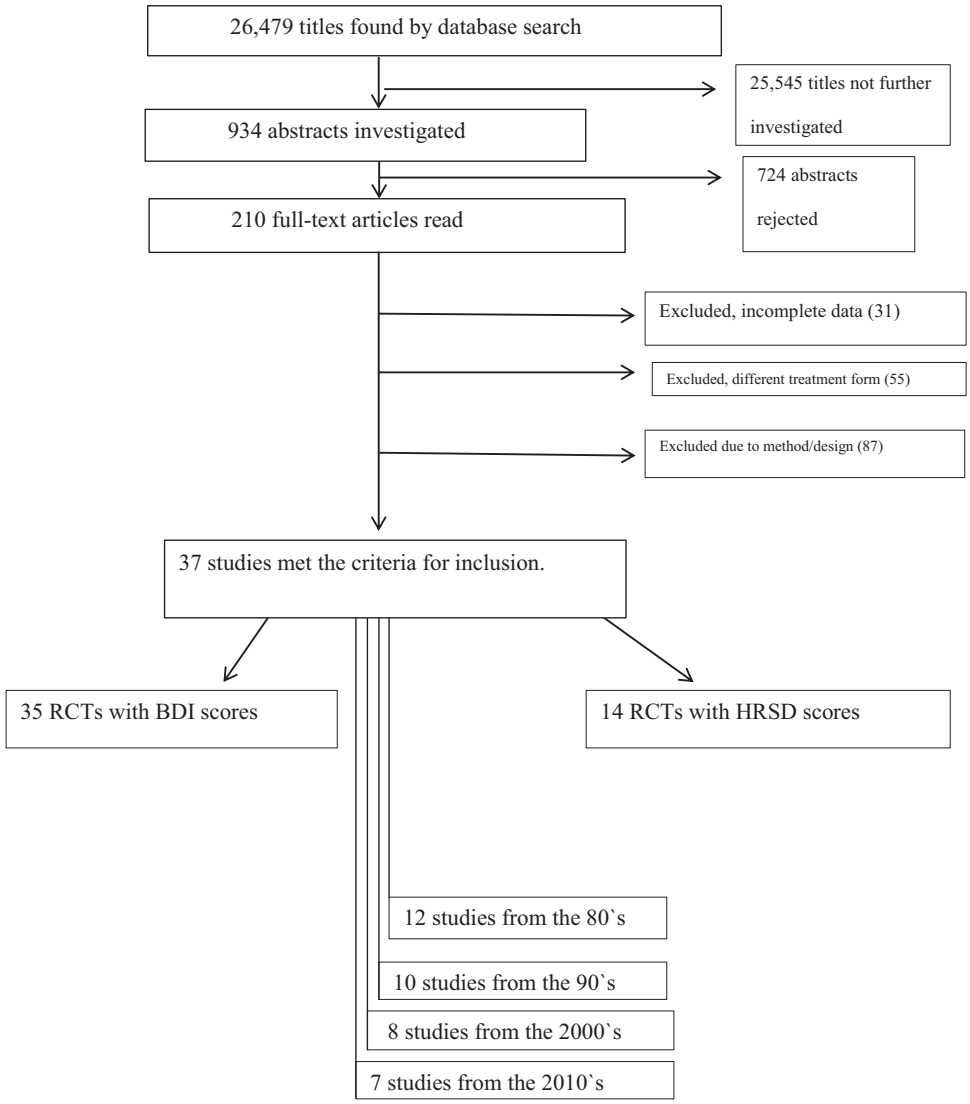


Figure 1. Flowchart of the search and selection procedure.

methodological quality of the published studies. The RCT-PQRS is a comprehensive instrument consisting of 24 items measuring six dimensions of study quality. Each item is assigned a score of 0 (*poor description, execution, or justification of a design element*), 1 (*brief description or either a good description or an appropriate method or criteria set, but not both*), or 2 (*well described and executed, and, where necessary, justified design elements*). The scale yields a total score ranging from 0 to 48, which was used in a subsequent metaregression analysis. A total quality score of 24 or higher indicates a minimally adequate level of quality (Thoma et al., 2012). Coding of all variables was completed by the first author.

ESs

Two procedures were used when calculating ESs in this analysis, both reliant on the BDI or the HRSD pre-/posttreatment scores: The primary analysis utilized change scores based on a within-study design (all studies), and the secondary analysis utilized change scores available from controlled trial designs (a subset). Although all studies in the present meta-analysis were originally RCTs, several of the RCT studies only included other treatment comparison groups (e.g., another psychotherapy group or medication treatment arm), and hence did not include a no-treatment control group (e.g., a wait-list group). These studies were coded as within-study designs. The remaining studies, which included a no-treatment control group, were thus coded as controlled trials in our analysis. In addition, these studies were also coded as within-studies and added to the separate within-study design pool. Thus, the two statistical conditions in this meta-analysis were kept separate from each other when calculating results.

For the first analysis, which included all 35 RCTs using the BDI, a standardized mean difference (SMD, or Cohen's *d*) was calculated for the intervention group (M_{pre}–M_{post}, divided by the

Table 1
A Descriptive Overview of the 37 Group Cognitive–Behavioral Therapy (GCBT) Studies Included

Study/author GCBT manual	Year	Sample	*N*	ES BDI (HRSD)	Sessions
Propst Own design	1980	Student	11	.80	8
Shaffer Own design	1981	Clinical	10	.89	10
Rush No	1981	Clinical	28	1.05	19
Kornblith Tested/validated	1983	Women	11	.65 (1.35)	24
Brown × Own design	1984	Advert.	25	1.26	24
Steuer Tested/validated	1984	Mixed	10	1.25 (2.66)	37
Baker No	1985	Advert.	31	1.65	14
Teri Tested/validated	1986	Clinical	47	1.60	24
Wierzbicki × Own design	1987	Advert.	9	.45	12
Rehm Own design	1987	Advert.	34	2.40 (1.93)	15
Hogg × No	1988	Student	13	1.83	16
Zettle No	1989	Advert.	10	.94 (.97)	22
Neimeyer No	1990	Advert.	24	.97 (1.40)	20
Wilson Tested/validated	1990	Inmates	5	.95	28
Wollersheim × Own design	1991	Advert.	10	.38	20
Beutler Tested/validated	1991	Mixed	21	1.23	20
Brand × Tested/validated	1992	Elderly	27	.89 (2.05)	16
Zettle No	1992	Advert.	14	1.91 (1.97)	20
Arean × Own design	1993	Elderly	19	1.20 (2.60)	24
Stravynski No	1994	Clinical	9	2.32 (1.96)	30
Bright Tested/validated	1999	Advert.	27	1.26 (1.44)	20
Oei Tested/validated	1999	Clinical	46	1.64	20
Rokke × Own design	2000	Mixed	18	1.27 (.69)	18
VanDam × Tested/validated	2003	Adverts	61	.87	24
Arean Tested/validated	2005	Elderly	17	(.25)	18
Hamamci × Tested/validated	2006	Student	10	1.98	12
Rohan × Own design	2007	SAD	13	**1.56** (2.22)	20
Faramarzi × No	2008	Women	29	1.64	10
Wong × Own design	2008	Clinical	48	.87	20
Hamdan × Own design	2009	Student	44	2.12	10
Hegerl Tested/validated	2010	Clinical	61	(1.10)	10
Manicvasgar Tested/validated	2010	Mixed	26	**.79**	16
Hunter No	2012	Addicts	140	**1.85**	36
Mokrue × No	2013	Student	54	**2.17**	8
Zamarinejad × No	2014	Student	9	**1.70**	16
Teismann × Own design	2014	Mixed	31	**1.49**	22
Milgrom Tested/validated	**2015**	**Postnatal**	**14**	**1.56**	**12**

Note. ES = effect size; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; Advert. = Advertisement. ES = Hedge's *g*. Studies with an asterisk are included in both the first and second analysis. ES in bold = BDI II as primary measure. *Authors* in italics = studies conducted outside the United States.

standard deviation of the change score). A Hedge's *g* correction was applied to the SMD, which reduces the SMD for studies having small sample sizes (Hedges & Olkin, 1985). All interventions consisted of GCBT that used a RCT design.

For the second analysis, which consisted of the 16 RCTs that included a no-treatment control group, the effect sizes were calculated from the difference between pre- and postscores on the BDI for the GCBT group and the no-intervention group, respectively, and then standardized using the change scores.

Standardization by change scores is the preferred choice when the aim is to measure change relative to preintervention scores (Kulinskaya et al., 2002). This procedure is often used in meta-analyses quantifying treatment effects (e.g., Abbass et al., 2013; Kishi et al., 2012; McGuire et al., 2014; Watts et al., 2013; Zoogman, Goldberg, Hoyt, & Miller, 2015). Because meta-analytic experts recommend imputing conservative prepost correlation for studies that do not report this (Balk, Earley, Patel, Trikalinos, & Dahabreh, 2012; Rosenthal, 1993), $r = .7$ was set for all such papers.

For the third analysis, the same procedure as for the first analysis was used (a within-study design), but this time the outcome measure was the HRSD.

Quantitative Data Synthesis and Statistical Calculations

The Comprehensive Meta-Analysis (CMA) software, version 2 (Borenstein et al., 2005) was used for all statistical analyses.

The average weighted effect sizes were estimated according to a random effects model, under the assumption that the true effect sizes would vary among studies because of the study-related factors. A Q test statistic (chi-square distributed) was calculated to examine whether the variance between studies was larger than the variance within studies, thus indicating that predictors (or moderators) might explain this between-study variation. Metaregression analyses were

used to estimate the coefficient of continuous moderator variables (such as publication year, patients' age, or study quality ratings). The unrestricted maximum likelihood method was used to calculate the *p* value for the coefficient, as it assumes an underlying random distribution of effect sizes. The moderator analyses for the categorical variables were based on a Q-test statistic, which examines whether the variability between categories (subgroups in the study) is larger than the variability within subgroups.

Heterogeneity, Publication Bias, and Identification of Outliers

Heterogeneity was calculated as I^2. This is an intuitive and simple expression of the inconsistency of the studies' results. The I^2 statistic aims to describe the percentage of variation across studies that is because of heterogeneity rather than chance (Higgins et al., 2003). Levels of 25%, 50%, and 75%, correspond, respectively, to low, medium, and high heterogeneity.

To identify any publication bias or undue outliers, visual inspections of the funnel and forest plots were performed, and Duval and Tweedie's trim-and-fill method was used.

Results

Studies and Participants

The search procedure yielded 37 eligible RCT studies conducted during the period 1980–2015 (average = 1996). The total number of patients was 1,016 (a descriptive overview is given in Table 1). The number of patients in the studies varied between five and 140 (M = 27, SD = 25). Males accounted for 24.3% of the patients, and the average age was 40.2 years (SD = 13.5). The mean prescore on the BDI was 24.2 (SD = 5.3). The patients received, on average, 18 sessions (one session = 45 min) of group therapy (SD = 6.4, range = 8–37). The methodological quality of the studies (the RCT-PQRS ratings) varied from 19 to 35 (M = 26, SD = 4.2).

Efficacy of Group CBT and Historical Time Trends

The average weighted BDI effect size for all studies, based on within ESs (k = 35), was g = 1.33 (95% confidence interval [CI] = 1.16 to 1.50). For the between-groups condition, the average BDI effect size (k = 16) was g = 1.14 (95% CI = .84 to 1.44). For the HRSD, using a within-group design (k = 14), g was 1.56 (95% CI = 1.20 to 1.90). Forest plots are found in Figures 2 and 3.

For the first analysis (a within-study design), the GCBT effect sizes based on all 35 RCT studies improved significantly with time as measured by the BDI (p = .02; Table 2; Figure 4). For the second analysis (the between-groups condition with the BDI as outcome measure), the ESs based on the 16 studies that included a no-intervention control group also improved significantly across time (p < .001; Figure 5). For the third analysis (a within-group design), effect sizes based on the 14 studies using the HRSD showed a nonsignificant trend toward a decline with the passage of time (p = .07; Table 2 and Figure 6).

Because the majority (k = 28) of the included studies used the outcome measure BDI-I, separate analyses were conducted on this measure exclusively. The results substantiated the finding of a positive improvement in time for the between-groups, controlled condition (p < .01). For the within-group condition, the metaregression analysis showed a nonsignificant, but still positive trend (p = .15; Table 2).

Because of the seemingly inconsistent results from the BDI and HRSD analyses, we decided to investigate this discrepancy further. For this analysis, the 12 studies with ESs from *both* outcome measures were included in a separate metaregression reanalysis based on within-group scores from the HRSD. The analysis now revealed a flat regression line (b = −0.01, p = .73; Table 2).

Publication Bias, Outliers, and Heterogeneity

The funnel plot based on the BDI seemed symmetrical (Figure 7), thus indicating little risk of publication bias or outliers. This observation was confirmed by Duval and Tweedie's trim-and fill method, where no adjustments to the original sample were made.

For the HRSD, the plot seemed slightly skewed downward right (Figure 8), indicating possible publication bias in the form of more studies with few participants having higher ES. The use of Duval and Tweedie's method partly confirmed this by identifying and trimming one study. As a consequence, the sample was reanalyzed without this possible outlier, without any substantial effect on the metaregression line.

Heterogeneity scores were in the moderate-to-high range for all analytic conditions, ranging from 46% to 84%. As could be expected, the analyses yielding the lowest levels of I^2 were found in the controlled groups design. The large heterogeneity scores indicated that it was appropriate to move on with the analysis of potential moderators.

Other Moderators

A separate moderator analysis, based on the BDI scores, for the variables study quality, age, sessions of therapy, and gender did not reveal any significant differences in treatment effects (Table 2), nor did the variables type of therapist, version of BDI (I or II), or diagnostic severity covary significantly with ES (Table 3).

For the moderator variable treatment manual, the analysis showed higher ESs for studies where no specific treatment manual was followed (p = .03). Furthermore, additional analyses showed that for trials *not* using a manual, there was a significant improvement in ES with time (p < .01). This relationship between ES and time was not found for trials that used a manual (Table 2).

The findings warranted further investigation. This time it was determined to investigate the relationship between study quality and the use of a manual. The analysis revealed that there were no differences in study quality related to the treatment manual (k = 25, M = 26, SD = 4.6) and the no manual (k = 11, M = 25, SD = 3.2) statistical conditions; $t(33)$ = .67; p = .51.

For the variable country, there was a tangible but nonsignificant tendency toward higher ESs for studies performed outside the United States compared with studies performed in the United States (p = .06, see Table 3). Separate follow-up analyses revealed that trials conducted exclusively in the United States had a nonsignificant tendency toward improvement in ES with time (p = .06), while trials conducted in the rest of the world showed no signs of improvement with time (Table 2).

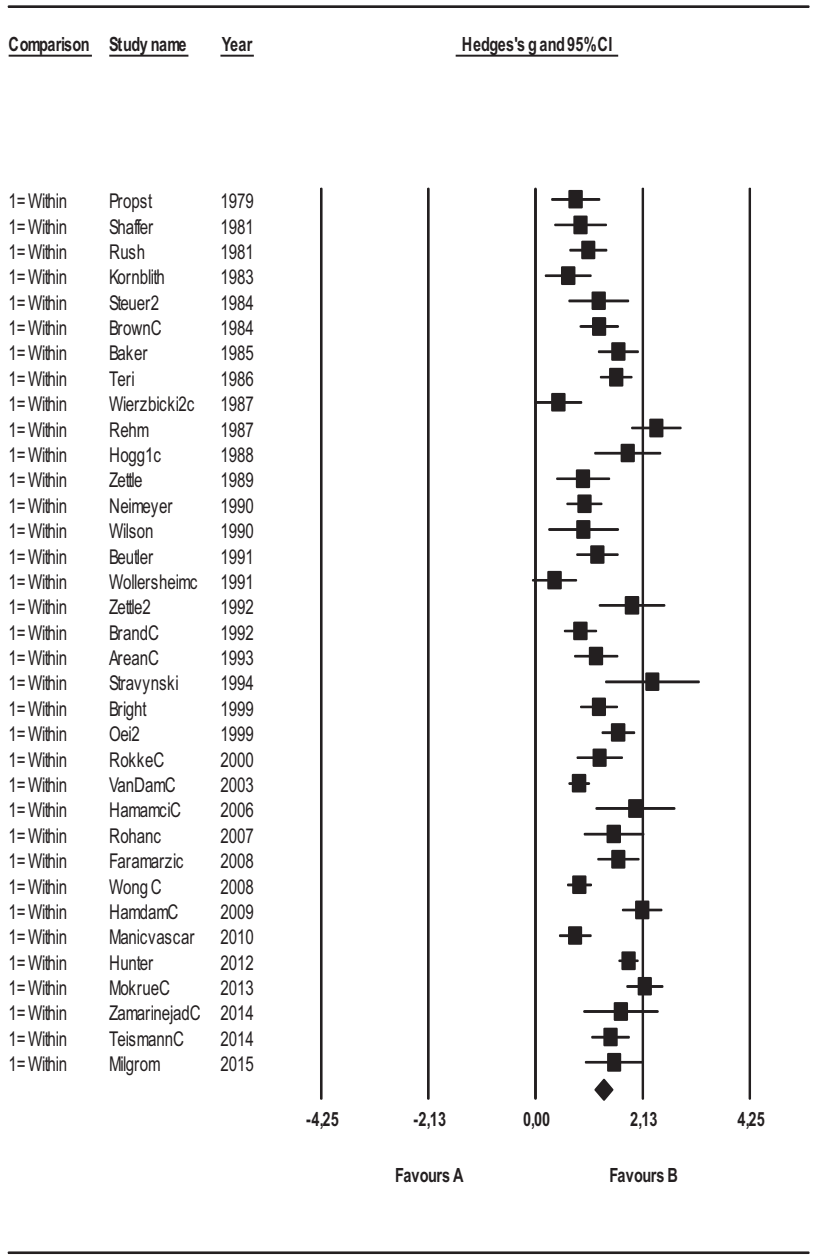


Figure 2. Forest plot for the Beck Depression Inventory (BDI) effect sizes in the within group condition.

Discussion

The Temporal Development of Treatment Effect

The main objective of the present meta-analysis was to examine the temporal evolution of the treatment efficacy of GCBT for unipolar depressive disorders. The results revealed a somewhat mixed picture. First, for three out of four comparisons, the analysis showed a significant increase in effect size across time, as measured by BDI. The last test also had a tendency—albeit a nonsignificant one—toward improvement. Perhaps more important, the two analyses considered most methodologically robust, based on a controlled group design, showed the largest and strongest improvements. The first was an analysis with a between-group design, exclusively consisting of studies with scores from the BDI-I. The second was a between-group controlled design consisting of studies with scores on both the BDI-I and the BDI-II. These two analytic conditions also seemed to have an acceptable comparability across the included trials, as judged by the moderate I^2 levels of

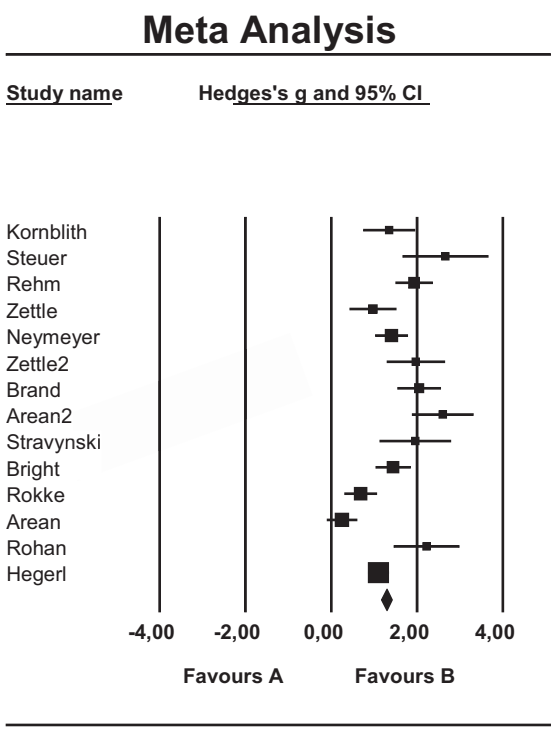


Figure 3. Forest plot for the Hamilton Rating Scale for Depression (HRSD) effect sizes. Trials are sorted after year of publication, with earliest on top.

.46, and .54, respectively. These I^2 scores were the lowest for all analytic conditions. All in all, the finding of an improvement in treatment effect based on scores from the BDI seems robust.

For the HRSD, the analysis did not confirm the findings of larger ES with time. On the contrary, the results showed a near-significant ($p = .07$) negative trend, indicating a possible decline in treatment effect with time. The discrepancy between the results from the two outcome measures was surprising, given the moderate to high correlation previously found between the BDI and the HRSD. In addition, recent metaregression analyses over the temporal development of antidepressive CBT have revealed a similar direction for the two outcome measures (Cristea et al., 2017; Johnsen & Friborg, 2015).

As a first step investigating the results, checks for any outlying studies based on the HRSD were performed. One study was consequently removed, as indicated by the funnel plot and Duval and Tweedie's trim-and-fill method. This procedure did not change the regression line, which still showed a nonsignificant negative trend. Next, the analysis was restricted to include only the studies with scores on both the HRSD and BDI. The result was now a very flat regression line, indicating no trend toward a change in ES over time. Furthermore, the analysis based on the HRSD presented the largest heterogeneity of all statistical conditions in the study, with an I^2 of 82.9. Lastly, this analysis also consisted of the fewest studies. Taken together, there is a case for a nonsignificant relationship between time and ES, as found in the initial HRSD-based analysis not being reliable. There are several possible reasons for this finding, for example, a truly nonsignificant effect in the population, error variance causing high heterogeneity, or the effects of other unknown variables.

Last, as described earlier, the BDI differs from the HRSD by focusing on cognitions (Wampold & Imel, 2015). Thus, divergent findings for the BDI and HRSD may be because of differences in content between the two instruments, suggesting that the increase in ES over time for the BDI could reflect larger treatment effects on depressive cognitions.

The BDI is the symptom checklist of choice for cognitive behavior therapists, and the analysis based on the BDI included over twice as many papers as did the HRSD. The BDI subgroup analytic conditions had lower percentages of heterogeneity and

Table 2
A Metaregression Analysis Examining the Association Between Continuous Moderators and Effect Sizes

Moderator variable	*K*	b_0	b_1	95% CI	Z (b_1)	I^2	*p*
Time, within group BDI	35	−30.63	.0160	.002, .030	2.23	81.2	.03
Time, between-group BDI	16	−105.34	.0532	.029, .077	4.46	54.9	<.001
Time, within HRSD	14	69.06	−.0338	−.073, .005	−1.71	82.9	.07
Time, common HRSD/BDI[a]	12	19.41	−.0090	−.059, .042	−.34	78.4	.73
Time, BDI I, between	12	−84.00	.0420	.016, .069	3.17	46.5	<.01
Time, BDI I, within	28	−28.71	.0150	−.001, .036	1.43	79.9	.15
Time, trials from United States	22	−40.99	.0212	−.001, .043	1.88	73.2	.06
Time, all other countries	12	−.01	9.8881	−.037, .029	−.25	86.4	.80
Time, no manual[b]	11	−42.47	.0221	.007, .037	2.88	76.3	<.01
Time, manual used[b]	24	−23.74	.0125	−.005, .030	1.39	79.9	.16
Age of participants	32	1.72	−.0117	−.026, .002	−1.65	79.3	.10
Gender (male %)	32	1.31	.0008	−.008, .009	.19	82.1	.85
Study quality (0–48)	35	1.79	−.0176	−.058, .023	−.40	82.1	.39

Note. b_0 = intercept (year 0 A.D.); b_1 = time slope (change coefficient); CI = confidence interval; Time = publication year; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression.
[a] In this conditions, all trials that included scores from both the HRSD and the BDI were included in a within-group design. [b] For these analytic conditions, trials using a treatment manual were separated from trials not using a manual, and analyzed separately with time (year of study) as the predictor. The BDI was the outcome measure.

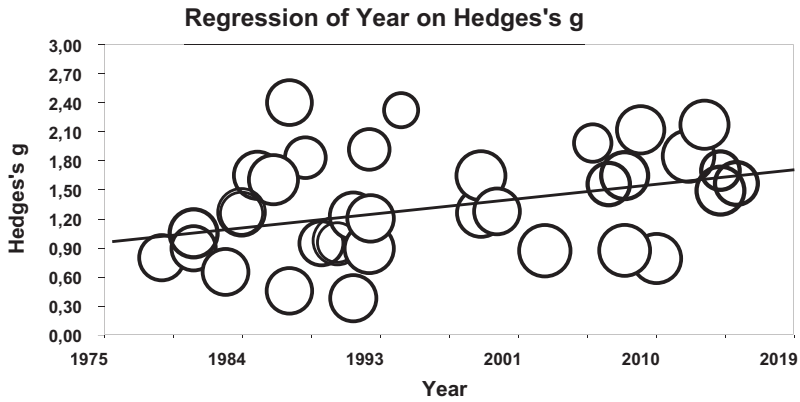


Figure 4. The plot portrays the positive change ($p < .001$) in Beck Depression Inventory (BDI) effect sizes across time for the first analysis ($k = 35$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

also a more symmetrical funnel plot, without any substantial deviations as presented by the Duval and Tweedie's trim-and-fill method. It can therefore be concluded that the most reliable and consistent findings in this analysis are those associated with the BDI.

Moderators of Treatment Efficacy

Consistent with previous meta-analyses (e.g., Cuijpers et al., 2008), the current study showed a strong within-group effect size of GCBT ($g = 1.33$), comparable to the ES of individual CBT as reported by Johnsen and Friborg (2015). In addition to publication year, the present study investigated several moderators of treatment effect that had been included in previous meta-analyses. The results showed that the variables diagnostic severity, age, gender, type of therapist, study quality, and sessions of therapy were not related to treatment efficacy. Because no group therapy studies included fewer than eight sessions, it was impossible to say whether the efficacy of brief GCBT (<8 sessions) would be different from the present selection.

This analysis found a negative relationship between the application of a treatment manual and outcome. Trials not following (or reporting) a manual had a significantly larger ES than those that did report use of a manual. This finding is consistent with previous studies that have found negative effects of the use of treatment manuals (Duncan & Miller, 2006). It should also be noted that two recent meta-analyses concerning the individual format of CBT (Cristea et al., 2017; Johnsen & Friborg, 2015) found no differences in ES between studies using the original Beck manual versus studies that did not use it.

Furthermore, subgroup analyses showed that for trials where no manual had been used, there was an increase in treatment effect sizes with the passage of time. However, for trials using a treatment manual, there was no increase with time. This finding, not explained by differences in study quality, could be interpreted in a number of ways, but indications are that in regard to treatment efficacy for GCBT, the specific factors (techniques applied in treatment) do have a substantial bearing on treatment outcome; when following a set routine, treatment effects do not improve with time. The standardiza-

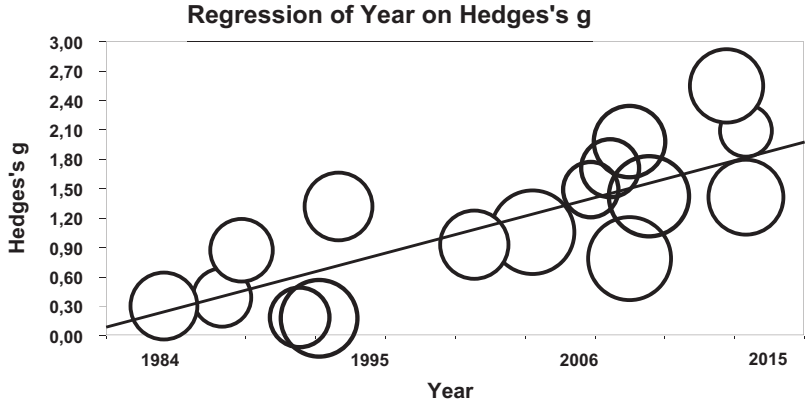


Figure 5. The plot portrays the positive change ($p < .001$) in Beck Depression Inventory (BDI) effect sizes across time for the second analysis ($k = 16$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

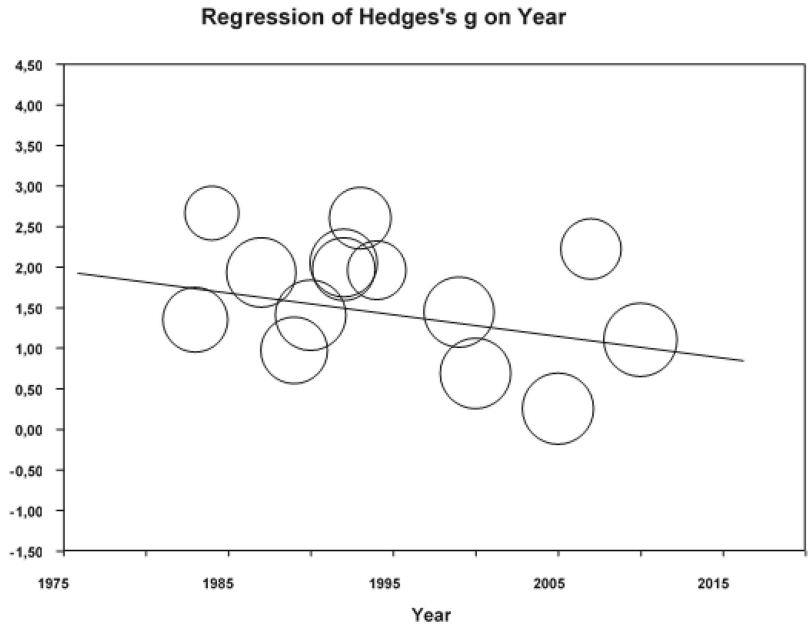


Figure 6. The plot portrays the development ($p = .07$) in Hamilton Rating Scale for Depression (HRSD) effect sizes across time ($k = 16$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

tion of treatment likely results in less variability in treatment effects across time. On the other hand, when varying the approach, ESs increase. The results lend support to the importance of updating treatment procedures and manuals so that they are in touch with general developments in society. This hypothesized cultural and social "goodness of fit" for psychotherapy is a concept that is attracting increased interest from both researchers and commentators (e.g., Burkeman, 2015). Because cultural shifts and developments (especially in the western world) have occurred at a rapid pace during the last decades, it seems wise to frequently update treatment manuals in accordance. Such a procedure could also facilitate another contributor to treatment effect, namely the placebo effect. Hypothesized as a

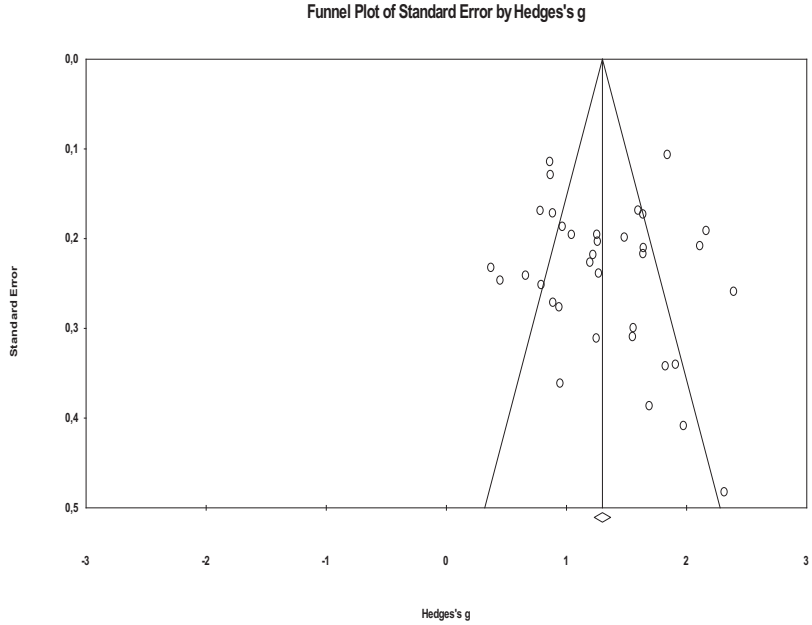


Figure 7. A funnel plot for the 35 studies using the Beck Depression Inventory (BDI) as measure for effect size (ES).

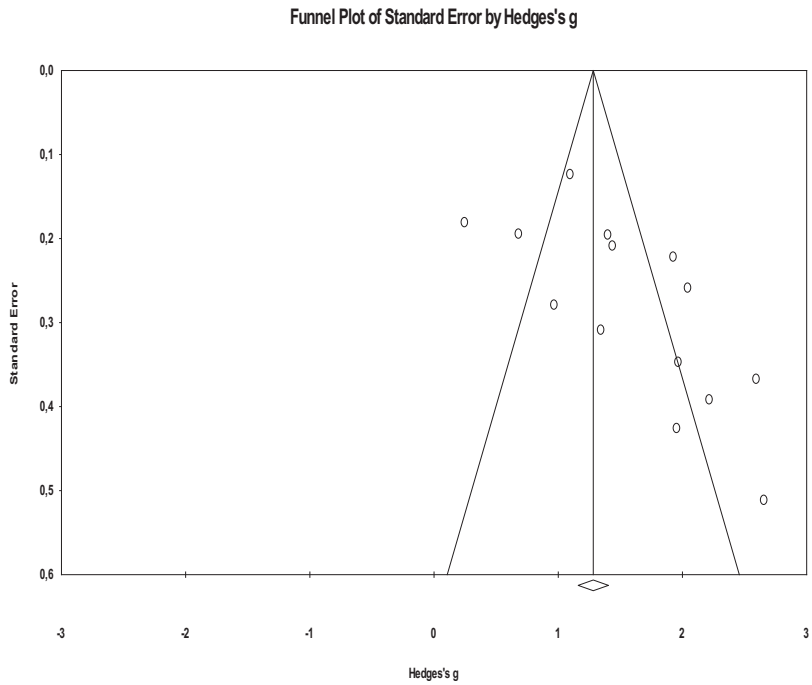


Figure 8. A funnel plot for the 14 studies using the Hamilton Rating Scale for Depression (HRSD) as measure for effect size (ES).

possible reason for the decline in individual CBT (Johnsen & Friborg, 2015), the loss of a placebo effect with time could, to some degree, be alleviated if manuals were updated and adjusted at timely intervals. Indications suggest that a substantial reason for the observed decline in ES over time for individual CBT could be related to the use of a manual from the 1970s for a population from the 2000s. The variables "time" and "manual" thus seem to have a complex effect on each other, where the importance of updating the latter (or being flexible as a therapist), becomes more evident with the passage of time. This is a topic where further research is necessary.

Table 3
A Subgroup Analysis of Dichotomous Variables and Effect Sizes Based on the Beck Depression Inventory (BDI)

Moderator	*k*	*G*	95% Confidence interval	Q_{df}	*p*	I^2
Diagnostic severity				$.26_2$	.88	79
Mild	9	1.18	.77–1.60			79
Moderate	21	1.27	1.07–1.41			69
Severe	5	1.38	.69–2.07			92
Country of trial				3.40_1	.06	82
World	13	1.54	1.26–1.83			86
United States	22	1.20	.98–1.42			80
Manual used				5.03_1	.03	82
No	11	1.60	1.31–1.89			76
Yes	24	1.21	1.01–1.77			80
Version of BDI				2.38_2	.12	82
BDI I	28	1.26	1.08–1.45			79
BDI II	7	1.58	1.22–1.95			84
Type of therapist[a]				$.71_1$	.40	79
Trained student	8	1.13	.78–1.48			73
Psychologist	17	1.33	1.03–1.65			82

Note. Q_{df} = Q value for the between group difference(s); *df* = associated degrees of freedom; I^2 = the degree of between study variance relative to total variance.
[a] The remaining studies ($k = 10$) used a combination of therapists, or type of therapist was not reported. These studies were excluded from the analysis.

The analysis also found a trend approaching significance ($p = .06$) related to country of origin, with trials conducted outside the United States showing higher ESs. Follow-up subgroup analyses showed that the ESs of studies conducted in the United States increased with time, an indication that was not evident for trials conducted in the rest of the world. With time, however, studies performed in the United States seem to have improved, reaching a level of efficacy comparable to that of trials conducted in the rest of the world. This development is interesting, particularly when considering a similar, yet opposite, finding from another recent meta-analysis concerning individual CBT (Cristea et al., 2017). In that study, ESs from studies performed in the United States decreased with time, while ESs for trials from the rest of the world remained unchanged. However, the authors concluded that the observed effects most likely were spurious, mainly because of high levels of heterogeneity. Nevertheless, the common connotation for both CBT and GCBT seems to be that the largest temporal developments in treatment effects, regardless of direction, are limited to the United States. One explanation for this finding could be that the range of publication years is larger for trials conducted in the United States, thus increasing the potential for significant results to occur.

Potential Reasons for the Main Finding

The primary objective of this study was to uncover temporal trends related to treatment effects. Efforts to fully explain the

observed improvements are mainly limited to exploratory hypotheses.

Regarding the previously identified temporal decline in ES for individual CBT, a recently published paper tied some of the decline to possible differences in the control groups for newer versus older trials of CBT (Dobson, 2016). Another study linked the decline to a lack of training, competency, treatment fidelity and/or adherence to cognitive principles by modern cognitive–behavioral therapists (Waltman et al., 2016). It is unclear whether these aspects would apply to the present finding regarding GCBTs increase in treatment effect. However, when looking at the different results in this analysis, there seems to be a beneficial effect in varying and adapting the GCBT interventions, perhaps according to the patient group. For GCBT, there is a wide variety regarding applying a manual, with no single manual being universally embraced or utilized across time. Researchers frequently develop their own manuals before the start of treatment, and many do not follow a set manual at all—with success to boot. This tendency of utilizing different manuals contrasts with individual CBT, where a large number of studies use the original Beck manual for CBT for depression (Beck et al., 1979). However, as shown by recent research, this approach does not seem too beneficial to the individual format either. Treatment effects are decreasing, and there are no signs of a beneficial effect from using the original manual (Cristea et al., 2017; Johnsen & Friborg, 2015).

It is important to keep in mind that all the included trials in this meta-analysis adhere to the cognitive principles and use traditional cognitive and behavioral techniques and approaches. Perhaps the key to further improvement in ES for GCBT (and individual CBT) lies in moving away from using highly structured treatment manuals (which could lead to rigid treatment), thus paving the way to a more creative, adaptable, and intuitive way of conducting GCBT? The hallmarks of an efficient group cognitive–behavioral therapist could very well be the ability to use therapeutic creativity and adaptability, yet within a framework consisting of high levels of competency, treatment fidelity, and adherence to general cognitive–behavioural principles and techniques. This notion is consistent with previous research regarding the association between flexibility and the use of a manual (Kendall, Chu, Gifford, Hayes, & Nauta, 1998). More recent findings also suggest that therapist flexibility is associated with better treatment outcomes (Owen & Hilsenroth, 2014).

Limitations

The self-report inventory BDI comes in two forms, the BDI-I and BDI-II. For some analytic conditions, these were treated as interchangeable. Although the two versions are very similar, there is still a possibility of a confounding effect resulting from this mixture. In addition, as the BDI and the HRSD do not measure improvement on a more general or global level, one cannot reliably generalize the effects to areas beyond symptom relief.

The analysis included a number of trials with few participants, which could be a possible confounder. Although the Hedge's g calculation of ES alleviates this concern to a large extent, it is not inconceivable that trials with very few participants are still subject to somewhat unreliable effect sizes.

Heterogeneity could be a concern for some subgroup analyses, with relatively high percentages noted for most of them. However, this is not an unusual phenomenon for published meta-analyses in the field of social and medical sciences, where about a quarter have I^2 scores above 50% (Higgins et al., 2003). Furthermore, it is important to emphasize that quantification of heterogeneity is only one component of a wider investigation of variability across studies, the most important being diversity in clinical and methodological aspects. This especially holds true for analyses based on small samples. Thus, to assume any systematic findings as random or spurious, there should be other clear indications present—as could be argued was the case for the analytic condition based on the HRSD.

When conducting meta-analyses based on prepost within-Group ESs, there is an inherent risk of bias in the form of data dependence, also known as type-1 error (Cuijpers et al., 2017). This risk would also apply to the current analysis. However, data dependence is not an issue with regards to between-groups calculated ESs. This meta-analysis uses both of these procedures to calculate effect sizes, revealing similar results for both the within group and between-groups analytic conditions. This result suggests that data dependence does not have an undue effect on the results. Furthermore, even if an issue of data-dependence did exist, this would mainly affect the overall estimated ES of treatment. Thus, there is little reason to assume that the time-trend results, which are the main focus of this study, would be differentially affected by the issue of Type-1 error.

Unfortunately, reliable measures of adherence or treatment fidelity were almost nonexistent for the included studies, making it impossible to perform any informative analyses in that regard. Finally, the current study was exploratory in nature, and no hypotheses were posed beforehand, which reduces the certainty with which one can draw conclusions.

Implications and Conclusion

The practical significance of this systematic review is to contribute to a heightened awareness and understanding of the current trends in group psychotherapy that applies CBT principles and to encourage further research on the topic of the temporal evolution of treatment effects in psychiatry in general. The fact that GCBT shows substantial ESs, comparable to the effects of individual CBT, may imply that this form of therapy represents a particularly efficient approach to managing treatment needs in the years to come. Considering the beneficial cost-effectiveness ratio of group therapy, this avenue should be developed further. Recent meta-analyses have shown that the monetary benefits involved in the application of Group CBT does not seem to compromise treatment efficacy significantly, as the outcome differences between individual and Group CBT trials are slight to moderate (Hans & Hiller, 2013; Huntley et al., 2012). However, these reviews do not take into consideration the fact that Group CBT seems to have increased in effect size over the years, while the opposite is found for individual CBT.

The results of this analysis also indicate that a shift in perspective should be considered in regard to the most efficient way of implementing GCBT. The highly structured manuals could be replaced by more adaptable forms of GCBT, or at least by frequently updated manuals. Further research focusing on the association among effect sizes, time-trends, and the use of a treatment manual is warranted.

Résumé

Cette méta-analyse se penche sur les changements temporels (tendances dans le temps) des effetsde la thérapie comportementale et cognitive de groupe (TCCG) utilisée pour traiter la dépression unipolaire. Cette étude exploratoire examine 37 études (regroupant1016 patients) menées entre 1980 et 2015 et quantifie les effets observés (EO) dans le cadre de celles-ciselon la valeur g de Hedge en fonction de l'inventaire de dépression de Beck (IDB) et de l'Échelle de dépression de Hamilton(HAM D). Dans les analyses principales, des comparaisons au sein du groupe (prépostérieures) (k _ 35 pour l'IDBet k _ 14 pour la HAM D) et des comparaisons sous contrôle (entre les groupes) (k _ 16 pour l'IDB) ontété menées. Les EO moyens observés au sein du groupe étaient de 1,33 (intervalle de confiance [IC] de 95 %_ 1,16 –1,19) en fonction del'IDB et de 1,56 en fonction de la HAM D (IC de 95 % _ 1,20 à 1,90). L'année de publication de chaque étude a été analyséeen tant qu'indicateur linéaire de méta-régression des EO. Les résultats ont montré une hausse significative des effets observésen recourant à la TCCG au fil du temps, l'IBD étant utilisée comme mesure des résultats. Toutefois, ceci n'a pas été observé pour la HAM Den tant que mesure des résultats. Des analyses de sous-groupes ont été effectuées sur des variables modératrices sélectionnéesafin de déterminer la présence de covariations en fonction des effets observés. Ces analyses ont révélé que, lors des essais menéssans observer un manuel établi, les effets observés augmentaient au fil du temps, une association qui n'a pas étérelevée pour les essais avec manuel. Les causes et conséquences possibles sont discutées.

Mots-clés : Thérapie comportementale et cognitive de groupe, efficacité, troubles dépressifs, méta-analyse, manuel sur le traitement.

References

*Articles included in the analysis.

Abbass, A. A., Rabung, S., Leichsenring, F., Refseth, J. S., & Midgley, N. (2013). Psychodynamic psychotherapy for children and adults: A meta-analysis of short-term psychodynamic models. *Journal of the American Academy of Child & Adolescent Psychiatry, 52,* 863–875. http://dx.doi.org/10.1016/j.jaac.2013.05.014

Andrews, G. (2001). Should depression be managed as a chronic disease? *British Medical Journal, 322,* 419–421. http://dx.doi.org/10.1136/bmj.322.7283.419

*Arean, P. A., Perri, M. G., Nezu, A. M., Schein, R. L., Christopher, F., & Joseph, T. X. (1993). Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology, 61,* 1003–1010. http://dx.doi.org/10.1037/0022-006X.61.6.1003

*Baker, A. L., & Wilson, P. H. (1985). Cognitive–behavior therapy for depression: The effects of booster sessions on relapse. *Behavior Therapy, 16,* 335–344. http://dx.doi.org/10.1016/S0005-7894(85)80001-0

Balk, E. M., Earley, A., Patel, K., Trikalinos, T. A., & Dahabreh, I. J. (2012). Empirical assessment of within-arm correlation imputation in trials of continuous outcomes (Methods research report no. AHRQ Publication No. 12(13)-EHC141-EF). Agency for Healthcare Research and Quality: Rockville, MD.

Beach, S. R. H., & O'Leary, D. K. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23,* 507–528. http://dx.doi.org/10.1016/S0005-7894(05)80219-9

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. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588–597. http://dx.doi.org/10.1207/s15327752jpa6703_13

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II: Beck Depression Inventory 2nd ed.* (manual). Psychological Corporation: San Antonio, TX.

Beck, A. T., Steer, R., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100. http://dx.doi.org/10.1016/0272-7358(88)90050-5

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571. http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

*Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology, 59,* 333–340. http://dx.doi.org/10.1037/0022-006X.59.2.333

Blanco, C., Okuda, M., Markowitz, J. C., Liu, S.-M., Grant, B. F., & Hasin, D. S. (2010). The epidemiology of chronic major depressive disorder and dysthymic disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of Clinical Psychiatry, 71,* 1645–1656. http://dx.doi.org/10.4088/JCP.09m05663gry

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

*Brand, E., & Clingempeel, W. G. (1992). Group behavioral therapy with depressed geriatric inpatients: An assessment of incremental efficacy. *Behavior Therapy, 23,* 475–482. http://dx.doi.org/10.1016/S0005-7894(05)80170-4

*Bright, J. I., Baker, K. D., & Neimeyer, R. A. (1999). Professional and paraprofessional group treatments for depression: A comparison of cognitive–behavioral and mutual support interventions. *Journal of Consulting and Clinical Psychology, 67,* 491–501. http://dx.doi.org/10.1037/0022-006X.67.4.491

*Brown, R. A., & Lewinsohn, P. M. (1984). A psychoeducational approach to the treatment of depression: Comparison of group, individual, and minimal contact procedures. *Journal of Consulting and Clinical Psychology, 52,* 774–783. http://dx.doi.org/10.1037/0022-006X.52.5.774

Burkeman, O. (2015). *Why CBT is falling out of favor*. Retrieved from https://www.theguardian.com/lifeandstyle/2015/jul/03/why-cbt-is-falling-out-of-favour-oliver-burkeman

Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., & Strauss, B. (2016). Outcome differences between individual and group formats when identical and nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy, 53,* 446–461. http://dx.doi.org/10.1037/pst0000090

Cristea, I. A., Stefan, S., Karyotaki, E., David, D., Hollon, S. D., & Cuijpers, P. (2017). The effects of cognitive behavioral therapy are not systematically falling: A revision of Johnsen and Friborg (2015). *Psychological Bulletin, 143,* 326–340. http://dx.doi.org/10.1037/bul0000062

Cuijpers, P., Huibers, M., Ebert, D. D., Koole, S. L., & Andersson, G. (2013). How much psychotherapy is needed to treat depression? A metaregression analysis. *Journal of Affective Disorders, 149,* 1–13. http://dx.doi.org/10.1016/j.jad.2013.02.030

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. http://dx.doi.org/10.1037/a0013075

Cuijpers, P., van Straten, A., & Warmerdam, L. (2008). Are individual and group treatments equally effective in the treatment of depression in adults? A meta-analysis. *The European Journal of Psychiatry, 22,* 38–51. http://dx.doi.org/10.4321/S0213-61632008000100005

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences, 26,* 364–368. http://dx.doi.org/10.1017/S2045796016000809

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57,* 414–419. http://dx.doi.org/10.1037/0022-006X.57.3.414

Dobson, K. S. (2016). The efficacy of cognitive–behavioral therapy for depression: Reflections on a critical discussion. *Clinical Psychology: Science and Practice, 23,* 123–125. http://dx.doi.org/10.1111/cpsp.12151

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,* 668–680. http://dx.doi.org/10.1037/a0020570

Duncan, B. L., & Miller, S. D. (2006). Treatment manuals do not improve outcomes. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 140–149). Washington, DC: American Psychological Association.

*Faramarzi, M., Alipor, A., Esmaelzadeh, S., Kheirkhah, F., Poladi, K., & Pash, H. (2008). Treatment of depression and anxiety in infertile women: Cognitive behavioral therapy versus fluoxetine. *Journal of Affective Disorders, 108,* 159–164. http://dx.doi.org/10.1016/j.jad.2007.09.002

Friborg, O., & Johnsen, T. J. (2017). The effect of cognitive–behavioral therapy as an antidepressive treatment is falling: Reply to Ljòtsson et al. (2017) and Cristea et al. (2017). *Psychological Bulletin, 143,* 341–345. http://dx.doi.org/10.1037/bul0000090

Friborg, O., Martinsen, E. W., Martinussen, M., Kaiser, S., Overgård, K. T., & Rosenvinge, J. H. (2014). Comorbidity of personality disorders in mood disorders: A meta-analytic review of 122 studies from 1988 to 2010. *Journal of Affective Disorders, 152–154*, 1–11. http://dx.doi.org/10.1016/j.jad.2013.08.023

Garfield, S. L. (1986). Research on client variables in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 213–256). New York, NY: Wiley.

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49,* 59–72. http://dx.doi.org/10.1016/S0165-0327(97)00199-7

*Hamamci, Z. (2006). Integrating psychodrama and cognitive behavioral therapy to treat moderate depression. *The Arts in Psychotherapy, 33,* 199–207. http://dx.doi.org/10.1016/j.aip.2006.02.001

*Hamdan-Mansour, A. M., Puskar, K., & Bandak, A. G. (2009). Effectiveness of cognitive-behavioral therapy on depressive symptomatology, stress and coping strategies among Jordanian university students. *Issues in Mental Health Nursing, 30,* 188–196. http://dx.doi.org/10.1080/01612840802694577

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23,* 56–62. http://dx.doi.org/10.1136/jnnp.23.1.56

Hans, E., & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology, 81,* 75–88. http://dx.doi.org/10.1037/a0031080

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557

*Hogg, J. A., & Deffenbacher, J. L. (1988). A comparison of cognitive and interpersonal-process group therapies in the treatment of depression among college students. *Journal of Counseling Psychology, 35,* 304–310. http://dx.doi.org/10.1037/0022-0167.35.3.304

Hollon, S. D., Shelton, R. C., & Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology, 59,* 88–99. http://dx.doi.org/10.1037/0022-006X.59.1.88

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41,* 159–164. http://dx.doi.org/10.1037/0003-066X.41.2.159

*Hunter, S. B., Witkiewitz, K., Watkins, K. E., Paddock, S. M., & Hepner, K. A. (2012). The moderating effects of group cognitive–behavioral therapy for depression among substance users. *Psychology of Addictive Behaviors, 26,* 906–916. http://dx.doi.org/10.1037/a0028158

Huntley, A. L., Araya, R., & Salisbury, C. (2012). Group psychological therapies for depression in the community: Systematic review and meta-analysis. *The British Journal of Psychiatry, 200,* 184–190. http://dx.doi.org/10.1192/bjp.bp.111.092049

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141,* 747–768. http://dx.doi.org/10.1037/bul0000015

Joutsenniemi, K., Laaksonen, M. A., Knekt, P., Haaramo, P., & Lindfors, O. (2012). Prediction of the outcome of short- and long-term psychotherapy based on socio-demographic factors. *Journal of Affective Disorders, 141,* 331–342. http://dx.doi.org/10.1016/j.jad.2012.03.027

*Kavanagh, D. J., & Wilson, P. H. (1989). Prediction of outcome with group cognitive therapy for depression. *Behaviour Research and Therapy, 27,* 333–343. http://dx.doi.org/10.1016/0005-7967(89)90003-X

Kendall, P. C., Chu, B., Gifford, A., Hayes, C., & Nauta, M. (1998). Breathing life into a manual: Flexibility and creativity with manual-based treatments. *Cognitive and Behavioral Practice, 5,* 177–198. http://dx.doi.org/10.1016/S1077-7229(98)80004-7

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11,* 289–299. http://dx.doi.org/10.1007/BF01186280

Khoshbooii, R. (2012). Comparison group and individual cognitive behavioral therapy in treatment of depression among Iranian women around menopause. *International Journal of Psychological Studies, 4,* 175–181. http://dx.doi.org/10.5539/ijps.v4n1p174

Kishi, T., Kafantaris, V., Sunday, S., Sheridan, E. M., & Correll, C. U. (2012). Are antipsychotics effective for the treatment of anorexia nervosa? Results from a systematic review and meta-analysis. *The Journal of Clinical Psychiatry, 73,* e757–e766. http://dx.doi.org/10.4088/JCP.12r07691

Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., . . . Leon, A. C. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry, 51,* 319–324. http://dx.doi.org/10.1016/j.comppsych.2009.07.001

*Kornblith, S., Rehm, L., O'Hara, M., & Lamparski, D. (1983). The contribution of self-reinforcement training and behavioral assignments to the efficacy of self-control therapy for depression. *Cognitive Therapy and Research, 7,* 499–527. http://dx.doi.org/10.1007/BF01172888

Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. (2002). *Meta-analysis: A guide to calibrating and combining statistical evidence*. Hoboken, NJ: Wiley.

Lambert, M. J. (2001). The status of empirically supported therapies: Comment on Westen and Morrison's (2001). multidimensional meta-analysis. *Journal of Consulting and Clinical Psychology, 69,* 910–913. http://dx.doi.org/10.1037/0022-006X.69.6.910

Ljótsson, B., Hedman, E., Mattsson, S., & Andersson, E. (2017). The effects of cognitive-behavioral therapy for depression are not falling: A re-analysis of Johnsen and Friborg (2015). *Psychological Bulletin, 143,* 321–325. http://dx.doi.org/10.1037/bul0000055

*Manicavasgar, V., Parker, G., & Perich, T. (2011). Mindfulness-based cognitive therapy vs. cognitive behaviour therapy as a treatment for non-melancholic depression. *Journal of Affective Disorders, 130,* 138–144. http://dx.doi.org/10.1016/j.jad.2010.09.027

McDermut, W., Miller, I. W., & Brown, R. A. (2001). The efficacy of group psychotherapy for depression: A meta-analysis and review of the empirical research. *Clinical Psychology: Science and Practice, 8,* 98–116. http://dx.doi.org/10.1093/clipsy.8.1.98

McGuire, J. F., Piacentini, J., Brennan, E. A., Lewin, A. B., Murphy, T. K., Small, B. J., & Storch, E. A. (2014). A meta-analysis of behavior therapy for Tourette Syndrome. *Journal of Psychiatric Research, 50,* 106–112. http://dx.doi.org/10.1016/j.jpsychires.2013.12.009

Mental Health Foundation. (2010). *Economic burden of mental illness cannot be tackled without research investment*. Retrieved from http://www.mentalhealth.org.uk/content/assets/PDF/campaigns/MHF-Business-case-for-MH-research-Nov2010.pdf

*Milgrom, J., Gemmill, A. W., Ericksen, J., Burrows, G., Buist, A., & Reece, J. (2015). Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: A randomised controlled trial. *Australian and New Zealand Journal of Psychiatry, 49,* 236–245. http://dx.doi.org/10.1177/0004867414565474

Miller, S. J., & Binder, J. L. (2002). The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 39,* 184–198. http://dx.doi.org/10.1037/0033-3204.39.2.184

*Mokrue, K., & Acri, M. (2013). Feasibility and effectiveness of a brief cognitive behavioral skills group on an ethnically diverse campus. *Journal of College Student Psychotherapy, 27,* 254–269. http://dx.doi.org/10.1080/87568225.2013.766114

Murphy, G. E., Carney, R. M., Knesevich, M. A., Wetzel, R. D., & Whitworth, P. (1995). Cognitive behavior therapy, relaxation training, and tricyclic antidepressant medication in the treatment of depression. *Psychological Reports, 77,* 403–420. http://dx.doi.org/10.2466/pr0.1995.77.2.403

National Institute of Mental Health. (2015). *Depression*. Retrieved from http://www.nimh.nih.gov/health/topics/depression/index.shtml#part_145396

*Neimeyer, R. A., & Feixas, G. (1990). The role of homework and skill acquisition in the outcome of group cognitive therapy for depression. *Behavior Therapy, 21,* 281–292. http://dx.doi.org/10.1016/S0005-7894(05)80331-4

Oei, T. P., & Dingle, G. (2008). The effectiveness of group cognitive behaviour therapy for unipolar depressive disorders. *Journal of Affective Disorders, 107,* 5–21. http://dx.doi.org/10.1016/j.jad.2007.07.018

*Oei, T. P., & Yeoh, A. E. (1999). Pre-existing antidepressant medication and the outcome of group cognitive-behavioural therapy. *The Australian and New Zealand Journal of Psychiatry, 33,* 70–76. http://dx.doi.org/10.1046/j.1440-1614.1999.00520.x

Okumura, Y., & Ichikura, K. (2014). Efficacy and acceptability of group cognitive behavioral therapy for depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 164,* 155–164. http://dx.doi.org/10.1016/j.jad.2014.04.023

Owen, J., & Hilsenroth, M. J. (2014). Treatment adherence: The importance of therapist flexibility in relation to therapy outcomes. *Journal of Counseling Psychology, 61,* 280–288. http://dx.doi.org/10.1037/a0035753

Patten, S. B. (2006). A major depression prognosis calculator based on episode duration. *Clinical Practice and Epidemiology in Mental Health, 2,* 13. http://dx.doi.org/10.1186/1745-0179-2-13

*Propst, L. R. (1980). The comparative efficacy of religious and nonreligious imagery for the treatment of mild depression in religious individuals. *Cognitive Therapy and Research, 4,* 167–178. http://dx.doi.org/10.1007/BF01173648

*Rehm, L. P., Kaslow, N. J., & Rabin, A. S. (1987). Cognitive and behavioral targets in a self-control therapy program for depression. *Journal of Consulting and Clinical Psychology, 55,* 60–67. http://dx.doi.org/10.1037/0022-006X.55.1.60

*Rohan, K. J., Roecklein, K. A., Tierney Lindsey, K., Johnson, L. G., Lippy, R. D., Lacy, T. J., & Barton, F. B. (2007). A randomized controlled trial of cognitive–behavioral therapy, light therapy, and their combination for seasonal affective disorder. *Journal of Consulting and Clinical Psychology, 75,* 489–500. http://dx.doi.org/10.1037/0022-006X.75.3.489

*Rokke, P. D., Tomhave, J. A., & Jocic, Z. (2000). Self-management therapy and educational group therapy for depressed elders. *Cognitive Therapy and Research, 24,* 99–119. http://dx.doi.org/10.1023/A:1005407125992

Rosenthal, R. (1993). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

*Rush, A. J., & Watkins, J. T. (1982). Group versus individual cognitive therapy: A Pilot Study. *Cognitive Therapy and Research, 5,* 95–103. http://dx.doi.org/10.1007/BF01172329

*Shaffer, C. S., Shapiro, J., Sank, L. I., & Coghlan, D. J. (1981). Positive changes in depression, anxiety, and assertion following individual and group cognitive behavior therapy intervention. *Cognitive Therapy and Research, 5,* 149–157. http://dx.doi.org/10.1007/BF01172523

Shaw, B. F. (1977). Comparison of cognitive therapy and behavior therapy in the treatment of depression. *Journal of Consulting and Clinical Psychology, 45,* 543–551. http://dx.doi.org/10.1037/0022-006X.45.4.543

Sobocki, P., Jönsson, B., Angst, J., & Rehnberg, C. (2006). Cost of depression in Europe. *Journal of Mental Health Policy and Economics, 9,* 87–98.

Solomon, D. A., Keller, M. B., Leon, A. C., Mueller, T. I., Lavori, P. W., Shea, M. T., . . . Endicott, J. (2000). Multiple recurrences of major depressive disorder. *The American Journal of Psychiatry, 157,* 229–233. http://dx.doi.org/10.1176/appi.ajp.157.2.229

*Steuer, J. L., Mintz, J., Hammen, C. L., Hill, M. A., Jarvik, L. F., McCarley, T., . . . Rosen, R. (1984). Cognitive–behavioral and psychodynamic group psychotherapy in treatment of geriatric depression. *Journal of Consulting and Clinical Psychology, 52,* 180–189. http://dx.doi.org/10.1037/0022-006X.52.2.180

*Stravynski, A., Verreault, R., Gaudette, G., Langlois, R., Gagnier, S., & Larose, M. (1994). The treatment of depression with group behavioural–cognitive therapy and imipramine. *Canadian Journal of Psychiatry, 39,* 387–390. http://dx.doi.org/10.1177/070674379403900701

*Teismann, T., von Brachel, R., Hanning, S., Grillenberger, M., Hebermehl, L., Hornstein, I., & Willutzki, U. (2014). A randomized controlled trial on the effectiveness of a rumination-focused group treatment for residual depression. *Psychotherapy Research, 24,* 80–90. http://dx.doi.org/10.1080/10503307.2013.821636

*Teri, L., & Lewinsohn, P. M. (1986). Individual and group treatment of unipolar depression: Comparison of treatment outcome and identification of predictors of successful treatment outcome. *Behavior Therapy, 17,* 215–228. http://dx.doi.org/10.1016/S0005-7894(86)80052-1

Thoma, N. C., McKay, D., Gerber, A. J., Milrod, B. L., Edwards, A. R., & Kocsis, J. H. (2012). A quality-based review of randomized controlled trials of cognitive–behavioral therapy for depression: An assessment and metaregression. *The American Journal of Psychiatry, 169,* 22–30. http://dx.doi.org/10.1176/appi.ajp.2011.11030433

Tucker, M., & Oei, T. P. S. (2007). Is group more effective than individual cognitive behavior therapy? The evidence is not solid yet. *Behavioural and Cognitive Psychotherapy, 35,* 77–91. http://dx.doi.org/10.1017/S1352465806003134

*Van Dam, E. A., Hosman, C. M., Hoogduin, C. A., & Schaap, P. D. (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,* 381–396. http://dx.doi.org/10.1016/S0005-7894(03)80007-2

Vos, T., Corry, J., Haby, M. M., Carter, R., & Andrews, G. (2005). Cost-effectiveness of CBT and drug interventions for major depression. *Australian and New Zealand Journal of Psychiatry, 39,* 683–692.

Waltman, S. H., Creed, T. A., & Beck, A. T. (2016). Are the effects of cognitive behavior therapy for depression falling? Review and critique of the evidence. *Clinical Psychology: Science and Practice, 23,* 113–122. http://dx.doi.org/10.1111/cpsp.12152

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate*. New York, NY: Routledge.

Wampold, B. E., Minami, T., Baskin, T. W., & Callen Tierney, S. (2002). A meta-(re)analysis of the effects of cognitive therapy versus "other therapies" for depression. *Journal of Affective Disorders, 68,* 159–165. http://dx.doi.org/10.1016/S0165-0327(00)00287-1

Watts, B. V., Schnurr, P. P., Mayo, L., Young-Xu, Y., Weeks, W. B., & Friedman, M. J. (2013). Meta-analysis of the efficacy of treatments for posttraumatic stress disorder. *The Journal of Clinical Psychiatry, 74,* e541–e550. http://dx.doi.org/10.4088/JCP.12r08225

Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. *Advances in Psychiatric Treatment, 16,* 219–227. http://dx.doi.org/10.1192/apt.bp.108.005744

*Wierzbicki, M., & Bartlett, T. S. (1987). The efficacy of group and individual cognitive therapy for mild depression. *Cognitive Therapy and Research, 11,* 337–342. http://dx.doi.org/10.1007/BF01186284

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24,* 190–195. http://dx.doi.org/10.1037/0735-7028.24.2.190

*Wilson, G. L. (1990). Psychotherapy with depressed incarcerated felons: A comparative evaluation of treatments. *Psychological Reports, 67,* 1027–1041. http://dx.doi.org/10.2466/PR0.67.7.1027-1041

*Wollersheim, J. P., & Wilson, G. L. (1991). Group treatment of unipolar depression: A comparison of coping, supportive, bibliotherapy, and delayed treatment groups. *Professional Psychology, Research and Practice, 22,* 496–502. http://dx.doi.org/10.1037/0735-7028.22.6.496

*Wong, D. F. (2008). Cognitive behavioral treatment groups for people with chronic depression in Hong Kong: A randomized wait-list control design. *Depression and Anxiety, 25,* 142–148. http://dx.doi.org/10.1002/da.20286

Wright, J. H., Wright, A. S., Albano, A. M., Basco, M. R., Goldsmith, L. J., Raffield, T., & Otto, M. W. (2005). Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. *The American Journal of Psychiatry, 162,* 1158–1164. http://dx.doi.org/10.1176/appi.ajp.162.6.1158

*Zamirinejad, S., Hojjat, S. K., Golzari, M., Borjali, A., & Akaberi, A. (2014). Effectiveness of resilience training versus cognitive therapy on reduction of depression in female Iranian college students. *Issues in Mental Health Nursing, 35,* 480–488. http://dx.doi.org/10.3109/01612840.2013.879628

*Zettle, R. D., Haflich, J. L., & Reynolds, R. A. (1992). Responsivity to cognitive therapy as a function of treatment format and client personality dimensions. *Journal of Clinical Psychology, 48,* 787–797. http://dx.doi.org/10.1002/1097-4679(199211)48:6<787::AID-JCLP2270480613>3.0.CO;2-D

*Zettle, R. D., & Rains, J. C. (1989). Group cognitive and contextual therapies in treatment of depression. *Journal of Clinical Psychology, 45,* 436–445. http://dx.doi.org/10.1002/1097-4679(198905)45:3<436::AID-JCLP2270450314>3.0.CO;2-L

Zoogman, S., Goldberg, S. B., Hoyt, W. T., & Miller, L. (2015). Mindfulness interventions with youth: A meta-analysis. *Mindfulness, 6,* 290–302. http://dx.doi.org/10.1007/s12671-013-0260-4

Received January 10, 2017
Revision received September 28, 2017
Accepted October 13, 2017 ■

Personality and Social Psyhology

Time trends in the effects of mindfulness-based cognitive therapy for depression: A meta-analysis

JENS C. THIMM[1] and TOM J. JOHNSEN[2]

[1]*Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway*
[2]*Private practice, Tromsø, Norway*

Thimm, J. C. & Johnsen, T. J. (2020). Time trends in the effects of mindfulness-based cognitive therapy for depression: A meta-analysis. *Scandinavian Journal of Psychology*.

Recent studies suggest that the effects of cognitive therapies for depression show systematic changes over time. A meta-analysis was conducted to explore the temporal development of the effect of mindfulness-based cognitive therapy (MBCT) for current depression in studies that used the Beck Depression Inventory (BDI) or the Hamilton Depression Rating Scale (HDRS) as outcome measures. A systematic search of research databases yielded 20 studies that were included in the analyses. The results showed that MBCT is effective in reducing depressive symptoms. The effect sizes of studies using the BDI or the HDRS as an outcome measure were not moderated by the time of publication. Funnel plots and the trim and fill method suggested that publication bias was low. However, the number of available studies was small, and the time period investigated relatively short. The results should therefore be considered preliminary.

Key words: Mindfulness-based cognitive therapy, depression, effectiveness, time trends, meta-analysis.

Jens C. Thimm, UiT The Arctic University of Norway, Department of Psychology, 9037 Tromsø, Norway. Tel: (+47) 77644348; e-mail: jcthimm@gmail.com

INTRODUCTION

With an estimated global prevalence of 4.4% (WHO, 2017), depression is one of the most common and frequently occurring mental disorders, with a high rate of relapse and recurrence (Steinert, Hofmann, Kruse, & Leichsenring, 2014). As such, depression represents a significant burden on the individual and society (Whiteford, Degenhardt, Rehm, Baxter, Ferrari, Erskine, & Vos, 2013).

A variety of psychological therapies, including cognitive therapy (CT), behavioral activation therapy, interpersonal therapy and short-term psychodynamic therapy, have been shown to be effective in the treatment of depression (Cuijpers, 2017). Recently, mindfulness-based cognitive therapy (MBCT; Crane, 2009; Segal, Williams, & Teasdale, 2002; Segal, Williams, Teasdale, & Kabat-Zinn, 2013) was developed as a modification of CT to specifically prevent the relapse and recurrence of depressive episodes in individuals who had recovered from depression (Lau, 2016). MBCT is a manual-based treatment that combines exercises in mindfulness training with cognitive techniques. The integration of mindfulness practice with cognitive interventions distinguishes MBCT from other mindfulness-based interventions (MBIs), such as mindfulness-based stress reduction (Kabat-Zinn, 1990). The overall goal of MBCT is to increase metacognitive awareness (Lau, Segal, & Williams, 2004) and, thereby, reduce cognitive and emotional reactivity (Gu, Strauss, Bond, & Cavanagh, 2015).

Studies have shown that MBCT is effective in reducing the relapse and recurrence of depression (Kuyken, Warren, Taylor, Whalley, Crane, Bondolfi, & Schweizer, 2016; Piet & Hougaard, 2011). MBCT seems to be equally effective in reducing risk of relapse as CBT (Farb, Anderson, Ravindran, Hawley, Irving, Mancuso, & Segal, 2018) and more effective than antidepressant medications in this regard (Kuyken *et al.*, 2016). Although, in an early paper, the developers of MBCT cautioned against using MBCT to treat treating patients with acute depression (Teasdale, Segal, Williams, Ridgeway, Soulsby, & Lau, 2000), MBCT has subsequently been extended to this group. The treatment of current unipolar depression with MBCT follows the original manual by Segal *et al.* (2013) and is delivered in a group format with up to 12 participants and one or two instructors. After an individual pretreatment interview in which the participant's history of depression is discussed and information about MBCT is provided, the treatment consists of eight weekly two-hour sessions (Baer & Walsh, 2016).

Several meta-analyses have shown that mindfulness-based interventions (MBIs) in general (e.g., Goldberg, Tucker, Greene, Davidson, Wampold, Kearney, & Simpson, 2018; Goyal, Singh, Sibinga, &, Singh, & Sibinga&&, 2014; Hedman-Lagerlöf, Hedman-Lagerlöf, & Öst, 2018; Hofmann, Sawyer, Witt, & Oh, 2010; Khoury, Lecomte, Fortin, Masse, Therien, Bouchard, & Hofmann, 2013; McCarney, Schulz, & Grey, 2012; Strauss, Cavanagh, Oliver, & Pettman, 2014; Wang, Li, Zheng, Xu, Ng, Ungvari, & Xiang, 2018), and MBCT in particular (Galante, Iribarren, & Pearce, 2013; Hofmann *et al.*, 2010; Klainin-Yobas, Cho, & Creedy, 2012; Lenz, Hall, & Bailey Smith, 2016), are effective in reducing depressive symptoms. For example, a recent meta-analysis of randomized controlled trials (RCTs) observed effect sizes (ESs) of $d = 0.59$ for MBIs vs. no treatment and $d = 0.38$ for MBIs vs. active control conditions (Goldberg *et al.*, 2018). For MBCT specifically, similar or higher ESs for the reduction of depressive symptom severity have been reported. For example, Hofmann *et al.* (2010) observed an average ES of 0.85 (Hedges's g) in nine pre-post studies. Lenz *et al.* (2016) reported mean ES of $g = 0.76$ and 0.54 for MBCT vs. waitlist or no treatment and for MBCT vs. alternative treatments, respectively,

in RCTs. Recently. Goldberg, Tucker, Greene, Davidson, Kearney, and Simpson (2019) found that MBCT was superior to non-specific control conditions (d = 0.71) at posttest but not more effective than other active treatments (d = 0.00).

In previous meta-analyses of MBCT for acute depression, the temporal development of ESs in treatment studies has received little attention. This may not be surprising, as MBCT is a relatively new development. However, the investigation of the relationship between time of study and ES is important, as it informs about time trends and developments that can be positive or negative and call for action. For example, a decline in ESs for individual CBT for depression has been observed in published studies over time using a version of the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and/or the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960) as outcome measures (Cristea, Stefan, Karyotaki, David, Hollon, & Cuijpers, 2017; Johnsen & Friborg, 2015). Several possible explanations for these findings have been discussed, including more heterogeneous and complex samples in more recent trials, therapist training, and lack of adherence to the treatment manual (Dobson, 2016; Johnsen & Friborg, 2015; Waltman, Creed, & Beck, 2016). In contrast, an increase in ESs for group CBT for depression was observed when the BDI was used as outcome measure but was not observed when the HDRS was used (Johnsen & Thimm, 2018). As MBCT is an anti-depressive treatment that shares key concepts and features with traditional CBT and usually is delivered in a group format, the investigation of the time-trends connected to this treatment form can give new and relevant insights. For example, combined with the previous research on the temporal development of ESs in studies of individual CBT and group CBT, analyses such as the present one can provide indications as to whether any time-trends can be connected to the treatment format (group vs. individual therapy), or, alternatively, can be related to the focus of the interventions rather than the treatment format.

Thus, the aim of the present exploratory study is to examine the effect of MBCT on the treatment of current depression and the development of ESs over time. Since the BDI and the HDRS are the most widely used instruments for evaluating the effectiveness of cognitive therapies for depression (Johnsen & Friborg, 2015), the analysis will focus on studies that used the BDI and/or the HDRS as outcome measures to allow for time trends of depression treatment to be compared with previous studies. In the analysis, studies comparing MBCT to control condition with and without active treatments and studies examining pre-post differences were included. Recently, it has been suggested that ESs based on within-group scores should be avoided in meta-analysis if possible (Cuijpers, Weitz, Cristea, & Twisk, 2017). While there is no doubt that between-group RCTs is the gold standard when it comes to conduct meta-analysis, there are also several salient reasons why the inclusion of within-group based ESs could be helpful - albeit used with caution when it comes to interpretation of the results. First, the scope of papers could be greatly expanded, which would increase the statistical power. Second, we believe that meta-analyses on time-trends are not as vulnerable to some of the pitfalls as meta-analyses measuring standard treatment effects. For example, one common objection to within-group pre-post standardized mean differences is that they are influenced by natural processes and characteristics of patients and settings, which cannot be discerned from the effects of the intervention. However, when it comes to research of temporal developments in treatment efficacy related to any particular treatment form, variations in the characteristics of patients and settings (as well as general environment and society) could very well be highly relevant moderators to consider when it comes to interpreting the reasons behind any temporal development of treatment effects. Identification of any characteristics or processes that change systematically with the passing of time, influencing treatment effects, are of major importance. Finally, the most accurate indicator of reliability for any pre-post ES-calculation is heterogeneity. If this index is at a satisfactorily level, within-group ES's could be an informative calculation of ESs. With these considerations in mind, we have for the present study chosen to perform a primary analysis utilizing between-group RCT-based ESs, and a secondary analysis utilizing within-group ESs. We expect that the outcomes of the two calculations regarding the temporal development of ESs would be similar and thus validate each other's results.

METHODS

To identify relevant studies, a systematic search was conducted in research databases MEDLINE, PsychINFO and EMBASE on January 20, 2018. The broad search query "mindfulness AND depress*" was used to minimize the risk of missing relevant studies. In addition, previous systematic reviews and published meta-analyses of MBIs for mental disorders were manually searched. After removal of duplicates, in the first stage of the study selection, the titles, abstracts, types of references, and language of publication were screened by the first author. In the second round, both authors assessed the full text of studies for eligibility. The following inclusion criteria were applied: 1) MBCT was given in a group format aimed at reducing depression; 2) participants were adults (≥18 years of age) diagnosed with depression or showing elevated scores on the BDI (> 13) or the HDRS (> 8), as a group; 3) a version of the BDI or the HDRS was used as an outcome measure; and 4) publication was in English and was in a peer-reviewed journal. Studies were excluded when 1) MBIs other than MBCT were examined, 2) no treatment effects for MBCT were investigated or reported, 3) depression was not the principal problem of the participants; 4) partial or complete sample overlap with a study already included in the meta-analysis was observed, 5) information necessary to calculate ES (i.e., means and standard deviations) was lacking, or 6) only dichotomous outcomes (e.g., relapse) were reported.

For each study included in the meta-analysis, the following information was extracted: 1) year of publication; 2) sample size of the MBCT group and the control group; 3) mean age and percentage of females in the MBCT group; 4) number of sessions; 5) modification of the treatment manual by Segal *et al.* (2002) or Segal *et al.* (2013); 6) use of the BDI or BDI-II as outcome measure; 7) no treatment vs. active treatment comparison groups; 8) randomization of participants; and 9) reporting results of intent-to-treat (ITT) analyses.

For the meta-analytic calculations, means and standard deviations of the BDI and/or the HDRS at pre-treatment and post-treatment were extracted for the treatment group and, if present, for the control group(s).

To assess the methodological quality of the studies included in the meta-analysis, the Jadad scale (Jadad, Moore, Carroll, Jenkinson, Reynolds, Gavaghan, & McQuay, 1996) was used. Both authors assessed the studies independently. Rater agreement was calculated using double entry intraclass correlation (McCrae, 2008). The coefficient for study quality was .91. Discrepant ratings were clarified and resolved through discussion.

To obtain the ES for each study, the standardized mean difference (SMD) between the intervention group and control group, and/or the pretest and the posttest was calculated correcting for bias (Hedges' g). Following the recommendations by Rosenthal (1993), a conservative pre-

post correlation of .7 was set. The mean ES across studies was calculated using a random effects model. The analyses were conducted separately for controlled studies (between-group) with and without active treatment comparisons and pre-post differences (within-group) and for the BDI/BDI-II and HDRS as outcome measures. When data for ITT samples were available, these were preferred over data from completer samples.

To examine publication year as moderator for the pooled ES, meta-regression analysis was used.

Heterogeneity among studies was assessed using Q tests and the I^2 statistic (Higgins, Thompson, Deeks, & Altman, 2003), which is a measure of the proportion of the total variance across studies that is due to heterogeneity. Higgins *et al.* (2003) suggest that I^2 values of 25% indicate low heterogeneity, 50% indicate moderate heterogeneity, and 75% indicate high heterogeneity between studies.

To assess publication bias, funnel plots were obtained, and Duval and Tweedie's (2000) trim and fill method was used to estimate the number of missing studies and the ESs after imputation of the missing studies.

All analyses were conducted in Comprehensive Meta-Analysis version 3 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2017).

RESULTS

Study selection

After duplicates were removed, the search resulted in 4,010 unique studies. In the screening process, 3869 studies were excluded. One hundred forty-one full-text articles were retrieved, 121 of which were excluded based on the eligibility criteria. Thus, 20 eligible studies were retained for the meta-analysis (see Figure 1 for the flowchart of the selection process).

Study characteristics

The study characteristics are displayed in Table 1. The included studies were published between 2007 (Kenny & Williams, 2007; Kingston, Dooley, Bates, Lawlor, & Malone, 2007 and 2017 (Greenberg, Shapero, Mischoulon, & Lazar, 2017). The average number of participants in the MBCT condition was 41.2 (SD = 46.2, range 6 to 212). The total number of participants was 824. In ten studies, the original MBCT manual was modified to adapt the treatment to the target group. Eighteen studies were RCTs (14) or included a control condition without randomization (4). Seven of these 18 studies included an active control condition (antidepressant medication in three studies, and psychological treatment in four studies). Two studies had two control groups (Hosseinian, Shahtaheri, Ebrahimi, Mahdavi, & Sepahvandi, 2016; Michalak, Schultze, Heidenreich, & Schramm, 2015). Two studies used a pre-post design. Treatment outcome was measured with a version of the BDI in 15 between-group studies and two within-group studies and with the HDRS in nine between-group studies. As to study quality, the average Jadad score was 1.90 (SD = 1.17, range 0–3).

Effects of MBCT on current depression and analysis of time trends

The average weighted ES for between-group studies using a no-treatment control group and the BDI as an outcome measure (n = 11) was g = 0.92 (95% CI [0.70, 1.14]; $Q(10)$ = 17.45, p = 0.065, I^2 = 42.7). The trim and fill method suggested that two studies were missing, and the imputed point estimate was g = 0.86 (95% CI [0.63, 1.08]). When an active treatment comparison group was included (n = 5), the ES for the BDI was g = 0.45 (95% CI [0.09, 0.80];$Q(4)$ = 11.16, p = 0.025, I^2 = 64.2). The trim and fill method suggested that no studies were missing. For the between-group studies using the HDRS and a no-treatment control group (n = 7), the ES was g = 0.80 (95% CI [0.61, 0.99]; $Q(6)$ = 7.15, p = 0.308, I^2 = 16.04). The trim and fill method suggested that two studies were missing, and the imputed point estimate was g = 0.72 (95% CI [0.51, 0.92]). For studies using an active treatment control group and the HDRS (n = 4), the mean weighted ES was g = 0.37 (95% CI [0.21, 0.54]; $Q(3)$ = 2.62, p = 0.454, I^2 = 0). The trim and fill method suggested that one study was missing, and the imputed point estimate was g = 0.34 (95% CI [0.12, 0.56]).

With respect to pre-post differences on the BDI, studies using a within-group design were pooled with between-group studies (n = 17), resulting in an ES of g = 0.90 (95% CI [0.70, 1.09]; $Q(16)$ = 120.36, p < .001, I^2 = 86.7). The trim and fill method suggested that no studies were missing. The ES of the Abolghasemi, Gholami, Narimani, and Gamji (2015) study was considerably larger than the ESs of the other studies (g = 5.18, 95% CI [3.70, 6.70]). When the Abolghasemi *et al.* (2015) study was removed from the analysis, the average weighted ES was g = 0.82 (95% CI [0.66, 0.99], $Q(15)$ = 84.27, p < .001, I^2 = 82.2). There were no within-group studies that used the HDRS as an outcome measure. ESs for the individual studies are presented in Figures 2 through 6.

Visual inspections of the funnel plots revealed largely symmetrical distributions. The funnel plots of the observed and imputed studies are provided in the supplemental material Figures S1 to S5.

Analysis of time trends showed no significant relationships between year of publication and ES for between-group studies with the BDI as an outcome measure and with no-treatment comparisons (b = −0.03, 95% CI [−0.11, 0.05], p = 0.440), active treatment comparisons (b = −0.02, 95% CI [−0.28, 0.24], p = 0.863), and in pre-post designs (b = −0.01, 95% CI [−0.07, 0.05], p = 0.657, and b = −0.03, 95% CI [−0.08, 0.01], p = 0.155 when the Abolghasemi *et al.* (2015) study was excluded). For studies using the HDRS, the associations between year of publication and ES were not significant for between-group comparisons with no treatment (b = −0.04, 95% CI [−0.13, 0.05], p = 0.348) and active treatment comparisons (b = −0.10, 95% CI [−0.95, 0.74], p = 0.810).

Analysis of other moderators

In addition to year of publication, sample size, average age, gender distribution, and baseline level of depression in the MBCT group, as well as study quality (Jadad score), were examined as moderators. None of these variables moderated the ESs of between-group studies with no-treatment controls using the BDI (sample size: b = 0.00, 95% CI [−0.01, 0.00], p = 0.389; age: b = −0.01, 95% CI [−0.07, 0.05], p = 0.758; gender: b = −0.01, 95% CI [−0.03, 0.02], p = 0.677; baseline depression: b = 0.02, 95% CI [−0.02, 0.06], p = 0.262; study quality: b = −0.07, 95% CI [−0.28, 0.15], p = 0.546) or the HDRS (sample size: b = 0.00, 95% CI [−0.01, 0.01], p = 0.624;

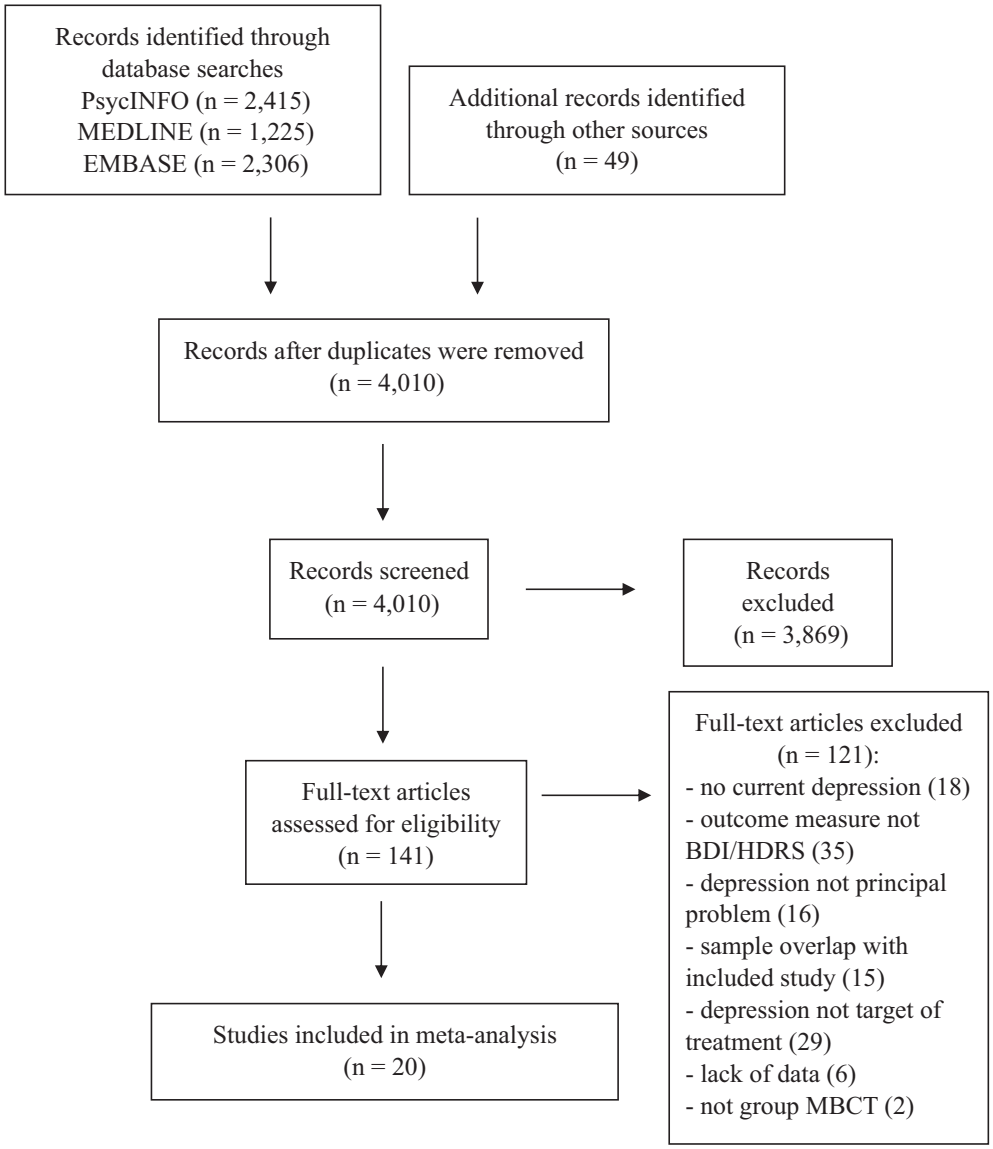


Fig. 1. Flowchart of the search and selection procedure.

age ($b = -0.04$, 95% CI [−0.12, 0.04], $p = 0.313$; gender: $b = 0.02$, 95% CI [0.00, 0.03], $p = 0.087$; baseline depression: $b = -0.02$, 95% CI [−0.05, 0.01], $p = 0.271$; study quality: $b = 0.07$, 95% CI [−0.48, 0.33], $p = 0.727$). No moderation of these variables was also found for studies with active treatment comparisons using the BDI (sample size: $b = 0.00$, 95% CI [−0.01, 0.01], $p = 0.835$; age: $b = -0.01$, 95% CI [−0.04, 0.03], $p = 0.752$; gender: $b = 0.02$, 95% CI [−0.01, 0.05], $p = 0.252$; baseline depression: $b = 0.00$, 95% CI [−0.06, 0.06], $p = 0.908$; study quality: $b = -0.30$, 95% CI [−0.78, 0.17], $p = .208$) and the HDRS (sample size: $b = 0.00$, 95% CI [0.00, 0.00], $p = 0.183$; age: not enough studies; gender: $b = -0.01$, 95% CI [−0.03, 0.01], $p = 0.462$; baseline depression: $b = 0.01$, 95% CI [−0.01, 0.04], $p = 0.177$; study quality: $b = -0.03$, 95% CI [−0.31, 0.25], $p = 0.854$).

For within-group comparisons using the BDI, sample size ($b = 0.00$, 95% CI [−0.01, 0.00], $p = 0.013$), age ($b = -0.08$, 95% CI [−0.12, −0.03], $p < 0.001$), and baseline depression ($b = 0.04$, 95% CI [0.01, 0.06], $p = 0.002$) were significant moderators, but sex ($b = -0.01$, 95% CI [−0.03, 0.01], $p = 0.353$) and study quality ($b = -0.14$, 95% CI [−0.29, 0.01], $p = 0.078$) were not. When the Abolghasemi *et al.* (2015) study was excluded from the analyses, sample size ($b = 0.00$, 95% CI [−0.01, 0.00], $p = 0.002$) and baseline depression ($b = 0.03$, 95% CI [0.00, 0.05], $p = 0.020$) were significant moderators but not age ($b = -0.03$, 95% CI [−0.07, 0.01], $p = 0.110$), sex ($b = 0.00$, 95% CI [−0.02, 0.02], $p = 0.976$), and study quality ($b = -0.10$, 95% CI [−0.24, 0.05], $p = 0.202$). Thus, smaller sample size and higher baseline depression was associated with higher ESs across statistical conditions.

DISCUSSION

The present meta-analysis explored the development of ESs for MBCT over time in the treatment of current depression in studies that used the BDI or the HDRS as outcome measures. Previous findings indicated significant changes in the ESs of individual and group CBT for depression over time (Cristea *et al.*, 2017; Johnsen

Table 1. *Overview of studies included in the meta-analysis*

Study	Sample	MBCT protocol modified	*N* MBCT condition	Comparison condition (*N*)	Random-ization	No. of treatment sessions	Depression measures	Jadad score
Abolghasemi *et al.* (2015)	Depression	yes	15	CT (15)	yes	12	BDI-II	1
Barnhofer, Crane, Hargus, Amarasinghe, Winder, and Williams (2009)[1]	Chronic depression	yes	16	TAU (15)	yes	8	BDI-II	3
Chiesa, Castagner, Andrisano, Serretti, Mandelli, Porcelli, and Giommi (2015)[1]	Depression	yes	23	Psycho-education (20)	yes	8	HDRS BDI-II	3
Crane, Barnhofer, Duggan, Hepburn, Fennell, and Williams (2008)	Past depression and active suicidal ideation	yes	19	Waitlist (23)	yes	8	BDI-II	3
De Raedt, Baert, Demeyer, Goeleven, Raes, Visser, and Speckens (2012)	Former depression	no	44	No intervention (26)	no	8	BDI-II	0
Eisendrath et al. (2008)[1]	Treatment-resistant depression	yes	55	none	n/a	8	BDI-II	0
Eisendraht et al. (2015)	Depression	yes	19	Antidepressant management (17)	no	8	HDRS	1
Geschwind, Peeters, Drukker, Os, and Wichers (2011)[1]	Past depression and residual depressive symptoms	no	63	Waitlist (66)	yes	8	HDRS	3
Godfrin and van Heeringen (2010)[1]	Recurrent depression	no	52	Waitlist (54)	yes	8	BDI-II HDRS	3
Greenberg *et al.* (2017)	Depression	no	12 (BDI) 16	(HDRS)	Waitlist (BDI: 13; HDRS: 9)	yes	8	BDI-II
HDRS								
	2							
Hamidian, Omidi, Mousavinasab, and Naziri (2013)	Dysthymia	no	22	Medication (22)	yes	8	BDI-II	2
Hosseinian *et al.* (2016)	Depression	no	12	- Metacognitive therapy (12) - nonspecified control (12)	yes	8	HDRS	1
Kenny and Williams (2007)	Treatment-resistant depression	no	46	none	n/a	8	BDI	0
Kingston *et al.* (2007)	Recurrent depression	no	6	Waitlist (11)	no	8	BDI-II	1
Kuyken, Hayes, Barrett, Byng, Dalgleish, Kessler, and Byford (2015)[1]	Recurrent depression	yes	212	Antidepressive medication (212)	yes	8	BDI-II HDRS	3
Manicavasgar, Parker, and Perich (2011)	Depression	yes	19	CBT (26)	yes	8	BDI-II	2
Mann, Kuyken, O'Mahen, Ukoumunne, Evans, and Ford (2016)[1]	Previous depression	yes	19	TAU (19)	yes	8	BDI-II	3
Michalak *et al.* (2015)[1]	Chronic depression	yes	36	- CBASP (35) - TAU (35)	yes	8	BDI-II HDRS	3
van Alderen et al. (2012)[1]	Recurrent depression	no	102	TAU (103)	yes	8	BDI HDRS	3
Verhoeven, Vrijsen, Oostrom, Speckens, and Rinck (2014)	Remitted depressed patients	no	28	Waitlist, patients treated for depression (26)	no	8	BDI-II	1

[1]Results from intent-to-treat analyses reported. n/a = not applicable.

& Friborg, 2015; Johnsen & Thimm, 2018). The main goal of the present study was, therefore, to examine if reported time trends of ESs for MBCT for depression could be observed in the different studies.

The results showed that the ESs of studies using between- and within-group designs and the BDI or the HDRS as an outcome measure were not moderated by the time of publication. In previous studies of time trends of ESs, diverging results for the BDI and HDRS have been observed (e.g., Johnsen & Thimm,

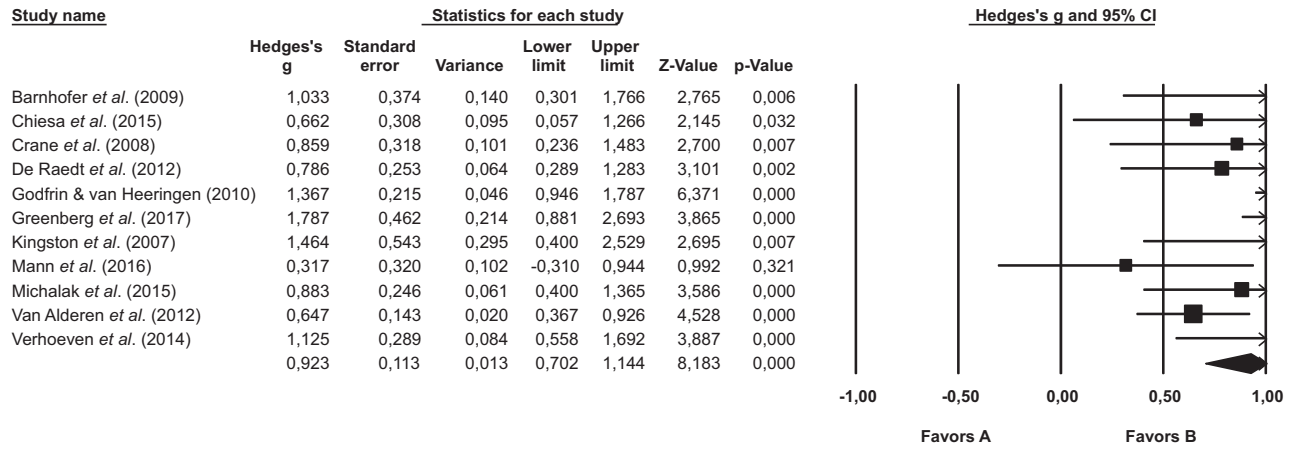


Fig. 2. Forest plot for between-group studies using the BDI and no-treatment control groups.

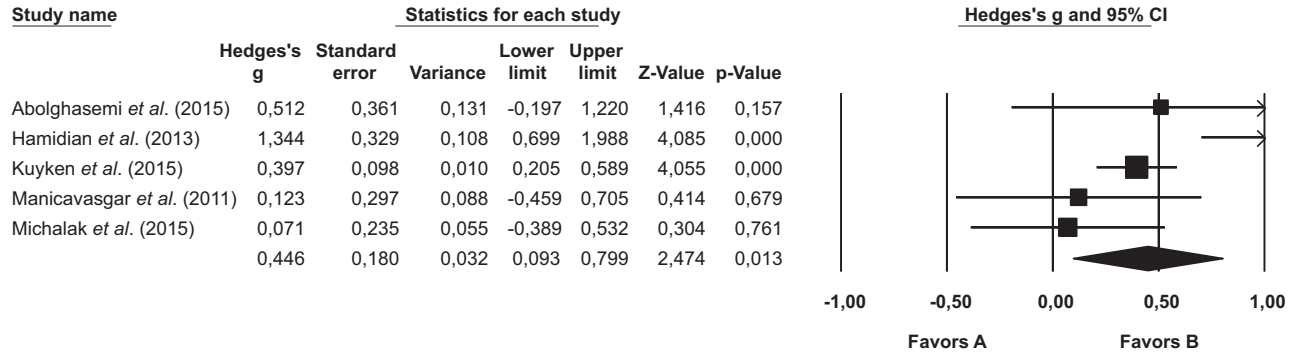


Fig. 3. Forest plot for between-group studies using the BDI and active treatment control groups.

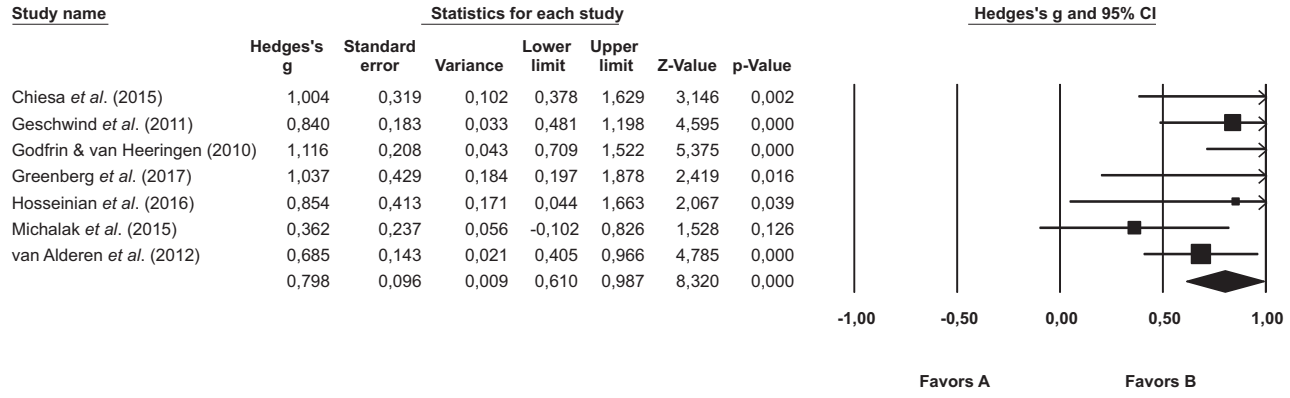


Fig. 4. Forest plot for between-group studies using the HDRS and no-treatment control groups.

2018) and have been related to the different ways of administration of the two instruments: the BDI is a self-report inventory, while the HDRS is rated by a clinician. In addition, the BDI assesses cognitive symptoms of depression to a higher degree than does the HDRS, which focuses more on somatic symptoms (cf. Wampold & Imel, 2015). Thus, while a decline in ESs for individual CBT for depression and an increase in ESs for group CBT have been observed (Cristea *et al.*, 2017; Johnsen & Friborg, 2015; Johnsen & Thimm, 2018), no effects of time for the ESs of MBCT were found. Neither were there any indications of potential trends towards a decline or increase in ES, as the regression line was nearly neutral (flat) for all statistical conditions. A probable reason for this finding is that studies of MBCT for depression have used heterogeneous samples from the beginning, i.e., included participants with various conditions in addition to depression and had no strict exclusion criteria (cf. Dobson, 2016). On the other hand, an improvement in ESs over time was not observed either. It can only be speculated whether the reported ESs of MBCT for current depression already represent the upper limit of its effectiveness or whether factors such as insufficient therapist training and supervision (cf. Waltman *et al.*, 2016) inhibit an increase of the effects.

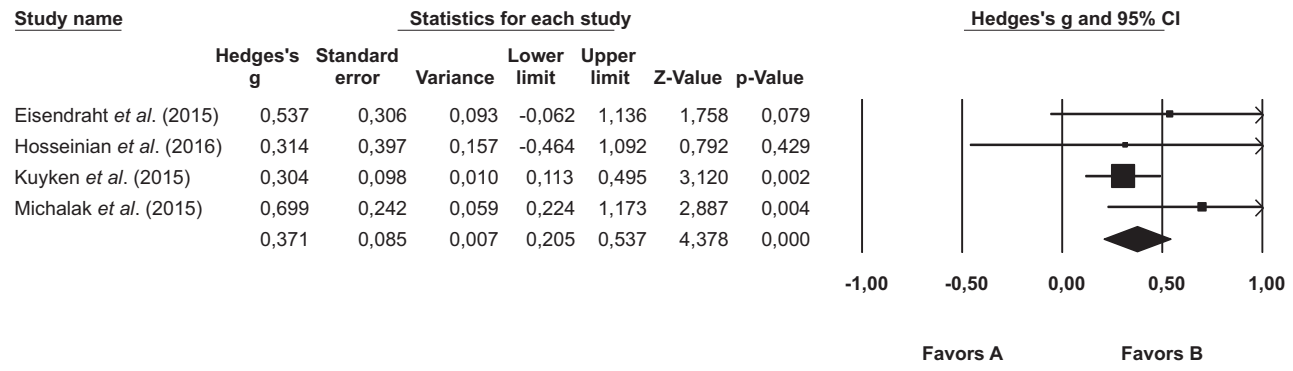


Fig. 5. Forest plot for between-group studies using the HDRS and active treatment control groups.

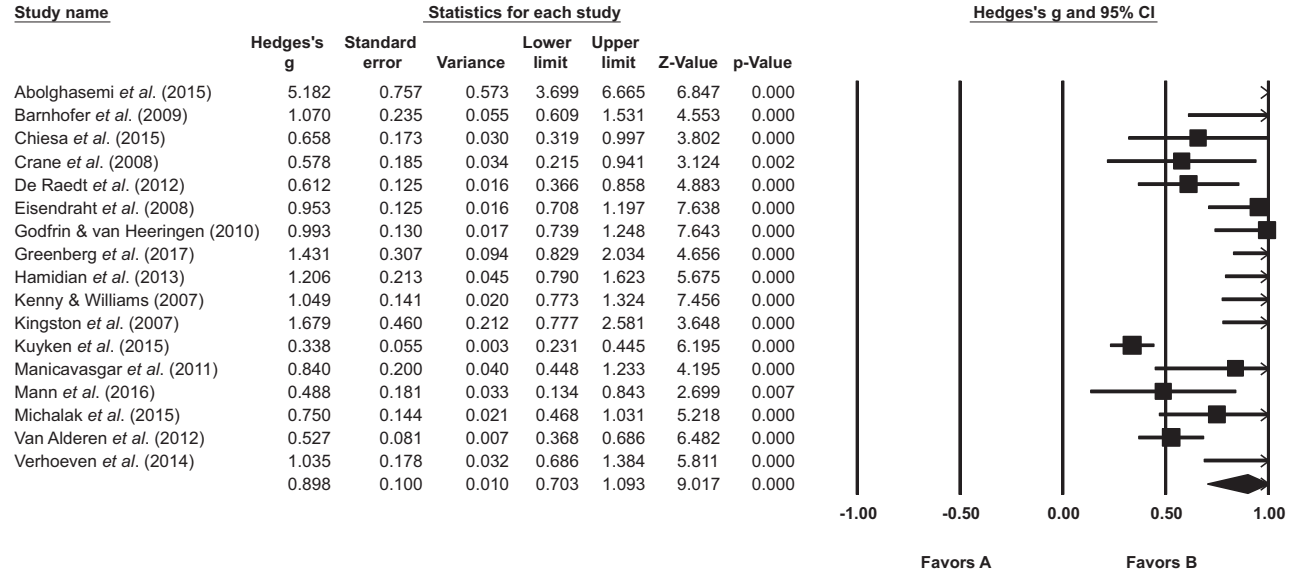


Fig. 6. Forest plot for within-group studies using the BDI.

As to the overall effect of MBCT for acute depression, the results of the present study are consistent with previous meta-analytic studies (e.g., Goldberg et al., in press; Lenz *et al.*, 2016), suggesting that MBCT is effective in reducing symptoms of depression. Applying Cohen's (1992) criteria, the average ESs for between-group studies comparing MBCT to no-treatment control conditions and pre-post studies were large for both outcome measures. Studies with active control conditions showed moderate average ESs in favor of MBCT. The trim and fill method indicated that publication bias was present for three of the five meta-analytic conditions. However, the estimated number of missing studies did not exceed two, suggesting that overall publication bias is low. Consistent with previous findings (e.g., Kühberger, Fritz, & Scherndl, 2014), there was a negative association between sample size and ES in within-group studies, i.e., studies with smaller samples tended to show higher ES than studies with larger samples. Similarly, higher baseline levels of depression were related to higher ESs. Based on the robust finding of the effectiveness of MBCT for current depression, it has been proposed that MBCT should be offered as a first-line treatment for depression on equal terms with other evidence-based treatments (Strauss *et al.*, 2014). However, more research is needed to support this claim. It should be noted that the average ES observed for MBCT when compared to no treatment comparisons is lower than those for other psychological treatments. For example, for individual and group CBT, average ESs of $g = 1.37$ and $g = 1.14$, respectively, have been reported for between-group studies using the BDI (Johnsen & Friborg, 2015;Johnsen & Thimm, 2018). The corresponding ES for MBCT in the current study was $g = 0.92$. Additionally, for pre-post comparisons, the average ES for MBCT observed in the present study ($g = 0.90$, $g = 0.82$ when the Abolghasemi *et al.* (2015) study was excluded) is smaller than those for individual and group CBT in clinical trials ($g = 1.65$ and $g = 1.33$, respectively; Johnsen & Friborg, 2015; Johnsen & Thimm, 2018) and in routine clinical practice ($d = 1.06$; Hans & Hiller, 2013).

When interpreting the results of the present investigation, several limitations have to be considered. Compared to previous examinations of temporal development of the effects of CBT for depression, the period in which the studies investigating the effects of MBCT for depression were conducted was relatively short. Further, the number of available studies was small. Chronicity of depression and an assessment of adherence to the MBCT manual was not reported in most publications and could,

therefore, not be included in the analyses. Particular caution is warranted when interpreting the results from within-studies due to data dependence (Cuijpers *et al.*, 2017). Heterogeneity was found to be significant for two of the analyses (between-group studies using an active treatment control group and the BDI and pre-post differences on the BDI). The I^2 index further indicated that heterogeneity was low for the two analytical conditions based on the HDRS ($I^2 = 0$, and $I^2 = 16.04$, respectively), while two of the conditions based on the BDI showed moderate ranges ($I^2 = 42.7$ and $I^2 = 64.2$, respectively). These values are highly acceptable, especially when taking into consideration that meta-analyses in the field of psychology are notorious for having large degrees of heterogeneity, as proven in a recent study examining rates of I^2 in 61 published meta-analyses in Psychological Bulletin between 1990 and 2013 (van Erp, Verhagen, Grasman, & Wagenmakers, 2017). The authors found that over half of the between-study meta-analyses showed $I^2 > 70$. For the final analysis in the present study, $I^2 = 86.7$ was found for the within–group condition. This is not uncommon, as higher degrees of heterogeneity are associated with within-group analyses. The finding may be due to the less rigid (and less precise) statistical requirements, as no control groups are implemented in the analysis, thus inherently leaving room for larger variability between the included studies. In addition, the high heterogeneity might be due to differences between studies in the efficacy of MBCT. Future studies should examine possible additional moderators, e.g., variables related to the implementation of the treatment. Finally, the present meta-analysis was restricted to studies that used a version of the BDI or HDRS as outcome measures. Including studies that used other instruments could lead to different findings.

In conclusion, the results of the present meta-analysis show that MBCT is effective in reducing symptoms of current depression and that study findings are stable over time. However, the relatively small number and short time range of the studies included in the analysis require further investigations in the future.

REFERENCES

Abolghasemi, A., Gholami, H., Narimani, M. & Gamji, M. (2015). The effect of beck's cognitive therapy and mindfulness-based cognitive therapy on sociotropic and autonomous personality styles in patients with depression. *Iranian Journal of Psychiatry and Behavioral Sciences*, *9*(4), e3665. https://doi.org/10.17795/ijpbs-3665

Baer, R.A. & Walsh, E. (2016). Treating acute depression with mindfulness-based cognitive therapy. In A. Wells, & P.L. Fisher (Eds.), *Treating depression* (pp. 344–368). New York: Wiley.

Barnhofer, T., Crane, C., Hargus, E., Amarasinghe, M., Winder, R. & Williams, J.M.G. (2009). Mindfulness-based cognitive therapy as a treatment for chronic depression: A preliminary study. *Behaviour Research and Therapy*, *47*, 366–373.

Beck, A.T., Steer, R.A. & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio: Psychological Corporation.

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

Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2017). *Comprehensive Meta-Analysis (version 3)*. Englewood: Biostat.

Chiesa, A., Castagner, V., Andrisano, C., Serretti, A., Mandelli, L., Porcelli, S. & Giommi, F. (2015). Mindfulness-based cognitive therapy vs. Psycho-education for patients with major depression who did not achieve remission following antidepressant treatment. *Psychiatry Research*, *226*, 474–483.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Crane, C., Barnhofer, T., Duggan, D.S., Hepburn, S., Fennell, M.V. & Williams, J. (2008). Mindfulness-based cognitive therapy and self-discrepancy in recovered depressed patients with a history of depression and suicidality. *Cognitive Therapy and Research*, *32*, 775–787.

Crane, R. (2009). *Mindfulness-based cognitive therapy: Distinctive features*. London: Routledge.

Cristea, I.A., Stefan, S., Karyotaki, E., David, D., Hollon, S.D. & Cuijpers, P. (2017). The effects of cognitive behavioral therapy are not systematically falling: A revision of Johnsen and Friborg (2015). *Psychological Bulletin*, *143*, 326–340.

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Canadian Psychology/Psychologie Canadienne*, *58*, 7–19.

Cuijpers, P., Weitz, E., Cristea, I.A. & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*, 364–368.

De Raedt, R., Baert, S., Demeyer, I., Goeleven, E., Raes, A., Visser, A., *et al*. (2012). Changes in attentional processing of emotional information following mindfulness-based cognitive therapy in people with a history of depression: Towards an open attention for all emotional experiences. *Cognitive Therapy and Research*, *36*, 612–620.

Dobson, K.S. (2016). The efficacy of cognitive-behavioral therapy for depression: Reflections on a critical discussion. *Clinical Psychology: Science and Practice*, *23*, 123–125.

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Eisendrath, S.J., Delucchi, K., Bitner, R., Fenimore, P., Smit, M. & McLane, M. (2008). Mindfulness-based cognitive therapy for treatment-resistant depression: A pilot study. *Psychotherapy and Psychosomatics*, *77*, 319–320.

Eisendrath, S.J., Gillung, E., Delucchi, K., Mathalon, D.H., Yang, T.T., Satre, D.D., *et al*. (2015). A preliminary study: Efficacy of mindfulness-based cognitive therapy versus sertraline as first-line treatments for major depressive disorder. *Mindfulness*, *6*, 475–482.

Farb, N., Anderson, A., Ravindran, A., Hawley, L., Irving, J., Mancuso, E., *et al*. (2018). Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. *Journal of Consulting and Clinical Psychology*, *86*, 200–204.

Galante, J., Iribarren, S.J. & Pearce, P.F. (2013). Effects of mindfulness-based cognitive therapy on mental disorders: A systematic review and meta-analysis of randomised controlled trials. *Journal of Research in Nursing*, *18*, 133–155.

Geschwind, N., Peeters, F., Drukker, M., Van Os, J. & Wichers, M. (2011). Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *79*, 618–628.

Godfrin, K.A. & van Heeringen, C. (2010). The effects of mindfulness-based cognitive therapy on recurrence of depressive episodes, mental health and quality of life: A randomized controlled study. *Behaviour Research & Therapy*, *48*, 738–746.

Goldberg, S.B., Tucker, R.P., Greene, P.A., Davidson, R.J., Kearney, D.J. & Simpson, T.L. (2019). Mindfulness-based cognitive therapy for the treatment of current depressive symptoms: A meta-analysis. *Cognitive Behavior Therapy*, *48*, 445–462.

Goldberg, S.B., Tucker, R.P., Greene, P.A., Davidson, R.J., Wampold, B.E., Kearney, D.J. & Simpson, T.L. (2018). Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *59*, 52–60.

Goyal, M., Singh, S., Sibinga, E.S., Gould, N.F., Rowland-Seymour, A., Sharma, R., *et al* (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. *JAMA Internal Medicine*, *174*, 357–368.

Greenberg, J., Shapero, B.G., Mischoulon, D. & Lazar, S.W. (2017). Mindfulness-based cognitive therapy for depressed individuals

improves suppression of irrelevant mental-sets. *European Archives of Psychiatry and Clinical Neuroscience*, *267*, 277–282.

Gu, J., Strauss, C., Bond, R. & Cavanagh, K. (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. *Clinical Psychology Review*, *37*, 1–12.

Hamidian, S., Omidi, A., Mousavinasab, S.M. & Naziri, G. (2013). Comparison of the effect of mindfulness-based cognitive therapy accompanied by pharmacotherapy with pharmacotherapy alone in treating dysthymic patients. *Iranian Red Crescent Medical Journal*, *15*, 239–244.

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

Hans, E. & Hiller, W. (2013). Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: A meta-analysis of nonrandomized effectiveness studies. *Journal of Consulting and Clinical Psychology*, *81*, 75–88.

Hedman-Lagerlöf, M., Hedman-Lagerlöf, E. & Öst, L.-G. (2018). The empirical support for mindfulness-based interventions for common psychiatric disorders: A systematic review and meta-analysis. *Psychological Medicine*, *48*, 2116–2129.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Hofmann, S.G., Sawyer, A.T., Witt, A.A. & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*, 169–183.

Hosseinian, S., Shahtaheri, E., Ebrahimi, M., Mahdavi, A. & Sepahvandi, M.A. (2016). Effectiveness of mindfulness-based cognitive therapy and metacognition therapy on reduction of symptoms of depression, core self-evaluation and emotion regulation difficulties in pregnant women with depression. *Acta Medica Mediterranea*, *32*, 2033–2037.

Jadad, A.R., Moore, R.A., Carroll, D., Jenkinson, C., Reynolds, D.J.M., Gavaghan, D.J. & McQuay, H.J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*, 1–12.

Johnsen, T.J. & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin*, *141*, 747–768.

Johnsen, T.J. & Thimm, J.C. (2018). A meta-analysis of group cognitive–behavioral therapy as an antidepressive treatment: Are we getting better? *Canadian Psychology/Psychologie Canadienne*, *59*, 15–30.

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York: Dell Publishing.

Kenny, M.A. & Williams, J.M.G. (2007). Treatment-resistant depressed patients show a good response to mindfulness-based cognitive therapy. *Behaviour Research and Therapy*, *45*, 617–625.

Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V., et al. (2013). Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review*, *33*, 763–771.

Kingston, T., Dooley, B., Bates, A., Lawlor, E. & Malone, K. (2007). Mindfulness-based cognitive therapy for residual depressive symptoms. *Psychology and Psychotherapy: Theory, Research and Practice*, *80*, 193–203.

Klainin-Yobas, P., Cho, M.A. & Creedy, D. (2012). Efficacy of mindfulness-based interventions on depressive symptoms among people with mental disorders: A meta-analysis. *International Journal of Nursing Studies*, *49*, 109–121.

Kühberger, A., Fritz, A. & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE*, *9*, e105825.

Kuyken, W., Hayes, R., Barrett, B., Byng, R., Dalgleish, T., Kessler, D., *et al*. (2015). Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): A randomised controlled trial. *Lancet*, *386*, 63–73.

Kuyken, W., Warren, F.C., Taylor, R.S., Whalley, B., Crane, C., Bondolfi, G., et al. (2016). Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: An individual patient data meta-analysis from randomized trials. *JAMA Psychiatry*, *73*, 565–574.

Lau, M. (2016). The theory underlying mindfulness-based cognitive therapy as a relapse prevention approach to depression. In A. Wells, & P.L. Fisher (Eds.), *Treating depression* (pp. 194–200). New York: Wiley.

Lau, M.A., Segal, Z.V. & Williams, J.M.G. (2004). Teasdale's differential activation hypothesis: Implications for mechanisms of depressive relapse and suicidal behaviour. *Behaviour Research and Therapy*, *42*, 1001–1017.

Lenz, A.S., Hall, J. & Bailey Smith, L. (2016). Meta-analysis of group mindfulness-based cognitive therapy for decreasing symptoms of acute depression. *Journal for Specialists in Group Work*, *41*, 44–70.

Manicavasgar, V., Parker, G. & Perich, T. (2011). Mindfulness-based cognitive therapy vs cognitive behaviour therapy as a treatment for non-melancholic depression. *Journal of Affective Disorders*, *130*, 138–144.

Mann, J., Kuyken, W., O'Mahen, H., Ukoumunne, O.C., Evans, A. & Ford, T. (2016). Manual development and pilot randomised controlled trial of mindfulness-based cognitive therapy versus usual care for parents with a history of depression. *Mindfulness*, *7*(5), 1024–1033.

McCarney, R.W., Schulz, J. & Grey, A.R. (2012). Effectiveness of mindfulness-based therapies in reducing symptoms of depression: A meta-analysis. *European Journal of Psychotherapy & Counselling*, *14*, 279–299.

McCrae, R.R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, *90*, 105–109.

Michalak, J., Schultze, M., Heidenreich, T. & Schramm, E. (2015). A randomized controlled trial on the efficacy of mindfulness-based cognitive therapy and a group version of cognitive behavioral analysis system of psychotherapy for chronically depressed patients. *Journal of Consulting and Clinical Psychology*, *83*, 951–963.

Piet, J. & Hougaard, E. (2011). The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review*, *31*, 1032–1040.

Rosenthal, R. (1993). *Meta-analytic procedures for social research*. Newbury Park: Sage.

Segal, Z.V., Williams, M. & Teasdale, J.D. (2002). *Mindfulness-based cognitive therapy for depression*. New York: Guilford.

Segal, Z.V., Williams, M., Teasdale, J.D. & Kabat-Zinn, J. (2013). *Mindfulness-based cognitive therapy for depression*. New York: Guilford.

Steinert, C., Hofmann, M., Kruse, J. & Leichsenring, F. (2014). The prospective long-term course of adult depression in general practice and the community. A systematic literature review. *Journal of Affective Disorders*, *152–154*, 65–75.

Strauss, C., Cavanagh, K., Oliver, A. & Pettman, D. (2014). Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: A meta-analysis of randomised controlled trials. *PLoS ONE*, *9*, e96110.

Teasdale, J.D., Segal, Z.V., Williams, J., 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.

van Aalderen, J., Donders, A., Giommi, F., Spinhoven, P., Barendregt, H. & Speckens, A. (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*, 989–1001.

van Erp, S., Verhagen, J., Grasman, R.P.P.P. & Wagenmakers, E.-J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in *Psychological Bulletin* from 1990–2013. *Journal of Open Psychology Data*, *5*, 4.

Verhoeven, J.E., Vrijsen, J.N., van Oostrom, I., Speckens, A.E. & Rinck, M. (2014). Attention effects of mindfulness-based cognitive therapy in

formerly depressed patients. *Journal of Experimental Psychopathology*, *5*, 414–424.
Waltman, S.H., Creed, T.A. & Beck, A.T. (2016). Are the effects of cognitive behavior therapy for depression falling? Review and critique of the evidence. *Clinical Psychology: Science and Practice*, *23*, 113–122.
Wampold, B.E. & Imel, Z.E. (2015). *The great psychotherapy debate*. New York: Routledge.
Wang, Y.-Y., Li, X.-H., Zheng, W., Xu, Z.-Y., Ng, C.H., Ungvari, G.S., Xiang, Y.-T. (2018). Mindfulness-based interventions for major depressive disorder: A comprehensive meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, *229*, 429–436.
Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *The Lancet*, *382*, 1575–1586.
World Health Organization (2017). *Depression and other common mental disorders: Global health estimates*. Geneve: World Health Organzation.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Fig S1. Funnel plot for observed and imputed between-group studies using the BDI and no-treatment control groups.
Fig S2. Funnel plot for observed and imputed between-group studies using the BDI and active treatment control groups.
Fig S3. Funnel plot for observed and imputed between-group studies using the HDRS and no-treatment control groups.
Fig S4. Funnel plot for observed and imputed between-group studies using the HDRS and active treatment control groups.
Fig S5. Funnel plot for observed and imputed within-group studies using the BDI.

Received 16 April 2019, accepted 26 February 2020