

Clinical effectiveness of cognitive therapy *v.* interpersonal psychotherapy for depression: results of a randomized controlled trial

L. H. J. M. Lemmens^{1*}, A. Arntz^{1,2}, F. Peeters^{3,4}, S. D. Hollon⁵, A. Roefs¹ and M. J. H. Huibers^{1,6}

¹Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

²Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

³Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, Maastricht, The Netherlands

⁴Faculty of Health, Medicine and Life Sciences, Maastricht University, School for Mental Health and Neuroscience, Maastricht, The Netherlands

⁵Department of Psychology, Vanderbilt University, Nashville, TN, USA

⁶Department of Clinical Psychology, VU University Amsterdam, Amsterdam, The Netherlands

Background. Although both cognitive therapy (CT) and interpersonal psychotherapy (IPT) have been shown to be effective treatments for major depressive disorder (MDD), it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder. This study examined the clinical effectiveness of CT *v.* IPT in a large sample of depressed patients seeking treatment in a Dutch outpatient mental health clinic. We tested whether one of the treatments was superior to the other at post-treatment and at 5 months follow-up. Furthermore, we tested whether active treatment was superior to no treatment. We also assessed whether initial depression severity moderated the effect of time and condition and tested for therapist differences.

Method. Depressed adults ($n = 182$) were randomized to either CT ($n = 76$), IPT ($n = 75$) or a 2-month waiting list control (WLC) condition ($n = 31$). Main outcome was depression severity, measured with the Beck Depression Inventory – II (BDI-II), assessed at baseline, 2, 3, and 7 months (treatment phase) and monthly up to 5 months follow-up (8–12 months).

Results. No differential effects between CT and IPT were found. Both treatments exceeded response in the WLC condition, and led to considerable improvement in depression severity that was sustained up to 1 year. Baseline depression severity did not moderate the effect of time and condition.

Conclusions. Within our power and time ranges, CT and IPT appeared not to differ in the treatment of depression in the acute phase and beyond.

Received 17 March 2014; Revised 21 December 2014; Accepted 21 December 2014

Key words: Cognitive therapy, depressive disorders, individual psychotherapy, interpersonal psychotherapy, outcome studies, randomized controlled trial.

Introduction

Both cognitive therapy (CT) and interpersonal psychotherapy (IPT) have proven to be well-standardized, effective treatments for major depressive disorder (MDD; Cuijpers *et al.* 2008). However, it is not clear yet whether one therapy outperforms the other with regard to severity and course of the disorder. Recently, Jakobsen *et al.* (2012) meta-analysed results of five randomized controlled trials (RCT) comparing post-treatment effects of CT and IPT on the Beck

Depression Inventory (BDI). Three trials examined individual CT and IPT (Elkin *et al.* 1989; Luty *et al.* 2007; Quilty *et al.* 2008), one study examined the effects of group therapy (Bellino *et al.* 2007), and one study included psychodynamic IPT instead of traditional IPT (Hardy *et al.* 1995). Meta-analysis of the three studies that examined individual CT and IPT showed no significant differences in post-treatment depression severity scores between the two interventions [mean difference in favour of CT = 0.62 points on the BDI-II, $p = 0.59$, 95% confidence interval (CI) 2.86–1.61][†]. Even though existing research consistently suggests that both treatments are equally effective, additional trials are necessary. As Jakobsen *et al.* (2012) pointed

* Address for correspondence: L. Lemmens, MSc, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, Maastricht, The Netherlands.

(Email: Lotte.Lemmens@Maastrichtuniversity.nl)

† The notes appear after the main text.

out, so far the data collected in randomized comparisons of CT and IPT are insufficient to reliably decide whether the two therapies have differential effects on the BDI. More specifically, trial sequential analysis on the three existing studies examining the effects of individual CT and IPT for depression ($n = 388$), indicated that data of at least 108 more patients is necessary to detect or reject an intervention effect with a minimal relevant difference of 4 BDI points (80% power). For an effect of 3 BDI points, this is much higher (791 additional patients)². Furthermore, all studies included in Jakobsen's meta-analysis were classified as having a high risk of bias (especially Elkin *et al.* 1989), and none of them included data after treatment termination, leaving long-term effects largely unknown. Another unresolved issue in the comparison of CT and IPT is whether one of the two treatments is indicated for severe depression, as two studies found contradictory evidence: Elkin *et al.* (1989) favoured IPT, whereas Luty *et al.* (2007) found that CT was more effective for severe depression. In addition, treatment moderators and therapist effects are less well studied, but may be at least as important as treatment protocol. Last, it has been frequently argued in psychotherapy research that the effects of treatment studies cannot be attributed to the intervention(s) studied unless a placebo or no-treatment group is included (Klein, 1990). Even though several studies have individually compared CT or IPT to a non-active waiting list control (WLC) condition, to date there are no randomized comparisons of CT, IPT and a WLC condition.

Study aims

To address these issues, the current study examined the clinical effectiveness of CT and IPT in a large sample of depressed patients seeking treatment in an outpatient mental health clinic in The Netherlands. The aim of the study was threefold: first, we tested whether one of the treatments was superior to the other in the reduction of depressive symptoms as assessed with the BDI-II and on a set of secondary outcomes at post-treatment (7 months) and up to 5-months follow-up. Second, we tested for therapist differences and examined several potential moderators including baseline depression severity and total number of sessions. Third, and final, we examined whether short-term effects of active treatment were superior to waiting-list control after 2 months. With this we add to previous studies conducted in the USA and New Zealand with a large European sample. In order to determine the relative contribution of our study to the field, we added our data to the existing evidence base of trials that examined individual CT and IPT (Elkin *et al.* 1989; Luty *et al.* 2007; Quilty *et al.* 2008),

and meta-analysed the findings. Furthermore, using the same set of four studies, trial sequential analysis was performed to determine whether our study added the information necessary to conclude equivalence between CT and IPT. Finally, the risk of bias in our study was assessed.

Method

Trial design

Details about the study design have been fully described elsewhere (Lemmens *et al.* 2011). The data came from a large single-centre RCT (parallel group design) examining the effectiveness and mechanisms of change of individual CT and IPT for depression. In this study, depressed outpatients were randomly allocated to one of three conditions: (a) CT, (b) IPT, (c) a 2-month WLC condition followed by treatment of choice. In reporting the follow-up period of the study, three phases can be distinguished; (1) the 'Treatment phase' (baseline to 7 months); the period when therapy was delivered; (2) the 'Trial follow-up phase' (months 8–12) in which depression severity was measured monthly; and (3) the 'Long-term follow-up phase' (months 12–24); the period covered by retrospective assessment at 24 months. The data presented here concern the acute outcome of therapy at the end of the 7-month treatment phase and data collected in the trial follow-up phase (henceforth called follow-up). The Medical Ethics Committee of Maastricht University approved the study protocol, and all participants provided written informed consent. The study is registered at The Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN 67561918).

Participants and recruitment

Patients were adult outpatients (18–65 years³) referred to the mood disorder unit of the Maastricht Community Mental Health Centre with a primary diagnosis of MDD as confirmed by the Structural Clinical Interview for DSM-IV Axis I disorders (SCID-I; First *et al.* 1997) conducted by a trained evaluator. Further inclusion criteria were: internet access, an email address, and sufficient knowledge of the Dutch language. Exclusion criteria were: bipolar or chronic (current episode >5 years) depression, elevated acute suicide risk, concomitant pharmacological or psychological treatment⁴, drugs and alcohol abuse/dependence, and mental retardation ($IQ < 80$).

Procedure

Participants were recruited during regular intakes at our clinical site. After informed consent was obtained

and a baseline assessment was completed, the researcher pressed the 'assign' button on the computer screen, after which the database randomly allocated the participant to one of three conditions using computer-generated block randomization (10:10:4). The random allocation sequence was generated by an independent computer scientist and was concealed from the researchers who were involved in the randomization procedure to prevent prediction of future assignment (Lemmens *et al.* 2011). Randomization was pre-stratified according to presence or absence of previous episodes. Blinding of patients and therapists for treatment condition was not possible. Sample size calculations of the active conditions CT and IPT were based on long-term expectations of CT *v.* IPT (Lemmens *et al.* 2011). By combining rudimentary findings from previous randomized trials, a 20% difference in relapse/recurrence rate between the conditions was expected at the end of the 2-year follow-up period. An *a-priori* power analysis indicated that with 75 patients per arm in the active conditions and taking 15% attrition into account, the study was powered at 80% (two-tailed $\alpha=0.05$) to detect a CT-IPT 20% difference in relapse rates. With this sample size, taking 15% attrition into account, the study is powered at 80% to detect a medium effect size difference (Cohen's $d=0.50$) in dimensional outcomes at $p=0.05$, as in the present study. A second power calculation (powered at 80%, $\alpha=0.05$) showed that 28 patients in the WLC condition would suffice to detect a statistically significant difference of six BDI-II points between the two active conditions combined and the waiting list at 2 months (Cohen's $d=0.58$)⁵. The main assessment points for the active conditions (CT and IPT) were baseline, 3, 7, 9 and 12 months. The primary outcome was also assessed at 8, 10 and 11 months. All assessments were administered on a computer. Pre- and post-treatment assessments (baseline and 7 months) were administered at the research centre. Mid-treatment (3 months) and follow-up assessments (8–12 months) took place via the internet. In addition, in order to examine the effect of the waiting list, depression severity of the total sample was assessed at 2 months⁶.

Interventions

Treatment consisted of 16–20 individual sessions of 45 min, depending on the progress of the individual patient. The protocol allowed flexibility in scheduling appointments less often than weekly. Patients were considered to have had an adequate dose of treatment if they attended at least 12 sessions⁷. Both interventions were described in a treatment manual. The CT protocol was based on the manual by Beck *et al.* (1979) which

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

Treatment integrity

Treatment integrity and boundaries of both therapies were carefully monitored during the therapist training phase and the study itself. To prevent contamination, therapists were uniquely assigned to one of the treatment conditions. Furthermore, prior to the study, they received additional training by Steven Hollon (CT) and John Markowitz (IPT), experts in the respective interventions. The training (2 × 8 h) addressed the theoretical framework and skills for each treatment using tailored case examples and role-play. During the study, therapists and the investigators who also participated as therapists in the study (M.H. for CT and F.P. for IPT) met bi-weekly in separate consultation sessions for each treatment condition to discuss their caseload. In addition, all sessions were videotaped. Four independent raters evaluated a random selection of 106 tapes on treatment adherence (content and quality) and competence. The raters (three psychotherapists and one psychologist) were highly experienced in providing, teaching and supervising CT and IPT. All raters were trained prior to the study and masked to treatment outcome. Nine tapes were double-coded in order to assess inter-rater reliability. To examine competence in both therapies, the overall quality scores of the Cognitive Therapy

Scale (Dobson *et al.* 1985) and a short version of the IPT Adherence and Quality Scale (Stuart, 2011) were used. In order to obtain a measure of adherence, the Collaborative Study Psychotherapy Rating Scale – version 6 (CSPRS-6; Hollon *et al.* 1984, 1988) was used. This 96-item questionnaire is able to distinguish between CT, IPT and clinical management (Hill *et al.* 1992). Items are rated on a 7-point Likert scale ('not at all' to 'extensively') and higher scores indicate better adherence. Following Luty *et al.* (2007), we modified the original version to distinguish only between CT and IPT by omitting the 20 items pertinent to clinical management and medication, reducing the scale to 76 items.

Outcomes

Primary outcome was depression severity as measured with the Beck Depression Inventory, second edition (BDI-II; Beck *et al.* 1996). Scores range from 0 to 63 with higher scores indicating more severe symptoms of depression. Secondary outcome measures included the following: general psychological distress was assessed with the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983). A high score is indicative of severe distress (range 0–212). To obtain a measure of impairment in social functioning due to depression, the Work and Social Adjustment Scale (WSAS; Mundt *et al.* 2002) was used. Scores range from 0 (no impairment) to 40 (severe impairment). Besides clinical outcome measures, several quality-of-life measures were included because they have been shown to provide an important additional dimension to the evaluation of treatments for depression (Kennedy *et al.* 2001; Jakobsen *et al.* 2012). The RAND-36 (van der Zee & Sanderman, 1993) was used to measure quality of life in terms of impairments due to physical and mental health status. Scores were transformed to a 100-point scale with higher scores indicating more positive experiential quality of life. The EuroQol-6D (EQ-6D; EuroQolGroup, 1990) was used to obtain an overall utility score for population-based quality of life (EQ-5D)⁸. Utilities refer to the preference for any particular set of health outcomes and are indicated by a number between 0 (worst imaginable condition) and 1 (perfect health; Dolan, 1997; Drummond *et al.* 2005). For the current study, the Dutch value set was used (Lamers *et al.* 2006). For a more detailed description of these instruments and their psychometric properties we refer to our protocol paper (Lemmens *et al.* 2011).

Data analysis

A detailed description of statistical procedures is presented in online Data Supplement I. First, we mapped

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

To examine whether CT and IPT differed in the reduction of depressive symptoms mixed (multilevel) regression analysis was used. We applied intention-to-treat analysis, with the BDI-II as the primary outcome. Our basic model was a two-level (patients and measurements) repeated-measures design with depression severity as dependent variable, condition as a between-subjects factor and time of measurement in weeks as a within-subjects factor. The difference between CT and IPT was represented by the time \times condition interaction in the model. Furthermore, in all analyses, we controlled for baseline EQ-5D and BDI-II scores because they showed considerable differences between the groups. Because visual inspection showed separate linear time slopes for the acute (0–7 months) and the follow-up (7–12 months) phase, change over time was assessed in two separate analyses (one for the acute phase, and one for the follow-up). Furthermore, change on secondary outcomes was assessed by testing the two-level basic model on dependent variables BSI, WSAS, RAND-36 and EQ-5D. For BDI-II and EQ-5D mixed gamma regression was used in the follow-up because of skewed distributions. Effect sizes, Cohen's d and r (Cohen, 1988) for the continuous primary and secondary outcomes were computed from the multilevel estimates.

Next, we tested whether initial depression severity moderated the effect of time and condition by adding the two- and three-way interaction(s) of baseline depression severity (BDI-II) with time and condition to the basic model of the primary outcome BDI-II. Subsequently, we checked for influence of therapist and number of sessions by univariately adding them as fixed factors (main effect, and interaction with time) to the final model. Several other baseline characteristics that displayed potentially relevant differences between the treatment groups (gender, work, and marital status) were added to the model as covariates to see whether they would affect the results. All effects were tested at the $p < 0.05$ level (two-tailed). Then, we examined whether therapy outperformed the waiting list by comparing change in BDI-II scores of patients in the active groups after 2 months of therapy with those of patients in the WLC condition after 2 months of no-treatment. All analyses were performed with

SPSS v. 21.0 (SPSS Inc., USA) and results are reported according to the CONSORT guidelines for reporting trials (Moher *et al.* 2010). Furthermore, the methodology of Jacobson & Truax (1991) was used to determine the proportion of patients that showed clinically meaningful change on the BDI-II. Response (the minimum amount of decrease in symptoms that has to be accomplished during therapy) was defined as a decrease of at least 9 BDI-II points during the treatment phase. Remission (the cut-off point between healthy and 'ill') was defined as an absolute value of ≤ 9 on the BDI-II. Frequency differences in response and remission rates between the groups were examined using mixed binary logistic regression.

To determine the relative contribution of our study to the field, we added our BDI-II data to the existing evidence base and meta-analysed findings from all four randomized trials that examined individual CT and IPT (Elkin *et al.* 1989; Luty *et al.* 2007; Quilty *et al.* 2008; current study) using the statistical program Open Meta Analysis (Wallace *et al.* 2012).

Since existing trials did not adjust their outcome variable according to baseline values, non-covariate-corrected means were included as estimates of the effects of the current study. As a sensitivity analysis, we repeated the analysis with (baseline BDI-II and EQ-5D) corrected post-treatment BDI-II score. The same set of four studies was used to perform trial sequential analysis on the BDI-II. Following Jakobsen *et al.* (2012), we conducted two analyses; one with a minimal relevant difference of 4 BDI points and 80% power, and one with more strict presumptions (BDI difference of 3 points and 90% power). Both analyses were based on a type I error of 5% and on the variance of all trials. Similar to the meta-analysis, effects of both non-covariate-corrected as well as covariate-corrected means as estimates of the effects of the current study were explored. In addition, following Jakobsen *et al.* (2012), bias risk was assessed with regard to sequence generation, allocation concealment, intention-to-treat analysis, blinding, drop-out, outcome measure reporting, presence of economic and academic bias (see online Data Supplement V for a full description of the criteria). An independent rater⁹ checked the generated table entries for accuracy.

Results

Description of the sample

Patient flow is shown in Fig. 1. Of the 1562 patients who were initially screened for eligibility, 1191 did not meet inclusion criteria [mainly because of the use of antidepressant medication ($n=362$) or MDD not being the primary diagnosis ($n=434$)], 78 patients

met inclusion criteria but declined to participate, and 111 were excluded for other reasons. A total of 182 patients were randomized ($n=76$, $n=75$, $n=31$ for CT, IPT and WLC, respectively).

Pre-treatment characteristics of the total sample stratified according to condition are displayed in Table 1. There were no relevant differences between the patients in the two treatment conditions combined and the WLC condition for any of the sociodemographic variables or depression specifiers. However CT and IPT showed considerable differences on the BDI-II and EQ-5D. Therefore, we controlled for this in all analyses. Mean age in the total group was 41.2 (s.d. = 12.1) years and 100 participants (66.2%) were female. Mean baseline BDI-II score was 29.8 (s.d. = 9.0) and 49% of the patients were diagnosed with recurrent depression. More than half of the patients (55.6%) suffered from severe depression (BDI-II score ≥ 29) and the majority (59.6%) was still actively employed.

Therapists

The 10 therapists (five for each intervention) who participated in the study were licensed psychologists, psychotherapists and psychiatrists. They ranged in age from 24 to 47 years and 80% were female. At study onset, therapists had an average of 9.1 (s.d. = 5.4) years of clinical experience (range 4–21 years). Each therapist treated between 7 and 30 patients, except for one therapist who left the clinical site after treating only two patients. No differences were found between the CT and IPT therapists with regard to years of experience ($t_8 = 1.76$, $p = 0.12$) or the number of patients treated ($t_7 = 0.07$, $p = 0.95$)¹⁰.

Treatment and study compliance, and treatment integrity

Recruitment took place from February 2007 to April 2012. Treatment was delivered until December 2012. Follow-up data was complete in April 2013. At treatment termination, 117 patients (78%) in the active groups completed at least 12 sessions of therapy, with the mean number of sessions being 17 (s.d. = 2.9) in both conditions. There were no significant differences in session scheduling between CT and IPT¹¹. Of the 34 patients that attended fewer than 12 sessions, eight were early completers (<12 sessions because they were no longer experiencing symptoms), two (one per condition) never started treatment (zero sessions), and the remaining 24 withdrew or were lost to follow-up. Non-completion rates were similar for the two intervention groups. There were no baseline differences between (early) treatment completers and non-completers.

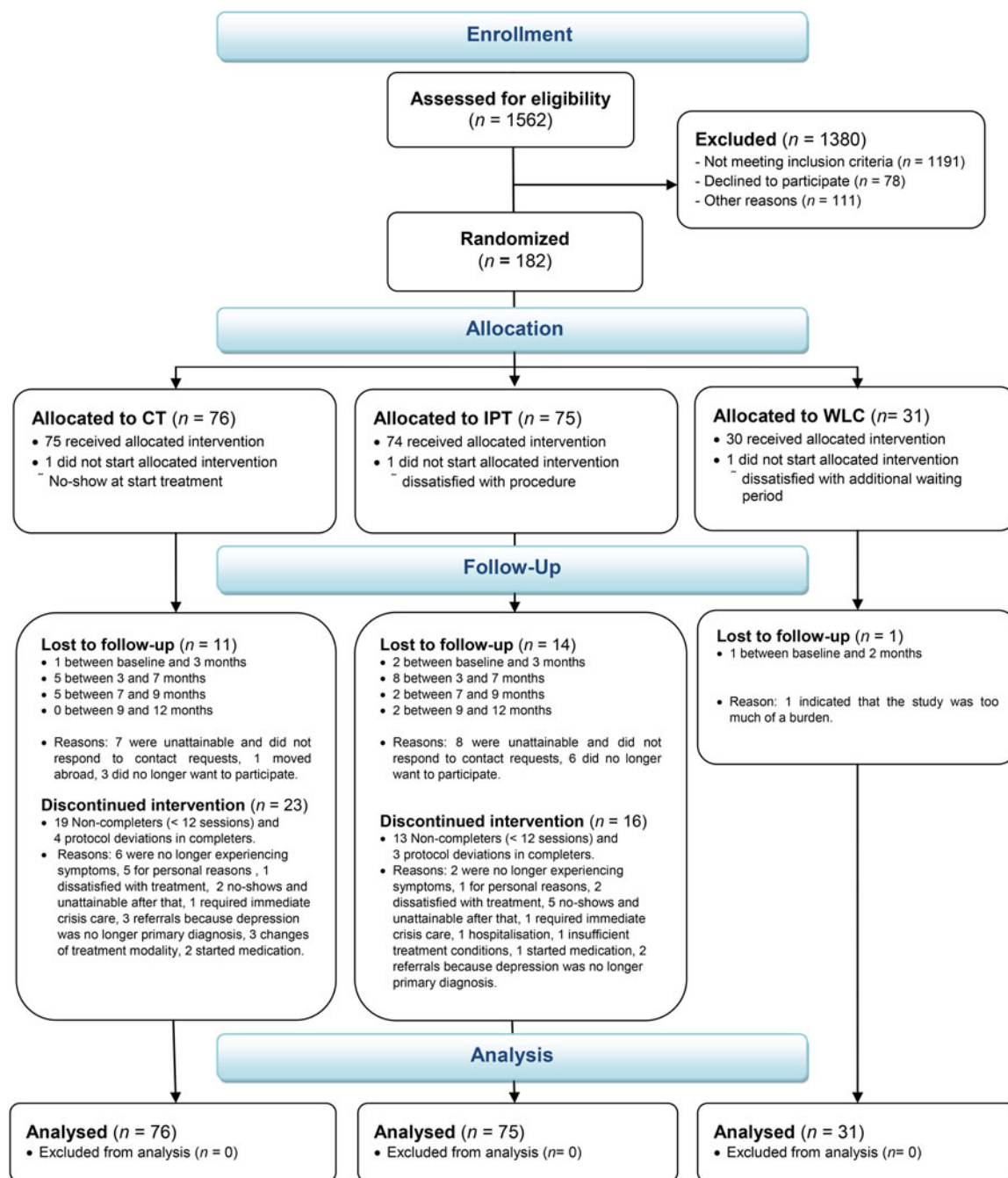


Fig. 1. Patient flow chart.

Quality of therapy was rated as being (very) good to excellent in 83.3% of the CT tapes. For IPT this was 90.4%. Analysis of CSPRS-6 scores revealed significant differences in therapy-specific behaviour between conditions, indicating that therapists adhered to protocol. CT-specific behaviour was significantly more evident in CT than in IPT (mean score of 80.80 *v.* 52.42, $t_{79.21} = 7.23$, $p < 0.001$). Scores on IPT subscales in turn, were higher in IPT than in CT (85.75 *v.* 44.57,

$t_{86.96} = 10.79$, $p < 0.001$). No significant differences between therapists were found. Mean intraclass correlation coefficient (ICC) across subscales of double-rated tapes was 0.63 (range 0.50–0.75).

The percentages of participants in the active conditions completing all assessments at 3 and 7 months were 98% and 89%, respectively. Follow-up assessments at 9 and 12 months were completed by 85% of the patients. Patients lost to follow-up were

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

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

BDI-II, Beck Depression Inventory, second edition.

^a Data unavailable for CT ($n=1$).

^b BDI-II score ≥ 29 .

significantly younger (36 *v.* 42 years; $t_{149}=2.13$, $p=0.03$) and more severely depressed at baseline (33.5 *v.* 29.0; $t_{149}=2.29$, $p=0.02$) compared to patients who completed all assessments. No significant differences in attrition rates emerged across conditions (see Fig. 1).

Treatment outcomes

Descriptive statistics

Table 2 presents the observed mean (95% CI) scores and corrected mixed-model estimated means (95% CI) on all outcome measures at each time-point in the treatment groups CT and IPT ($n=151$). Estimated means from mixed regression are more valid than observed values because they correct for baseline BDI-II and EQ-5D differences and take missing values into consideration.

Results on primary and secondary outcome measures

Table 3 shows the final models of the mixed regression analysis on the primary and secondary outcomes in the treatment phase¹² and in the follow-up. All time \times condition interactions were non-significant, indicating that there were no differential effects in symptom reduction between the two intervention groups. In line with this, no group differences were found on the mean scores at each time point (see Table 2). The analysis of the basic model (time, condition, time \times condition controlling for baseline BDI-II and EQ5D) on the primary outcome BDI-II further showed a significant main effect of time in the treatment phase, indicating that depressive

symptoms significantly decreased during treatment. The non-significant effect of time in the follow-up phase shows that BDI-II scores remained stable up until 5 months after finishing treatment. A graphic representation of change in depressive symptoms over time as measured with the BDI-II is shown in Fig. 2a. Analyses on secondary outcomes yielded a similar pattern of results, except for the BSI which also showed a significant effect of time in the follow-up phase (see Table 3).

Effect sizes

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

Moderation analyses

The moderation analysis of baseline severity (BDI-II) in the acute phase (results not shown in Table 3) did not show a time \times condition \times baseline severity interaction ($F_{1,755}=0.05$, $p=0.83$). After removing the three-way interaction from the model, the two-way interaction between treatment condition and baseline severity was still not significant ($F_{1,757}=1.10$, $p=0.29$), indicating that baseline severity did not moderate treatment outcome in the acute phase. The effect of therapist was not

Table 2. Observed mean scores (95% CI), corrected mixed regression-based estimated means (95% CI) and mean differences (95% CI, *p*) between CT and IPT for all outcome measures in the intention-to-treat sample (*n* = 151)

	Observed means		Estimated means*		CT – IPT	<i>p</i>
	CT (<i>n</i> = 76)	IPT (<i>n</i> = 75)	CT (<i>n</i> = 76)	IPT (<i>n</i> = 75)		
Beck Depression Inventory – II						
Baseline	28.4 (26.4–30.4)	31.2 (29.1–33.2)	29.6 (28.2–31.0)	29.6 (28.2–31.0)		
3 months	22.4 (19.8–24.9)	24.1 (21.0–27.1)	22.6 (20.4–25.1)	20.9 (18.8–23.2)	1.73 (–1.50 to 4.96)	0.29
7 months	13.7 (11.1–16.3)	16.0 (12.7–19.3)	14.1 (11.8–16.8)	13.9 (11.6–16.7)	0.14 (–3.44 to 3.72)	0.94
9 months	13.3 (10.5–16.2)	15.8 (12.2–19.4)	14.3 (11.8–17.3)	13.2 (10.8–16.1)	1.11 (–2.72 to 4.93)	0.57
12 months	12.6 (9.7–15.6)	17.2 (13.3–21.0)	13.4 (11.0–16.3)	14.9 (12.2–18.1)	–1.46 (–5.41 to 2.49)	0.47
Brief Symptom Inventory						
Baseline	65.9 (58.9–73.0)	71.1 (64.4–77.7)	66.2 (61.6–71.0)	62.7 (58.4–67.3)	3.49 (–3.06 to 10.04)	0.30
7 months	38.6 (12.9–17.9)	45.5 (13.0–17.5)	39.8 (33.4–47.4)	40.3 (33.8–48.2)	–0.54 (–10.54 to 9.46)	0.92
12 months	29.5 (21.8–37.2)	42.5 (32.5–52.5)	31.0 (25.0–8.3)	37.6 (30.2–46.7)	–6.60 (–17.15 to 3.94)	0.22
Work and Social Adjustment Scale						
Baseline	23.2 (21.4–25.0)	22.4 (20.7–24.0)	23.5 (21.9–25.2)	21.7 (20.2–23.2)	1.83 (–0.42 to 4.08)	0.11
7 months	15.4 (12.9–17.9)	15.2 (13.0–17.5)	15.6 (13.5–18.1)	14.3 (12.3–16.6)	1.34 (–1.77 to 4.46)	0.40
12 months	12.6 (10.1–15.1)	14.0 (11.5–16.4)	13.3 (11.1–15.8)	13.2 (11.0–15.8)	0.07 (–3.31 to 3.45)	0.97
RAND-36 Quality of Life						
Baseline	46.7 (42.9–50.5)	43.8 (40.6–47.0)	45.7 (42.5–49.0)	45.2 (42.0–48.5)	0.49 (–4.18 to 5.15)	0.84
7 months	56.9 (51.6–62.1)	54.5 (49.3–59.6)	56.4 (51.7–61.0)	55.7 (50.9–60.4)	0.72 (–5.95 to 7.39)	0.83
12 months	58.3 (52.9–63.7)	56.2 (51.0–61.5)	58.1 (53.2–63.0)	56.7 (51.7–61.7)	1.37 (–5.63 to 8.37)	0.70
EuroQol-5D Utility score						
Baseline	0.63 (0.58–0.69)	0.55 (0.49–0.62)	0.63 (0.57–0.67)	0.58 (0.52–0.63)	0.04 (–0.03 to 0.12)	0.26
7 months	0.77 (0.72–0.83)	0.77 (0.70–0.82)	0.78 (0.72–0.83)	0.78 (0.72–0.83)	0.00 (–0.08 to 0.08)	0.99
12 months	0.79 (0.73–0.85)	0.79 (0.75–0.86)	0.78 (0.70–0.83)	0.81 (0.74–0.85)	–0.03 (–0.12 to 0.05)	0.48

CT, Cognitive Therapy; IPT, Interpersonal Psychotherapy; CI, confidence interval.

*Mixed regression-based estimated means corrected for BDI-II and EQ-5D baseline scores (except for models in which BDI-II or EQ-5D were dependent variables); Mean difference scores (95% CI, *p*) are based on the corrected mixed-model estimates because they are more valid than observed values because they correct for baseline BDI-II and EQ-5D differences and take missing values into consideration; Data unavailable for 3, 17, 23, and 25 patients at 3, 7, 9, and 12 months, respectively.

significant either ($F_{1,267} = 0.35$, $p = 0.93$; see online Data Supplement II), indicating that individual differences between therapists did not influence treatment outcome. In addition, adjustment for number of sessions or potential confounds (gender, work, and marital status) did not affect the results in (see online Data Supplement II). In the follow-up phase, adjustment for baseline severity, as well as for the other potential confounds (total number of sessions, therapist, gender, work, and marital status) did not affect the results of the effectiveness analysis as well (see online Data Supplement III).

Active treatment groups v. waiting list

Response to therapy exceeded response in the WLC condition. Patients in the WLC condition showed minimal changes in depression severity across the 2-month waiting list period, suggesting that there was no

spontaneous recovery (see Fig. 2b). Improvement in the active treatment conditions after 2 months of therapy (mean number of sessions: 6.5, *s.d.* = 1.7) was significantly larger (mean BDI-II change difference between active conditions combined and WLC condition: 6.16, $t_{173} = 2.47$, $p < 0.02$, 95% CI 1.21–10.05, $r = 0.18$).

Response and remission

A complete overview of the observed and mixed regression based estimated response and remission rates of participants in each group for each time point, can be found in online Data Supplement IV. Overall, observed clinical improvement after 3 months was modest, yielding remission in 13% of the patients. After 7 months, remission in the total sample was 34%. Remission rates of 37% at 9 and 12 months show that treatment effects remained stable after treatment

Table 3. Results of analyses on primary and secondary outcome measures in the treatment phase (0–7 months) and trial follow-up phase (7–12 months)

	Treatment phase (0–7 months)						Trial FU phase (7–12 months)					
	<i>B</i>	95% CI	<i>F</i>	d.f.	<i>r</i> *	<i>p</i>	<i>B</i>	95% CI	<i>F</i>	d.f.	<i>r</i> *	<i>p</i>
Primary outcome analysis												
Beck Depression Inventory – II ^a												
Intercept	2.71	(2.59 to 2.83)	37.59	274	0.35	<0.001	2.72	(2.60 to 2.85)	10.72	758	0.12	<0.001
Baseline severity	0.35	(0.28 to 0.43)	88.74	274	0.49	<0.001	0.34	(0.21 to 0.46)	27.38	758	0.19	<0.001
Baseline QoL	–0.10	(–0.17 to –0.02)	6.30	274	0.15	0.01	–0.18	(–0.31 to –0.06)	8.27	758	0.10	<0.001
Time	–0.02	(–0.03 to –0.02)	59.81	274	0.42	<0.001	0.00	(–0.00 to 0.00)	0.12	758	0.01	0.73
Condition	–0.01	(–0.25 to 0.23)	0.01	274	0.01	0.94	0.05	(–0.20 to 0.30)	0.14	758	0.01	0.71
Time × condition	0.00	(–0.01 to 0.02)	0.38	274	0.04	0.54	0.01	(–0.00 to 0.01)	2.06	758	0.05	0.15
Analyses on secondary outcome measures												
Brief Symptom Inventory ^b												
Time	–0.016	(–0.019 to –0.012)	59.66	277	0.42	<0.001	–0.007	(–0.012 to 0.002)	6.51	252	0.16	0.01
Condition	0.013	(–0.230 to 0.257)	0.011	277	0.01	0.92	0.188	(–0.109 to 0.485)	1.55	252	0.08	0.22
Time × condition	0.002	(–0.006 to 0.010)	0.31	277	0.03	0.58	0.009	(–0.002 to 0.019)	2.82	252	0.11	0.09
Work and Social Adjustment Scale ^b												
Time	–0.013	(–0.016 to –0.010)	58.51	279	0.42	<0.001	–0.005	(–0.010 to 0.000)	3.81	254	0.12	0.05
Condition	–0.084	(–0.279 to 0.111)	0.72	279	0.05	0.40	–0.005	(–0.242 to 0.232)	0.002	254	0.00	0.97
Time × condition	–0.000	(–0.007 to 0.006)	0.004	279	0.00	0.95	–0.005	(–0.005 to 0.016)	1.09	254	0.07	0.30
RAND-36 Quality of Life ^b												
Time	0.351	(0.240 to 0.462)	39.13	137	0.47	<0.001	0.063	(–0.077 to 0.203)	0.79	126	0.08	0.38
Condition	–0.621	(–7.320 to 6.077)	0.034	137	0.02	0.86	–0.634	(–7.804 to 6.535)	0.031	130	0.02	0.86
Time × condition	–0.008	(–0.230 to 0.214)	0.005	137	0.01	0.94	–0.029	(–0.309 to 0.252)	0.041	126	0.02	0.84
EuroQol-5D utility score ^b												
Time	–0.019	(–0.024 to –0.014)	56.62	281	0.41	<0.001	–0.003	(–0.011 to 0.005)	0.504	257	0.04	0.48
Condition	–0.003	(–0.332 to 0.327)	0.000	281	0.00	0.99	–0.139	(–0.525 to 0.248)	0.500	257	0.04	0.48
Time × condition	–0.004	(–0.014 to 0.006)	0.564	281	0.04	0.45	–0.004	(–0.019 to 0.012)	0.202	257	0.03	0.65

CI, Confidence interval; FU, follow-up.

^a Mixed models gamma regression with log-link on Beck Depression Inventory – II (BDI-II)+1; Baseline severity is standardized BDI-II score at baseline; Baseline Quality of Life (QoL) is standardized EQ-5D utility score at baseline; Condition is CT *v.* IPT, centred at –0.5 and 0.5, respectively; Time is the linear trend in weeks, with week = 0 at 7 months for the treatment phase and week = 0 at 12 months for the trial FU phase; Time effects in the treatment phase represent change from baseline to 7 months; Time effects in the trial FU phase represent change from 7 to 12 months; Condition main effects represent the difference between conditions at 7 months in the treatment phase and 12 months for the trial FU phase.

^b Mixed model regression on RAND-36 Quality of Life; Mixed models gamma regression with log-link on Brief Symptom Inventory (BSI)+1, Work and Social Adjustment Scale (WSAS)+1; (–1 × EQ-5D utility score+1.01) × 100. Condition is CT *v.* IPT, centred at –0.5 and 0.5, respectively; Time is the linear trend in weeks, with week = 0 at 7 (respectively 12) months; Time effects represent change from baseline to 7 months (respectively 7–12 months), Condition main effects represent the difference between conditions at 7 (respectively 12) months. All models are controlled for standardized baseline BDI-II and EQ-5D scores (except for model in which EuroQol was the dependent variable).

* Effect size $r = \sqrt{F/(F+df)}$; Data unavailable for 3, 17, 23, 23, 23, 25, and 25 patients at 3, 7, 8, 9, 10, 11 and 12 months, respectively.

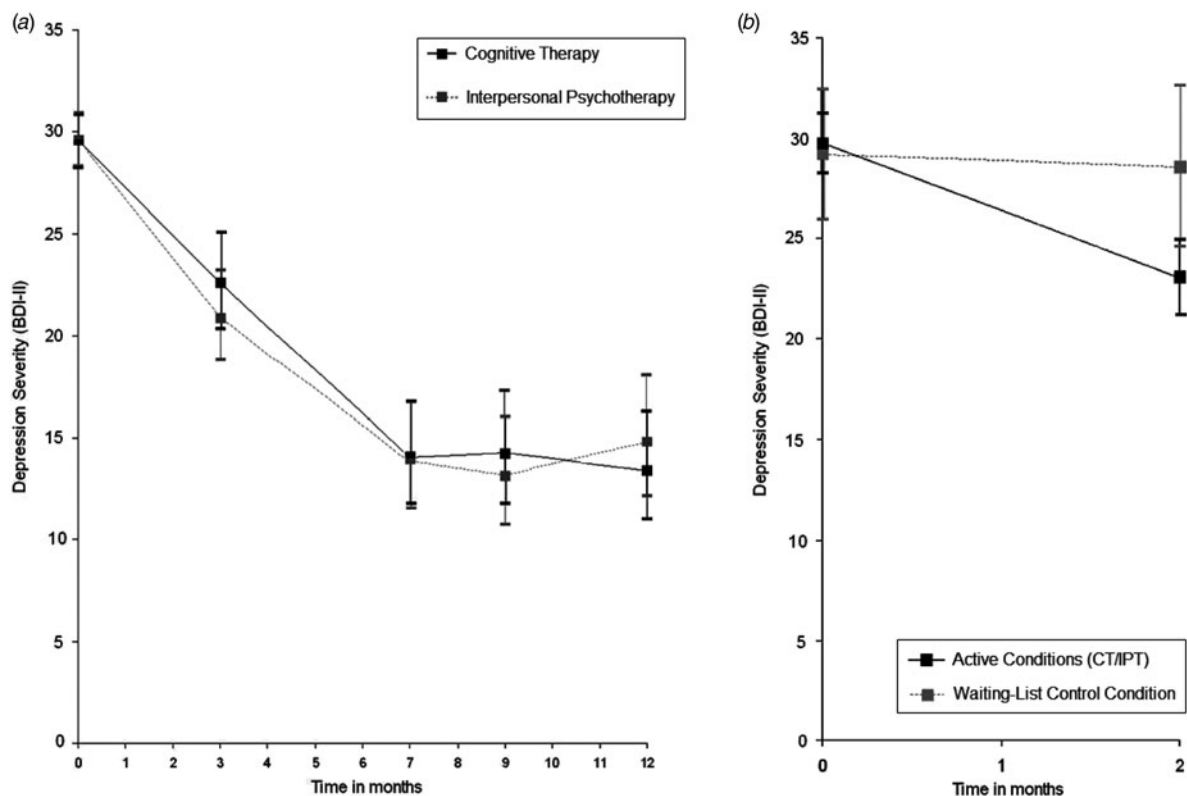


Fig. 2. Course of depression over time. (a) Mixed regression-based estimated means and 95% confidence interval (CI) (corrected for baseline severity and quality-of-life utility score) for the Beck Depression Inventory – II at baseline, 3, 7, 9 and 12 months for cognitive therapy (CT) and interpersonal psychotherapy (IPT). (b) Mixed regression-based estimated means and 95% CI for the Beck Depression Inventory – II at baseline and 2 months for the active conditions (CT/IPT) combined and the waiting-list control condition.

Table 4. Corrected mixed-model estimates of within and between condition changes and Cohen’s *d* based on the BDI-II

	Improvement effect size ^a				Between-group effect size ^b	
	CT		IPT		CT v. IPT	
	Mean change	<i>d</i>	Mean change	<i>d</i>	Mean change difference	<i>d</i>
Baseline – 7 months	–15.52	1.71	–15.66	1.73	–0.14	0.02
Baseline – 12 months	–16.17	1.78	–14.71	1.62	1.46	–0.16

BDI-II, Beck Depression Inventory – II; CT, cognitive therapy; IPT, interpersonal psychotherapy.

^a $(M_{t0} - M_{ti}) / S.D._{t0}$.

^b Difference in improvement effect sizes between two groups; all models corrected for baseline severity and baseline quality of life utility score.

termination. Even though mixed regression analysis yielded lower remission rates at 3 months, results at subsequent assessment points resembled those of the observed values. There were no significant differences between the two intervention groups (all p 's > 0.12). The trend in favour of CT on the observed remission rates at 12 months (42.1% v. 32.0%) disappeared after

controlling for baseline differences in the mixed model analysis.

Relative contribution to the field

Meta-analysis with the random-effects model on non-covariate-corrected post-treatment BDI data from the

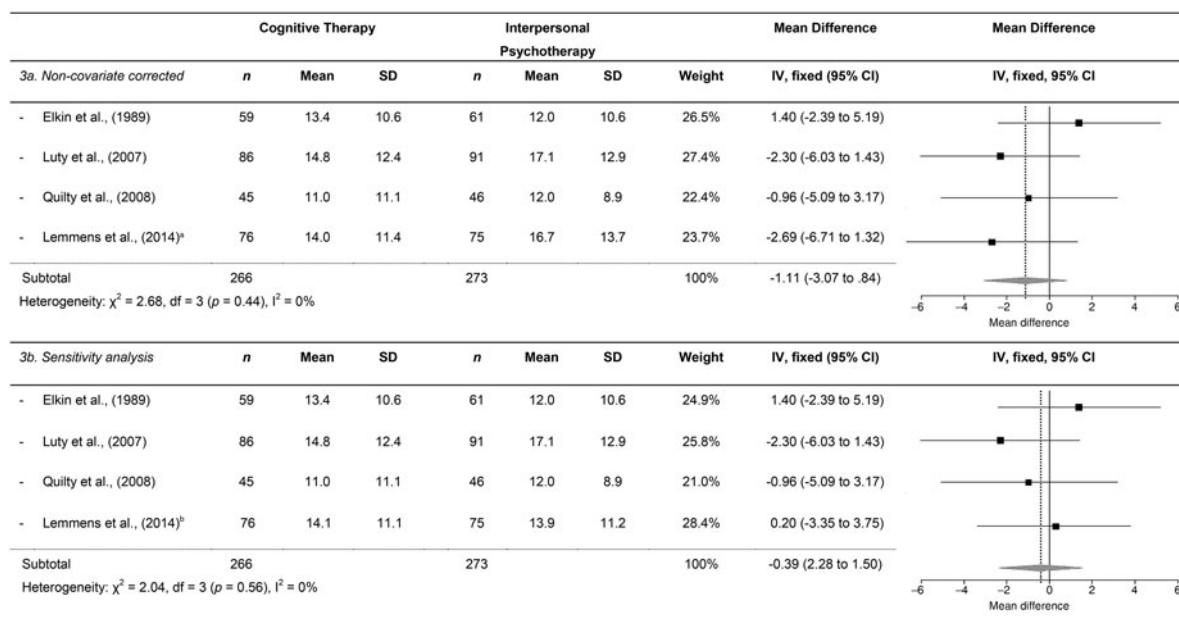


Fig. 3. Post-treatment effect of cognitive therapy *v.* interpersonal psychotherapy on the Beck Depression Inventory [random-effects model using the unstandardized mean difference score with a 95% confidence interval (CI)]. (a) Since existing trials did not adjust their outcome variable according to baseline values, non-covariate-corrected means were included as estimates of the effects of the current study. (b) As a sensitivity analysis, we repeated the analysis with (baseline BDI-II and EQ-5D) corrected post-treatment BDI-II scores. S.D., Standard deviation; IV, inverse variance; df, degrees of freedom. ^a Non-covariate-corrected estimated means; ^b covariate-corrected estimated means.

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

Discussion

Main results

In this study it was found that CT and IPT do not differ in the treatment of depression in the acute phase and beyond. By demonstrating that both treatments exceeded response in the WLC condition and led to considerable improvement in depressive symptoms that was sustained 5 months beyond the end of treatment, our study does not only replicate, but also extends findings of other RCTs that examined individual CT and IPT head-to-head (Elkin *et al.* 1989; Luty *et al.* 2007; Quilty *et al.* 2008). A substantial number of patients (67%) responded to therapy, and the overall improvement in depressive symptoms as assessed with the BDI-II was about 50% in each group. At the end of the 7-month treatment phase, 34% of patients were in remission. Even though this is lower than the rates of response, it is within the range of reported effects in RCT efficacy contexts (Keitner *et al.* 2006). However, it does indicate that the majority of patients experienced residual symptoms at the end of treatment, making them vulnerable for relapse. Pre- to post-treatment effect sizes in our trial were within range of those reported by Elkin *et al.* (1989), Luty *et al.* (2007), and Quilty *et al.* (2008). It should be noted that our

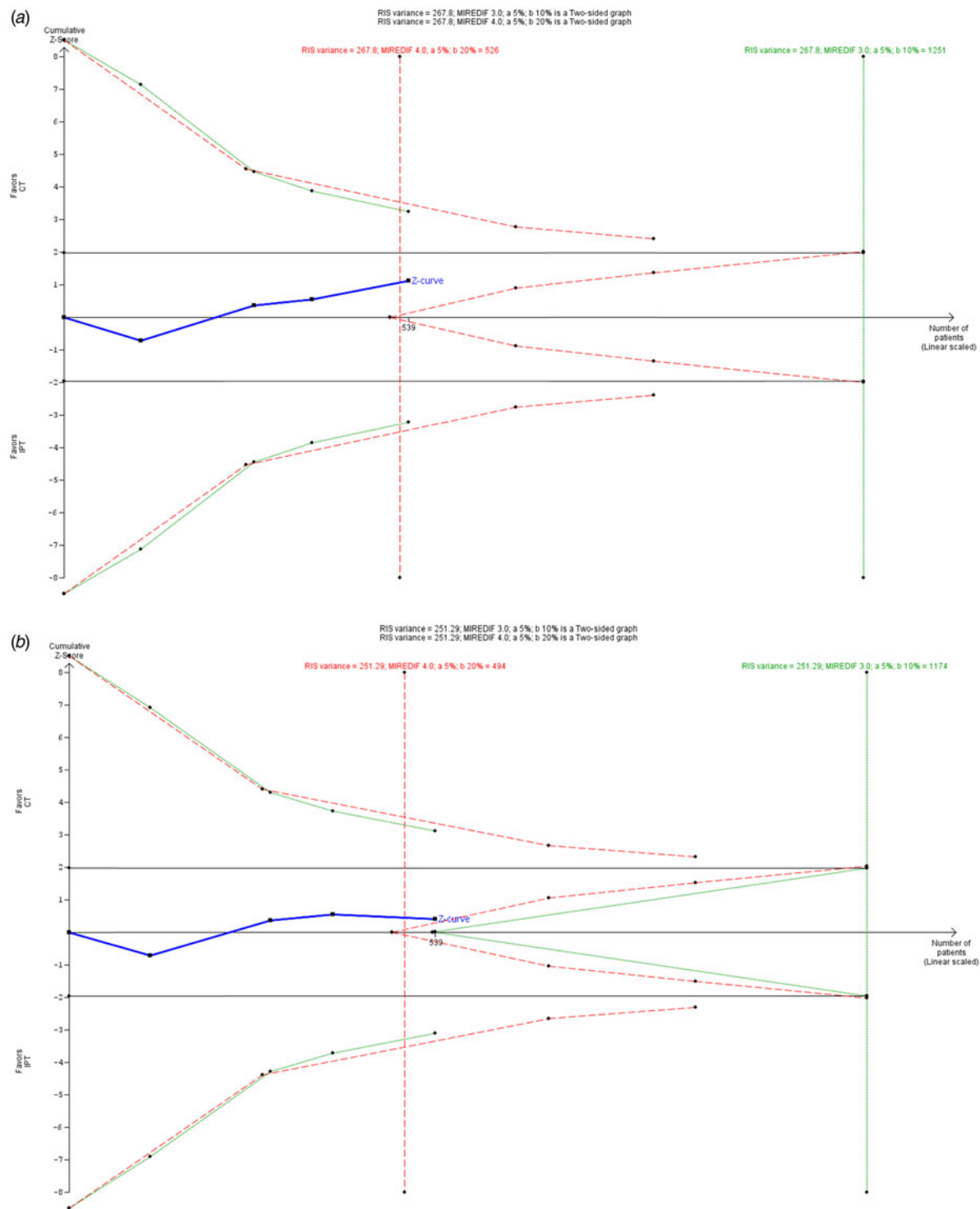


Fig. 4. Trial sequential analysis (TSA) of the cumulative meta-analysis of the effect of cognitive therapy (CT) *v.* interpersonal psychotherapy (IPT) for depression on the Beck Depression Inventory (BDI). (a) TSA using non-covariate-corrected means as estimates of the effects of the current study; the required information size of 526 participants (red/dotted lines) is calculated based on an intervention effect compared with IPT of 4 BDI points, and a power of 80%. With these presumptions, the cumulated Z curve (blue/bold curve) crosses the futility boundary, implying that there are no significant differences in effect between the two interventions and no more trials are needed. The required information size of 1251 (green/etched lines) is calculated based on an intervention effect compared with IPT of 3 BDI points, and a power of 90%. With these presumptions, the cumulated Z curve (blue/bold curve) does not cross the trial sequential monitoring boundaries, implying that there is no firm evidence for a beneficial effect of CT compared with IPT. Both analyses based on a variance of 267.8 and a risk of type I

treatment phase was somewhat longer than in other trials, which gave patients more time to improve. Therapist effects were found to be non-significant, indicating that experienced therapists in a research-oriented routine clinical setting who receive additional training and are carefully monitored throughout, show comparable results. This is in line with previous research by Crits-Christoph *et al.* (1991) that concludes that therapist effects are more likely to be found in outcome studies with minimal training and supervision.

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

Methodological considerations

This study has both strengths and limitations. First, our study is one of the largest clinical trials in the field. With a sample size of $n = 151$ in the active groups, we provided the final information necessary to conclude equivalence of individual CT and IPT for adult depression within a 4 BDI-point limit. Furthermore, it is the first study that replicates research results obtained in the USA and New Zealand in a European sample, and it thereby extends knowledge

about the relative effectiveness of these treatments around the world. In addition, our study was the first to add a WLC condition to a randomized comparison of CT and IPT. The inclusion of an untreated control group diminished uncertainty about whether or not the observed effect was a result of natural course. Given that the WLC condition showed minimal spontaneous remission, the within-condition estimates of the active groups are good estimators of the true effects of CT and IPT. With 85% of patients providing data at the 12-month assessment, attrition rates were low. We therefore are confident that only minimal biases occurred as a result of missing data. By training therapists in a research-oriented routine clinical setting we took care of both the trial's quality as well as generalizability of the trial findings. Other strengths include a broad range of outcomes, the use of modern statistical techniques, intention-to-treat analyses, assessment of bias risk, and an extensive integrity check. However, with regard to the integrity check, it should be noted that even though the overall ICC was acceptable, the range was somewhat less satisfactory, which is a limitation of the present study.

Some other limitations should be mentioned as well. First, as outlined by Cuijpers *et al.* (2010) it is important to include both clinician-rated and self-report measures of depression to assess improvement in depressive symptoms over the course of psychotherapy. Although structured diagnostic interviews ensured accurate classification of patient's diagnoses at baseline, all follow-up measures were based on self-report. We therefore have no information on actual diagnoses at follow-up. Furthermore, inter-rater reliability data on the SCID-I diagnoses is lacking. Third, the duration of the WLC condition was significantly shorter than the treatment time. Therefore a comparison of the full effects of treatment *v.* no treatment was not possible. However, given the study population (patients with major psychopathology who already applied for treatment on their own accord), and the distress and risks related to depression, we considered it unethical to include a full WLC condition. We therefore decided to include a waiting list that was as short as possible. Fourth, because we used a fixed post-treatment assessment and allowed flexibility in scheduling treatment sessions, not all patients finished treatment within the 7-month

error of 5%. (b) TSA using covariate-corrected means (baseline BDI-II and EQ-5D) as estimates of the effects of the current study; the required information size of 494 participants (red/dotted lines) is calculated based on an intervention effect compared with IPT of 4 BDI points, and a power of 80%. With these presumptions, the cumulated Z curve (blue/bold curve) crosses the futility boundary, implying that there are no significant differences in effect between the two interventions and no more trials are needed. The required information size of 1174 (green/etched lines) is calculated based on an intervention effect compared with IPT of 3 BDI points, and a power of 90%. With these presumptions, the cumulated Z curve (blue/bold curve) does not cross the trial sequential monitoring boundaries, implying that there is no firm evidence for a beneficial effect of CT compared with IPT. Both analyses based on a variance of 251.3 and a risk of type I error of 5%.

treatment phase. As a result, the score they provided at the 7-month assessment is a reflection of their progress up until that point, and cannot be considered an absolute post-treatment score. However, because these patients were in the final stage of therapy and only received one or two additional sessions on average, we do not think that this has led to marked changes in outcomes. This was further underlined by the results at 9 months, the point by which all patients had finished treatment, which yielded the same pattern of results. Furthermore, the exclusion of patients receiving concomitant treatment, and those using antidepressant medications in particular, may have reduced generalizability to the whole population of treatment-seeking depressed patients. However, because medication is not a preferred option for initial treatment of episodic depression in The Netherlands (see Spijker *et al.* 2013), only a small group was excluded from the trial for this reason. We therefore think that the negative effects on generalizability for the Dutch situation are relatively small. However, it should be noted that this is probably different for countries in which the use of medication plays a more prominent role in the treatment of MDD. Last, we did not control for multiple testing on the secondary outcomes.

Clinical implications and future research

The present study is the fourth to conclude that no significant differences could be detected between individual CT and IPT in treating depression in the acute phase across the full range of severity. What does this mean for clinical practice? Although it may seem attractive to shift the focus to cost-effectiveness, we think that this would be too early because several crucial aspects for treatment evaluation are still unknown. First of all, existing studies are powered to demonstrate a medium effect between conditions at a modest power level (e.g. 80%). This leaves room for smaller differences between the two approaches that could not be detected. True equivalence trials are necessary to test whether CT and IPT are really equivalent. However, it should be noted that these studies usually require very large sample sizes (depending on what difference is viewed as clinically relevant), and therefore are a challenge in itself. Furthermore, in spite of general equality of treatments, it might be the case that one treatment is superior to the other with regard to certain subgroups of depressed patients. It is therefore important to identify (patient) characteristics that predict differential treatment response. Unfortunately these moderators and predictors of treatment success are still largely unknown. In addition, the fact that therapies that show a distinguishable theory and use specific therapeutic techniques do not necessarily

lead to specific outcomes, calls for specification of their underlying mechanisms. One could argue that different specific pathways lead to similar results (specificity hypothesis). However, it is also possible that change is driven by more common factors such as motivation and therapeutic alliance (non-specificity hypothesis). Process research is needed to settle this question (Kazdin, 2009). Finally, existing studies, including the present, only focus on the short-term effectiveness. It might be the case, however, that even though both treatments are equally effective in the acute phase of treatment, differences will come forward in the long-term with regard to relapse prevention. It is therefore also important to examine long-term effects more closely, which we plan to do in a subsequent paper.

Supplementary material

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0033291715000033>.

Acknowledgements

This research is funded by the research institute of Experimental Psychopathology (EPP), The Netherlands, and the Academic Community Mental Health Centre (RIAGG) in Maastricht, The Netherlands. Both organizations have no special interests in specific outcomes of the trial. We acknowledge the contribution of participating patients and therapists at RIAGG Maastricht. Without their cooperation the trial would have been impossible to conduct. Furthermore, we gratefully thank Annie Raven and Annie Hendriks for their assistance during the study. In addition we also acknowledge Carolien Christ, Kosse Jonker, Meike Krüger and Dina Snippe for their contributions to the treatment integrity check; Heleen van Teeseling for her earlier contributions to the study; and Rosanne Janssen for the development of the infrastructure for online data collection.

Declaration of Interest

None.

Notes

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

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

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

- ⁴ This was done because the study was also designed to examine mechanisms of change. It was therefore important that patients only received one type of treatment at a time.
- ⁵ Further details about power calculations can be found in our design paper (Lemmens *et al.* 2011).
- ⁶ Effects of treatment given after wait in the WLC condition were assessed as well, but are beyond the scope of the current paper and will be presented elsewhere.
- ⁷ Twelve sessions is a commonly used cut-off for the definition of completion in clinical practice and research (see e.g. Beck, 1995; Weissman *et al.* 2000).
- ⁸ The EuroQol-6D assesses six health state dimensions (mobility, self-care, usual activity, pain/discomfort, anxiety/depression, cognitive impairment). The resulting profiles of these dimensions were converted in utilities. However, the norms for calculating utility scores are only available for a five-dimension version of the EuroQol (EuroQol-5D). As a result, the dimension cognition is not taken in consideration in the utility score.
- ⁹ Given his role as editor of the Cochrane Hepato-Biliary Group (CHBG), and expertise in assessment of bias risk, we asked Dr J. Jakobsen (PhD, MD) for his opinion.
- ¹⁰ Calculation based on nine therapists. The therapist that left after treating only two patients was not included in this analysis.
- ¹¹ Even though patients in the IPT condition were a bit faster in scheduling sessions 1–6 (CT = 59 days v. IPT = 53 days) and sessions 1–12 (CT = 133 days v. IPT = 121 days), these differences were not statistically significant ($t_{140} = 1.87$, $p = 0.06$, and $t_{118} = 1.61$, $p = 0.11$ for sessions 6 and 12, respectively). The difference in overall duration of treatment (CT = 178 days v. IPT = 197 days) was also not significant ($t_{147} = -1.51$, $p = 0.13$).
- ¹² Due to the possibility of scheduling sessions less often than weekly, 35% of patients were still in treatment at the end of the acute phase (7 months). The mean number of sessions at the 7-month assessment was 14.5 (s.d. = 3.1) and on average patients had an additional 1.7 (s.d. = 2.6) sessions after that. Since we consider these patients treatment completers in the final stage of therapy at the time of 7-month assessment we included them in the analyses.

References

- Ablon JS, Jones EE (1999). Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology* **67**, 64–75.
- Beck AT, Rush AJ, Shaw BF, Emery G (1979). *Cognitive Therapy of Depression*. Guilford Press: New York.
- Beck AT, Steer R, Brown GK (1996). *Beck Depression Inventory II: Manual*. Harcourt Brace: Boston.
- Beck JS (1995). *Cognitive Therapy: Basics and Beyond*. Guilford Press: New York.
- Bellino S, Zizza M, Rinaldi C, Bogetto F (2007). Combined therapy of major depression with concomitant borderline personality disorder: comparison of interpersonal and cognitive psychotherapy. *Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie* **52**, 718–725.
- Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum: Hillsdale, NJ.
- Crits-Christoph P, Baranackie K, Kurcias J, Beck A, Carroll K, Perry K, Luborsky L, McLellan A, Woody G, Thompson L, Gallagher D, Zitrin C (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research* **1**, 81–91.
- Cuijpers P, Li J, Hofmann SG, Andersson G (2010). Self-reported *versus* clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. *Clinical Psychology Review* **30**, 768–778.
- Cuijpers P, van Straten A, Andersson G, van Oppen P (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* **76**, 909–922.
- Derogatis LR, Melisaratos N (1983). The Brief Symptom Inventory: an introductory report. *Psychological Medicine* **13**, 595–605.
- Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL, Jacobson NS (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.
- Dobson KS, Shaw BF, Vallis TM (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology* **24**, 295–300.
- Dolan P (1997). Modeling valuations for EuroQol health states. *Medical Care* **35**, 1095–1108.
- Driessen E, Cuijpers P, Hollon SD, Dekker JJM (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.
- Drummond MF, Schulpher MJ, Torrance GW, O'Brien B, Stoddart GL (2005). *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edn. Oxford University Press: Oxford.
- Elkin I, Shea MT, Watkins JT, Imber SD (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Archives of General Psychiatry* **46**, 971–982.
- EuroQolGroup (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* **16**, 199–208.
- First MB, Spitzer RL, Gibbon M, Williams JBW (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. Biometrics Research Department New York State Psychiatric Institute: New York.
- Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton MC, Fawcett J (2010). Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Journal of the American Medical Association* **303**, 47–53.
- Hardy GE, Barkham M, Shapiro DA, Stiles WB, Rees A, Reynolds S (1995). Impact of Cluster C personality disorders on outcomes of contrasting brief psychotherapies

- for depression. *Journal of Consulting and Clinical Psychology* **63**, 997–1004.
- Hill CE, O'Grady KE, Elkin I** (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology* **60**, 73–79.
- Hollon SD, Evans MD, Auerbach A, DeRubeis RJ, Elkin I, Lowery A, Kriss M, Grove W, Tuason VB, Piasecki J** (1988). Development of a system for rating therapies for depression: differentiating cognitive therapy, interpersonal psychotherapy, and clinical management pharmacotherapy (unpublished manuscript).
- Hollon SD, Waskow IE, Evans M, Lowery HA** (1984). *Systems for Rating Therapies for Depression*. Annual Convention of the American Psychiatric Association: Los Angeles, CA.
- Jacobson NS, Truax P** (1991). Clinical significance: a statistical approach to define meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* **59**, 12–19.
- Jakobsen JC, Hansen JL, Simonsen S, Simonsen E, Gluud C** (2012). Effects of cognitive therapy *versus* interpersonal psychotherapy in patients with major depressive disorder: a systematic review of randomized clinical trials with meta-analyses and trial sequential analyses. *Psychological Medicine* **42**, 1343–1357.
- Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC** (1999). Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Archives of General Psychiatry* **56**, 431–437.
- Kazdin AE** (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research* **19**, 418–428.
- Keitner GI, Ryan CE, Solomon DA** (2006). Realistic expectations and a disease management model for depressed patients with persistent symptoms. *Journal of Clinical Psychiatry* **67**, 1412–1421.
- Kennedy SH, Eisfeld BS, Cooke RG** (2001). Quality of life: an important dimension in assessing the treatment of depression? *Journal of Psychiatry and Neuroscience* **26**, 23–28.
- Klein DF** (1990). NIMH collaborative research on treatment of depression. *Archives of General Psychiatry* **47**, 682–684.
- Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES** (1984). *Interpersonal Psychotherapy for Depression*. Basis Books: New York.
- Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV** (2006). The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* **15**, 1121–1132.
- Lemmens LHJM, Arntz A, Peeters FPML, Hollon SD, Roefs A, Huibers MJH** (2011). Effectiveness, relapse prevention and mechanisms of change of cognitive therapy *vs.* interpersonal therapy for depression: study protocol for a randomised controlled trial. *Trials* **12**, 150–162.
- Luty SE, Carter JD, McKenzie JM, Rae AM, Frampton CMA, Mulder RT, Joyce PR** (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry* **190**, 496–502.
- Markowitz JC, Weissman MM** (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry* **3**, 136–139.
- Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG** (2010). Research methods & reporting: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *British Medical Journal (Clinical Research Edition)* **340**, c869.
- Mundt JC, Marks IM, Greist JH, Shear K** (2002). The Work and Social Adjustment Scale: a simple accurate measure of impairment in functioning. *British Journal of Psychiatry* **180**, 461–464.
- Quilty LC, McBride C, Bagby RM** (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine* **38**, 1531–1541.
- Spijker J, Bockting CLH, Meeuwissen JAC, van Vliet IM, Emmelkamp PMG, Hermens MLM, van Balkom ALJM** (2013). *Multidisciplinary guideline for depression (third revision): guideline for assessment, treatment and counselling of adults with mood disorders*. Trimbos-instituut: Utrecht.
- Stuart S** (2011). *IPT Adherence and Quality Scale*. Interpersonal Psychotherapy Institute: Iowa (unpublished manuscript).
- Wallace BC, Daharreh IJ, Trikalinos TA, Lau J, Trow P, Schmid CH** (2012). Closing the gap between methodologists and end-users: R as a computational back-end. *Journal of Statistical Software* **49**, 1–15.
- Weissman MM, Markowitz JC, Klerman GL** (2000). *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books: New York.
- Willemse Y, Trijsburg RW** (2005). Cognitive Behavioural Therapy and Interpersonal Psychotherapy: and analysis of critical success factors. *Tijdschrift voor Psychiatrie* **47**, 593–602.
- Zee van der KI, Sanderma R** (1993). *Assessing the general health condition using the RAND-36: a manual*. Northern Center for health questions: Groningen.