Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression

Psychological Medicine

cambridge.org/psm

Long-term outcomes of acute treatment with cognitive therapy *v.* interpersonal psychotherapy for adult depression: follow-up of a randomized controlled trial

Lotte H.J.M. Lemmens[1], Suzanne C. van Bronswijk[2], Frenk Peeters[2], Arnoud Arntz[3], Steven D. Hollon[4] and Marcus J.H. Huibers[5,6]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; [2]Department of Psychiatry and Psychology, Faculty of Health, Medicine and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; [3]Department of Clinical Psychology, University of Amsterdam, PO Box 19268, 1000 GG Amsterdam, The Netherlands; [4]Department of Psychology, Vanderbilt University, 306 Wilson Hall, Nashville, Tennessee, USA; [5]Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands and [6]Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA, 19104-6241, USA

Original Article

Cite this article: Lemmens LHJM, van Bronswijk SC, Peeters F, Arntz A, Hollon SD, Huibers MJH (2018). Long-term outcomes of acute treatment with cognitive therapy *v.* interpersonal psychotherapy for adult depression: follow-up of a randomized controlled trial. *Psychological Medicine* **49**, 465–473. https://doi.org/10.1017/S0033291718001083

Received: 2 August 2017
Revised: 20 January 2018
Accepted: 4 April 2018
First published online: 24 May 2018

Key words:
Cognitive therapy; interpersonal psychotherapy; outcome studies; long-term outcomes; major depression; relapse; sustained response.

Author for correspondence:
Lotte Lemmens, E-mail: Lotte.Lemmens@Maastrichtuniversity.nl

Abstract

Background. Although equally efficacious in the acute phase, it is not known how cognitive therapy (CT) and interpersonal psychotherapy (IPT) for major depressive disorder (MDD) compare in the long run. This study examined the long-term outcomes of CT *v.* IPT for MDD.

Methods. One hundred thirty-four adult (18–65) depressed outpatients who were treated with CT ($n = 69$) or IPT ($n = 65$) in a large open-label randomized controlled trial (parallel group design; computer-generated block randomization) were monitored across a 17-month follow-up phase. Mixed regression was used to determine the course of self-reported depressive symptom severity (Beck Depression Inventory II; BDI-II) after treatment termination, and to test whether CT and IPT differed throughout the follow-up phase. Analyses were conducted for the total sample ($n = 134$) and for the subsample of treatment responders ($n = 85$). Furthermore, for treatment responders, rates of relapse and sustained response were examined for self-reported (BDI-II) and clinician-rated (Longitudinal Interval Follow-up Evaluation; LIFE) depression using Cox regression.

Results. On average, the symptom reduction achieved during the 7-month treatment phase was maintained across follow-up (7–24 months) for CT and IPT, both in the total sample and in the responder sample. Two-thirds (67%) of the treatment responders did not relapse across the follow-up period on the BDI-II. Relapse rates assessed with the LIFE were somewhat lower. No differential effects between conditions were found.

Conclusions. Patients who responded to IPT were no more likely to relapse following treatment termination than patients who responded to CT. Given that CT appears to have a prophylactic effect following successful treatment, our findings suggest that IPT might have a prophylactic effect as well.

Introduction

Cognitive therapy (CT) and interpersonal psychotherapy (IPT), the two best studied and commonly practiced psychological interventions for the treatment of major depressive disorder (MDD), have shown to be effective treatments for many depressed patients (Cuijpers *et al.* 2011; Barth *et al.* 2013; Cuijpers *et al.* 2013*a*; Cuijpers *et al.* 2016). With initial response rates up to 60%, they have shown to be at least as efficacious as antidepressant medications (ADM) in the acute phase of the disorder (Cuijpers *et al.* 2013*b*, 2013*c*). However, even when treated effectively in the acute phase[†1], depression has an unfavorable prognosis. It is estimated that at least 50% of those who recover from a first episode of MDD will have one or more additional episodes later on in life, and the risk of recurrence progressively increases with each additional episode (Solomon *et al.* 2000; Burcusa and Iacono, 2007; Eaton *et al.* 2008). It is therefore important that treatments do not only reduce symptoms in the acute phase, but also produce enduring effects.

CT has consistently been shown to have an enduring effect that lasts beyond the end of treatment, with survival rates higher than those associated with (prior) pharmacological treatment[2] (Vittengl *et al.* 2007; Cuijpers *et al.* 2013*b*). Research in IPT is less extensive. Even though IPT has shown to prevent relapse and recurrence when continued or maintained (Cuijpers *et al.* 2016), only one older study has examined whether it has a prophylactic effect

[†]The notes appear after the main text.

following treatment termination (Shea *et al.* 1992). This was the follow-up to the NIMH Treatment of Depression Collaborative Research Program (TDCRP), a placebo-controlled randomized comparison among CT, IPT, and ADM that found comparable rates of relapse between prior IPT and prior CT (33% *v.* 36%) that were each non-significantly lower than prior ADM (50%). These findings must be interpreted with caution, since sample sizes were small, ADM was continued for 6 months following the end of acute treatment, and the difference between prior CT and prior ADM was among the smallest reported in the literature, but they are suggestive of a possible enduring effect for prior IPT. Additional research into the extent to which the effects of IPT persist following the cessation of treatment is needed.

Recently, we conducted a large randomized controlled trial (RCT) investigating the effects of individual CT and IPT for adult depression, primarily designed to compare long-term outcomes of both therapies in a research-oriented routine clinical setting (Lemmens *et al.* 2011; 2015; 2017). CT and IPT were both superior to a waiting-list control (WLC) condition over the first 2 months of treatment, and did not differ from another across the rest of the 7-month treatment phase (Lemmens *et al.* 2015) – as has been the case in most acute phase comparisons between the two modalities (Jakobsen *et al.* 2012). We now report on the long-term outcomes of these two interventions over the next 17 months, through the end of 24 months post-randomization. We expected depression scores to be relatively stable across the follow-up period. Furthermore, we expected relapse rates in CT to be similar to those reported in the previous studies; approximately 30%. Following earlier findings (Shea *et al.* 1992), one would not expect large differences between CT and IPT. However, since CT has a stronger tradition in focusing on relapse prevention compared with IPT, we expected that CT would do somewhat better.

Methods

Design and participants

Data come from a single-center RCT (parallel group design) into the clinical effects and mechanisms of change of individual CT and IPT for MDD. In this study, 182 depressed adults were randomly allocated to CT ($n = 76$), IPT ($n = 75$), or a 2-month WLC condition followed by treatment of choice ($n = 31$). In the present study, we only included the patients who were assigned to one of the two active conditions (CT and IPT) and who provided data at post-treatment (month 7; $n = 134$; CT = 69, IPT = 65; henceforth *total sample*).

Details concerning study design, participants, interventions, and acute outcomes have been fully described elsewhere (Lemmens *et al.* 2011; 2015), and will therefore only briefly be summarized here. Participants were adult outpatients referred to the mood disorder treatment program of the Academic Community Mental Health Centre Maastricht. All patients had a primary diagnosis of MDD, as ascertained by the Dutch version of the Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I; First *et al.* 1997). Further inclusion criteria were: internet access, an e-mail address, and sufficient knowledge of the Dutch language. Patients receiving ADM or other psychological treatment at baseline were excluded from the study, as were those at imminent risk for suicide. Other exclusion criteria were: bipolar or chronic depression (current episode >5 years), IQ lower than 80, and substance abuse/dependence.

All participants provided informed consent. Randomization took place via computer-generated block randomization (10:10:4) and was pre-stratified according to the presence or absence of prior episodes. The random allocation sequence was generated by an independent computer scientist and was concealed from the researchers. Blinding of patients and therapists for treatment condition was not possible. As outlined elsewhere (Lemmens *et al.* 2011), sample size calculations were based on long-term expectations of CT *v.* IPT. An *a priori* power analysis indicated that 75 patients per arm in the active conditions (taking 15% attrition into account) would provide 80% power (two-tailed $\alpha = 0.05$) to detect an expected 20% difference in the relapse rate between CT and IPT (favoring CT) at the end of the follow-up period.

Treatment consisted of 16–20 individual 45-min sessions. The CT protocol followed the guidelines laid out by Beck *et al.* (1979) and included homework assignments. The IPT protocol was based on the manual by Klerman *et al.* (1984). Therapists were uniquely assigned to one of the treatment conditions to prevent contamination. All therapists had several years of clinical experience in the field of depression and with the assigned intervention. Prior to the study, all therapists received 16 hours of additional training by experts in the field. During the study, therapists and researchers met biweekly in consultation sessions to discuss their caseloads (separate sessions for CT and IPT). The study was approved by the Maastricht University's Ethical Board, and is registered at the Netherlands Trial Register, part of the Dutch Cochrane Centre (ISRCTN67561918). Patients completed an average of 17 therapy sessions (S.D. = 2.9). Independent assessors rated the quality of therapy, measured with the Cognitive Therapy Scale (Dobson *et al.* 1985) for CT and the Short version of the IPT Adherence and Quality Scale (Stuart, 2011) for IPT, as being '(very) good' to 'excellent' in both conditions (see Lemmens *et al.* 2015 for more details). Furthermore, significant differences in therapy-specific behavior between conditions were found (as measured with the Collaborative Study Psychotherapy Rating Scale – version 6; Hollon *et al.* 1984; Hollon *et al.* 1988), indicating that therapists adhered to the protocol.

Both treatments led to considerable improvement in depressive symptom severity as measured with the Beck Depression Inventory II (BDI-II; Beck *et al.* 1996: pre–post-treatment effect size $d = 1.72$ in the pooled active conditions). Response to the therapy exceeded response in the WLC condition. No differential effects between the active treatments were found (Lemmens *et al.* 2015).

Outcomes

Self-reported depression severity

Self-reported depressive symptom severity was measured with the BDI-II. The BDI-II is a 21-item questionnaire with strong psychometric properties (Beck *et al.* 1996; Van der Does, 2002). Items are rated on a four-point Likert scale (0–3), with higher scores indicating higher levels of depression severity (range 0–63).

Clinician-rated depression severity

The MDD section of the Longitudinal Interval Follow-up Evaluation (LIFE; Keller *et al.* 1987), a semi-structured interview for assessing the longitudinal course of psychiatric illness using a retrospective rating system, was used to obtain a clinician-rated measure of depression. The LIFE uses DSM-IV diagnostic criteria to classify depression retrospectively over the course of a pre-

https://doi.org/10.1017/S0033291718001083

determined follow-up period (in our case 1 year; see further). Ratings are made on a six-point scale, ranging from meeting the full criteria of MDD (ratings of 5 or 6) to no residual symptoms (rating of 1). The LIFE has shown to be a reliable and valid instrument for identifying the course of several mental disorders examined retrospectively over the period of 1 year (Warshaw *et al.* 1994; Warshaw *et al.* 2001).

Procedure

BDI-II assessments were completed at post-treatment (month 7), monthly thereafter for the next 5 months (month 7–12), and then again at the end of the follow-up (month 24). All assessments were administred on a computer. The post-treatment assessment took place at the research center (Maastricht University). All other assessments were administred online. The LIFE interview took place after the 24-month assessment and addressed retrospectively the period between 12 and 24 months[3]. A rating was made for each 2-week period between months 12 and 24, resulting in a total of 26 retrospective observations. A schematic overview of the study design and the data points used in this study can be found in Fig. 1.

The majority of LIFE interviews (90%) were administered by a clinical psychology graduate student. The remaining 10% was administred by a resident in psychiatry. Both LIFE assessors had several years of clinical experience in the field of depression. Prior to the study, assessors studied relevant literature, the original set of LIFE training materials, and the detailed instruction manual that was developed for the current study. Furthermore, they conducted several pilot interviews to familiarize themselves with the rating system. During the study, regular consensus sessions took place, in order to discuss interpretation and pitfalls. The interviewers were blind to condition, treatment-adherence, -satisfaction, and -outcome. Interviews took place face-to-face or by telephone[4] and ratings were made after the interview. Ratings made by the psychiatric resident were discussed with the other rater until consensus was reached.

Since patients were free to pursue additional treatment during the follow-up phase, we assessed whether patients received additional psychological support for MDD (conservatively defined as having one or more sessions with a general practitioner (GP) or a mental health care professional for depressive symptoms) or used ADM (⩾2 weeks) throughout the 17-month follow-up period. Information on health care status was obtained during the LIFE interview and at the 12 and 24 months assessment[5] with the periodic retrospect health care consumption questionnaire (de Graaf *et al.* 2008).

Definition of response and relapse

When investigating the clinical course of a disorder after acute phase treatment, one needs to carefully consider the definitions of response and relapse. Response is often conceptualized as a pre-determined change score representing a clinical significant improvement over the course of treatment (Jacobson and Truax, 1991). Even though this method is useful in the majority of cases, in some cases, it leads to somewhat peculiar classifications. For example, the approach excludes patients who reach remission without the necessary drop in symptoms as treatment responders (e.g. a drop from 14 to 8 on the BDI-II). Furthermore, it includes patients who do show a clinical significant improvement, but still report high depression scores at the end of treatment (e.g. a drop from 61 to 48), hereby indicating that treatment had some effect, but worked insufficiently to reach (partial) remission. In order to take these variations into account, we defined response to treatment as *either* (1) a post-treatment BDI-II score lower than 10 (the cut-off for remission in our trial determined with the method of Jacobson and Truax, 1991; see Lemmens *et al.* 2015); *or* (2) an overall change of at least 9 BDI-II points (the cut-off for reliable change in our trial determined with the method of Jacobson and Truax, 1991; see Lemmens *et al.* 2015) *and* a post-treatment BDI-II score lower than 20 (the border between moderate and mild depression on the BDI-II, see[6]). Since the BDI-II and the LIFE assess different aspects of depression (depressive symptom severity *v.* DSM-IV classification), we formulated two separate definitions for relapse. Relapse on the BDI-II was defined as losing ⩾50% of the improvement that occurred over the course of treatment at any point during the follow-up (7, 8, 9, 10, 11, 12, 24 months). This was done to account for individual symptom severity change. Following Hollon *et al.* (2005), relapse on the LIFE occurred as soon as the patients met full DSM-IV criteria for a depressive episode (rating of 5 or 6) on one of the 26 retrospective data points. Definitions of response and relapse are summarized in Table 1.

Data analysis

First, for all patients ($n = 134$), we mapped out study compliance (LIFE and BDI-II) across the follow-up period (7–24 months), and compared patients with and without complete data on each of the outcome measures in terms of baseline characteristics (gender, age, education level, work- and marital status, first/recurrent depression) and post-treatment BDI-II score. We used χ^2 tests for categorical data and independent samples *t* tests for continuous data. In addition, for each of the outcome measures (BDI-II and LIFE), BDI-II scores of patients with incomplete data were

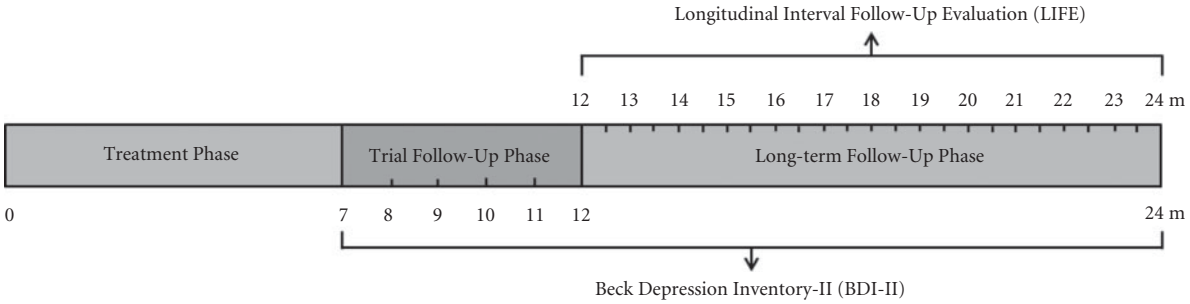

Fig. 1. Overview of BDI-II and LIFE data points used in the current study.

https://doi.org/10.1017/S0033291718001083

Table 1. Overview of definitions of Response, Relapse and Sustained Response on the BDI-II and LIFE

Beck Depression Inventory-II (BDI-II)	
Response	Post-treatment BDI-II score <10; *or* improvement of ⩾9 BDI-II points during treatment *and* post-treatment BDI-II score <20
Relapse	Losing ⩾50% of the improvement that occurred during treatment at any point during follow-up
Sustained response	Response to treatment (see above) *and* no relapse (see above) during follow-up
Longitudinal Interval Follow-Up Evaluation (LIFE)	
Relapse	Meeting full DSM-IV criteria for MDD (rating of 5 or 6) on one of the 26 retrospective data points

BDI-II, Beck Depression Inventory-II; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; MDD, Major Depressive Disorder

plotted to explore potential patterns in depression severity prior to drop-out. For the BDI-II, reliability at each time point was assessed using Cronbach's α.

After that, we determined the course of self-reported depressive symptom severity after treatment termination, and tested whether one of the treatments was superior to the other across the follow-up period. For this, we used a linear mixed-effects (multilevel) model with repeated BDI-II scores as the dependent variable, and time, condition (CT = −0.5, IPT = 0.5), and the time × condition interaction as the independent variables (Diggle *et al.* 2002). Because mixed regression takes the nested structure of the data into consideration and can deal with autocorrelation and missing values (see e.g. Schafer and Graham, 2002; Singer and Willett, 2003; Snijders and Bosker, 2012), missing values were not imputed. Since CT and IPT differed in depression severity (BDI-II) and quality of life (EQ5D utility score; EuroQol-Group, 1990) at baseline, albeit not significant (see Lemmens *et al.* 2015), we added their standardized baseline scores as covariates to the model. Visual inspection of BDI-II change scores over time showed separate linear patterns for the 7–12 and 12–24 months intervals. Therefore, for the fixed effects, the slopes were modeled separately for each interval (piecewise regression; see online Supplementary material I). An autoregressive covariance structure was applied to factor in the correlation between measurement points. Intercepts and slopes (for the time variable) were allowed to be correlated and to vary randomly over subjects. Robust standard errors were applied. Effect sizes Cohen's d and r were computed from the multilevel estimates. Within-condition change was defined as Cohen's d = (post-treatment mean−mean at time i)/(pooled post-treatment s.d.), with the estimated means derived from the mixed regression analysis. Between-group effect sizes were determined by calculating the difference between the within-condition effect sizes of CT and IPT at time i. The r was defined as $\sqrt{(F/(F+\mathrm{df})}$, with F and df values from the fixed part of the mixed regression analysis. Effect sizes were classified as being small (±0.2), medium (±0.5), and large (±0.8; Cohen, 1988).

Subsequently, we calculated response rates at 7 months and compared pre-treatment characteristics (similar to those described above) and post-treatment BDI-II scores between responders and non-responders. After that, we continued our analyses with the responder sample only (responder analysis; $n = 85$). First, we re-ran the previously described linear mixed-effects model and computed effect sizes r and d. Second, relapse rates (separate analyses for the BDI-II and LIFE) were examined using Cox regression models with condition as an independent variable and standardized baseline BDI-II and EQ5D utility scores as covariates (Cox and Oakes, 1984). The proportional hazard assumption was tested with Schoenfeld Residuals Test. Between-condition survival rates were compared using the log-rank test. Drop-outs were censored after the last measurement. In addition, for the BDI-II, we examined the rates of *sustained response*: the number of patients who responded to treatment and remained well (no relapse) during the follow-up period (see Table 1). Following Hollon *et al.* (2005), percentages of sustained response rates were adjusted for missing observations and calculated from the '*baseline*' sample (all patients that were initially assigned to CT or IPT, regardless of their enrollment in the current study; $n = 151$). By doing this, percentages of sustained response reflect pre-treatment probabilities of enduring treatment effects. CT/IPT differences in sustained response were examined using a χ^2 test of independence.

Finally, sensitivity analyses were performed on all models by adding the following variables sequentially as (centered) covariates to each model: gender, work- and marital status, number of sessions received in acute phase, therapist, and additional psychological support for MDD, and/or use of ADM in the follow-up period. Multilevel analyses were carried out in SPSS (version 21). Other analyses were performed in STATA (version 13.1). All effects were tested at the $p < 0.05$ level (two-tailed).

Results

Patient flow and attrition

Of the 134 patients that were enrolled in the current study, 119 (88.8%) completed all BDI-II assessments[7]. Baseline characteristics of patients with incomplete data on the BDI-II [eight in CT and seven in IPT: χ^2 (1, 134) = 0.02, $p = 0.88$] were not significantly different from those with complete data (all $p > 0.23$). In addition, exploration of the course of depression indicated no distinctive patterns of findings between patients with incomplete and complete BDI-II data. The depression severity pattern for patients with missing BDI-II data appeared to be random. A total of 98 LIFE interviews were administered. Among the 36 patients whose LIFE interview was missing, six dropped out in an earlier phase of the study, 14 were unattainable and did not respond to contact requests, and 16 indicated that they no longer wanted to participate. Furthermore, LIFE data of two patients were incomplete. The conditions did not differ with respect to whether there were (complete) LIFE data [20 in CT *v.* 18 in IPT: χ^2 (1, 134) = 0.03, $p = 0.87$]. Relative to those with complete LIFE data, patients without (complete) LIFE data were significantly more likely to have an intermediate (vocational) level of education [χ^2 (2, 134) = 8.62, $p = 0.01$]. Furthermore, they reported higher post-treatment BDI-II scores [$M = 18.7$ (s.d. = 14.7) *v.* $M = 13.3$ (s.d. = 10.6); t (132) = 2.33, $p = 0.02$]. In addition, patients with incomplete LIFE data had somewhat higher BDI-II scores at all time points as compared with patients with

https://doi.org/10.1017/S0033291718001083

complete LIFE data, indicating a more severe pattern of depression in general. BDI-II reliability coefficients ranged from $\alpha = 0.96$ (at 7 and 8 months) to $\alpha = 0.97$ (12 months). A total of 54 patients (40.3%; CT = 33, IPT = 21) had one or more sessions with their GP or a mental health care professional during the follow-up period[8]. Twenty-nine patients (21.6%; CT = 14, IPT = 15) used ADM. As some patients received both psychological and pharmacological support, the total number of patients with some form of additional support was 63 [47.0%; CT = 36, IPT = 27; χ^2 (1, 128) = 2.01, $p = 0.16$].

Course of depressive symptomatology in the total sample

Table 2 presents the observed mean BDI-II scores (95% CI) and mixed regression-based estimated means (95% CI) over the course of follow-up for the total sample ($n = 134$), stratified according to the treatment condition. In addition, the results of the mixed-effects model and effect sizes r and d are reported. As can be seen in Table 2, symptom scores remained stable across the follow-up for both conditions. Effect sizes were small.

Responder analyses

At post-treatment (7 months), 85 patients (63.4%) met criteria for response. No between-condition differences were found [65.2% in CT *v.* 61.5% in IPT; χ^2 (1, 134) = 0.20, $p = 0.66$]. Responders did not differ from non-responders at baseline, but reported significantly lower BDI-II scores at post-treatment [$M = 7.5$ (S.D. = 6.0) *v.* $M = 27.6$ (S.D. = 9.0); t (132) = 15.48, $p < 0.0001$], and fewer received additional support for MDD throughout the follow-up [26/85 responders *v.* 37/49 non-responders; χ^2 (1, 128) = 25.87, $p < 0.0001$].

Outcomes on the BDI-II

The linear mixed-effects model on the BDI-II for the responder sample (Table 3) revealed significant time × condition and time × year × condition interactions. The time × condition interaction points toward more favorable outcomes for CT up to month 12 (BDI-II scores showed a slight decrease in CT, whereas they increased in IPT), and the negative time × year × condition interaction reflects a subsequent drop in scores for IPT in the second year. At 17 months follow-up, these opposite effects resulted in comparable overall outcomes for CT and IPT. Effect sizes of change throughout the follow-up period (7–24 months) were small for both CT and IPT. Two-thirds of treatment responders (57/85; 67.1%) completed the 17-month follow-up phase without meeting criteria for relapse on the BDI-II. Cumulative survival rates per treatment condition are shown in Figure 2*a*. Relapse rates were 13 (28.9%) for CT and 15 (37.5%) for IPT. A log-rank test [χ^2 (1, 85) = 0.99, $p = 0.32$] and a Cox regression model (HR

Table 2. Results of the linear mixed-effects model on the BDI-II for the total sample ($n = 134$) + observed and mixed regression-based estimated mean BDI-II scores (95% CI) and (for the estimated means) effect size d over the course of follow-up (7–24 months), stratified by condition

BDI-II	Observed means		Estimated means				
	CT ($n = 69$)	IPT ($n = 65$)	CT ($n = 69$)		IPT ($n = 65$)		
7 months	13.8 (11.2–16.3)	16.0 (12.7–19.3)	14.7 (12.4–17.0)	*Change d**	15.1 (12.4–17.7)	*Change d**	*CT-IPT d**
8 months	13.6 (10.9–16.3)	16.6 (13.1–20.2)	14.6 (12.3–16.9)	0.01	15.2 (12.6–17.9)	−0.01	0.02
9 months	13.3 (10.5–16.2)	15.8 (12.2–19.4)	14.4 (12.1–16.7)	0.03	15.4 (12.8–18.1)	−0.03	0.06
10 months	14.3 (11.4–17.2)	16.8 (13.1–20.5)	14.3 (11.9–16.7)	0.04	15.6 (12.8–18.4)	−0.05	0.09
11 months	13.2 (10.4–16.0)	16.7 (13.0–20.4)	14.1 (11.7–16.6)	0.06	15.8 (12.9–18.7)	−0.07	0.13
12 months	12.7 (9.7–15.6)	17.2 (13.3–21.0)	14.0 (11.4–16.6)	0.07	16.0 (12.9–19.1)	−0.09	0.14
24 months	11.7 (8.9–14.5)	14.9 (11.2–18.5)	12.5 (9.7–15.2)	0.22	13.4 (10.2–16.6)	0.17	0.05

Linear mixed-effects model									
	Fixed effects						Random effects		
Variable	β	Robust SE	F	df	r**	p	Var.	S.E.	CI_{95}
Intercept	14.88	0.89	6.65	884	0.09	<0.001	111.22	15.73	84.28–146.75
Baseline severity	3.34	0.82	16.45	884	0.14	<0.001			
Baseline quality of life	−2.89	1.09	7.080	884	0.09	0.01			
Time	0.02	0.13	0.022	884	0.001	0.88	0.37	0.07	0.26–0.53
Condition	0.33	1.79	0.035	884	0.01	0.85			
Time × condition	0.33	0.25	1.712	884	0.04	0.19			
Time × year	−0.19	0.17	1.277	884	0.04	0.26			
Time × year × condition	−0.42	0.33	1.576	884	0.04	0.21			

BDI-II, Beck Depression Inventory-II; CT, cognitive therapy; IPT, interpersonal psychotherapy; 95% CI, 95% confidence interval; baseline severity is standardized BDI-II score at baseline; baseline quality of life is standardized EuroQol (EQ5D) Utility Score at baseline; condition is CT *v.* IPT centered at −0.5 and 0.5, respectively; time effects represent the linear trend from 7 to 24 months, with week = 0 at 7 months; year is first *v.* second year of the study, coded as 0 < 12 and 1 ⩾ 12 months; data unavailable for 7 (4 CT; 3 IPT), 6 (4 CT; 2 IPT), 6 (3 CT; 3 IPT), 8 (4 CT; 4 IPT), 8 (4 CT; 4 IPT) and 11 (7 CT; 4 IPT) patients at 8, 9, 10, 11, 12, and 24 months, respectively; * = $(M_{t7}-M_{ti})/SD_{t7}^{pooled}$; ** = effect size $r = \sqrt{(F/(F + df))}$.

https://doi.org/10.1017/S0033291718001083

Table 3. Results of the linear mixed-effects model on the BDI-II (in sample predictions) for the responder sample (n = 85) + observed and mixed regression-based estimated mean BDI-II scores (95% CI) and (for the estimated means) effect size *d* over the course of follow-up (7–24 months), stratified by condition

BDI-II	Observed means		Estimated means				
	CT (*n* = 45)	IPT (*n* = 40)	CT (*n* = 45)		IPT (*n* = 40)		
7 months	7.7 (5.8–9.6)	7.3 (5.5–9.2)	8.1 (6.4–9.9)	*Change d**	7.8 (5.9–9.8)	*Change d**	*CT-IPT d**
8 months	7.9 (5.7–10.2)	9.4 (6.4–12.5)	8.0 (6.3–9.7)	0.02	8.4 (6.4–10.3)	−0.10	0.12
9 months	7.7 (5.3–10.1)	9.4 (5.9–12.9)	7.8 (6.1–9.5)	0.05	8.9 (6.7–11.0)	−0.17	0.22
10 months	7.9 (5.5–10.3)	9.8 (6.4–13.2)	7.6 (5.9–9.3)	0.08	9.4 (7.1–11.7)	−0.25	0.33
11 months	7.4 (5.3–9.4)	9.6 (6.7–12.5)	7.4 (5.6–9.2)	0.11	9.9 (7.3–12.4)	−0.33	0.44
12 months	6.7 (4.5–8.9)	10.3 (7.2–13.4)	7.2 (5.2–9.2)	0.14	10.4 (7.6–13.2)	−0.41	0.55
24 months	8.9 (6.1–11.6)	8.6 (5.0–12.2)	8.9 (6.2–11.7)	−0.13	8.3 (4.8–11.8)	−0.08	−0.05

Linear mixed-effects model

	Fixed effects						Random effects		
Variable	*β*	Robust SE	*F*	df	*r***	*p*	Var.	S.E.	CI_{95}
Intercept	7.99	0.66	4.08	555	0.09	<0.001	39.69	8.33	26.3–59.9
Baseline severity	2.15	0.62	12.10	555	0.15	0.001			
Baseline quality of life	−0.16	0.71	0.05	555	0.01	0.82			
Time	0.16	0.13	1.62	555	0.05	0.20	0.34	0.08	0.22–0.53
Condition	−0.31	1.34	0.05	555	0.01	0.82			
Time × condition	0.70	0.25	7.56	555	0.12	0.006			
Time × year	−0.18	0.17	1.06	555	0.04	0.31			
Time × year × condition	−1.01	0.34	8.63	555	0.12	0.003			

BDI-II, Beck Depression Inventory-II; CT, cognitive therapy; IPT, interpersonal psychotherapy; 95% CI, 95% confidence interval; baseline severity is standardized BDI-II score at baseline; baseline quality of life is standardized EuroQol (EQ5D) Utility Score at baseline; condition is CT *v.* IPT centered at −0.5 and 0.5, respectively; time effects represent the linear trend from 7 to 24 months, with week = 0 at 7 months; year is first *v.* second year of the study, coded as 0 < 12 and 1 ⩾ 12 months; data unavailable for 6 (3 CT; 3 IPT), 4 (2 CT; 2 IPT), 4 (2 CT; 2 IPT), 5 (2 CT; 3 IPT), 4 (2 CT; 2 IPT) and 9 (5 CT; 4 IPT) patients at 8, 9, 10, 11,12, and 24 months, respectively; * = $(M_{t7}-M_{ti})/SD_{t7}^{pooled}$; ** = effect size $r = \sqrt{(F/(F+df))}$.

= 1.47, S.E. = 0.57, $p = 0.32$, $CI_{95} = 0.69–3.14$) indicated that there were no significant differences in relapse rates between CT and IPT. For those who did relapse ($n = 28$), the mean time to relapse was 61.4 weeks (S.E. = 5.7) after baseline. On average, patients in the IPT condition relapsed somewhat faster than those treated with CT [mean time to relapse of 54.1 (S.E. = 6.9) for IPT *v.* 69.8 weeks (S.E. = 9.2) for CT]. Patients were slightly more likely to show sustained response[9] in CT (32 of 45 = 42.1%) than in IPT (25 of 40 = 33.3%), but differences were not significant [χ^2 (1, 151) = 1.24, $p = 0.27$].

Outcomes on the LIFE

LIFE data were available for 65 of 85 responders (76.5%; CT = 33, IPT = 32); 55 completed the follow-up phase without meeting criteria for relapse. Figure 2*b* shows – separately for CT and IPT – the cumulative proportion of treatment responders without relapse on the LIFE. Survival rates of CT and IPT were not significantly different [log-rank test: χ^2 (1, 65) = 0.43, $p = 0.51$; Cox regression analysis: HR = 1.57, S.E. = 1.03, $p = 0.48$, $CI_{95} = 0.44–5.66$]. For the patients that did show relapse on the LIFE ($n =$ 10; four in CT and six in IPT), mean time to relapse was 75.4 weeks (S.E. = 4.3) after baseline [67.5 (S.E. = 5.4) and 80.7 (S.E. = 5.5) weeks in CT and IPT, respectively]. For all models, the proportional hazard assumption was not violated.

Sensitivity analyses

Sensitivity analyses indicated that the total number of sessions, therapist, work- and marital status did not influence the findings. None of these covariates were significant in any of the models. Gender was a significant covariate in the survival analysis on the LIFE, but did not change the conclusions. Additional psychological and/or pharmacological support for MDD also did not change the conclusions. However, these variables were significant in some of the models as well: in both multilevel models and in the survival analysis on LIFE. Results indicated that patients who received psychological support for MDD and/or used ADM during the follow-up reported higher BDI-II scores at 24 months and were more likely to meet criteria for relapse on the LIFE than those without additional support. In the majority of cases, additional support was requested after relapse occurred ($n = 6$ *v.* $n = 2$ for support before relapse).

Discussion

The current study evaluated the long-term outcomes of acute phase CT *v.* IPT for MDD. In the context of a large RCT, we determined the course of self-reported depressive symptom severity up to 17 months after treatment termination, and tested whether CT and IPT differed throughout the follow-up phase.

https://doi.org/10.1017/S0033291718001083

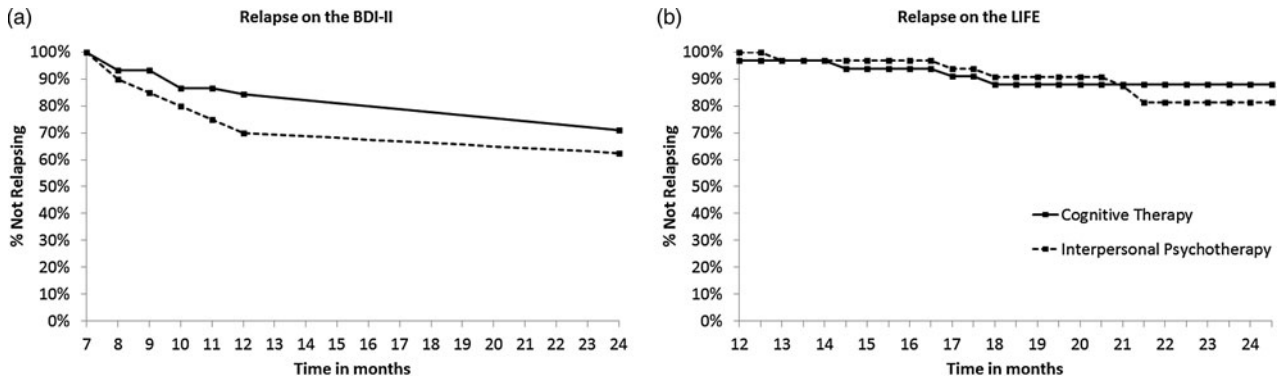

Fig. 2. Cumulative proportion of treatment responders surviving without relapse over the course of follow-up. Separate panels for relapse on the Beck Depression Inventory-II (BDI-II (*a*), $n = 85$) and the Longitudinal Interval Follow-Up Evaluation (LIFE (*b*), $n = 65$). Results stratified per condition. Note: *time in months* starts the month after baseline (baseline = month 0).

Furthermore, for treatment responders, rates of relapse and sustained response were examined for self-reported (BDI-II) and clinician-rated (LIFE) depression. On average, the symptom reduction achieved during the 7-month treatment phase was maintained across the follow-up period (7–24 months) for both CT and IPT. Effect sizes of change throughout the follow-up period were small. No differential effects between conditions were found. Two-thirds of the treatment responders completed the follow-up phase without meeting criteria for relapse on the BDI-II. Relapse rates assessed with the LIFE were somewhat lower. Patients who responded to IPT were no more likely to experience a return of symptoms than patients who responded to CT. The between-condition differences that were observed favored prior CT slightly but were not significant. This is important because CT has been shown to have an enduring effect that lasts beyond the end of treatment (relative to prior ADM) whereas IPT has not. Our findings are far from conclusive, but the findings from this trial suggest that IPT just might have a prophylactic effect.

One is always careful to make too much of what are essentially null findings, but there are several reasons why we think we might be justified in doing so in this instance. First, relapse rates were low in both conditions (around 33%), and are within the range of those reported in the previous CT studies (see e.g. Vittengl *et al.* 2007 for an overview) and similar to those obtained in the previous IPT studies (Shea *et al.* 1992). Furthermore, evidence for CT's enduring effect is relatively robust (Cuijpers *et al.* 2013*b*). While it is possible that we implemented CT in a less than adequate fashion, independent raters could tell CT from IPT in our trial and rated the quality of implementation as good (see Lemmens *et al.* 2015). In addition, while we relied on cross-sectional monthly assessments on the BDI over the first 5 months of follow-up and year-long retrospective assessments on the LIFE for the second year, it is unlikely that we missed many relapses or recurrences since the temporal intervals on the BDI were short and we detected more recurrences at the end of the interval covered by the LIFE than the start. The fact that CT did not outperform IPT suggests that IPT might have enduring effects as well. In the absence of a control condition that does not have an enduring effect – such as ADM – we cannot conclude that both IPT and CT were prophylactic, but it remains a real possibility.

Our study has several strengths. First of all, it was the first to examine the long-term effects of IPT since the initial study by Shea *et al.* (1992). Because we included a larger sample, provided high-quality IPT, and used more sophisticated statistical analyses techniques, our study goes beyond the initial study. Second, our RCT design provided a unique opportunity to directly compare the long-term outcomes of IPT with those of CT. Third, because we included both the BDI-II and the LIFE, our study provided information in terms of self-reported symptom severity change and in terms of clinician-rated DSM classifications. Moreover, by examining both relapse and sustained response rates, our study does not only provide information about the prognosis after successful initial treatment (i.e. what can a treatment responder expect after treatment termination?), but also about the prognosis at the start of treatment (i.e. what are the chances for successful and stable treatment effects when patients enter the clinic?). This is valuable information for clinical practice. Other strengths include the repeated-measures design and the inclusion of a series of sensitivity analyses. In addition, by using carefully considered and rather stringent definitions for response, relapse, and treatment status, we feel confident that we have not overestimated our effects.

There were limitations as well. Inherent to conducting a study with a long follow-up period, we were confronted with missing data. However, drop-out rates in our study were relatively low. Moreover, we accounted for missing data in our analyses by using mixed (multilevel) regression models, which are suitable to handle missing data (Schafer and Graham, 2002; Singer and Willett, 2003; Snijders and Bosker, 2012). Another factor complicating the interpretation of results was the naturalistic setting of the follow-up phase. Even though we addressed this by controlling for additional professional support for MDD, it is impossible to exactly control for all influencing parameters. However, this approach resembles clinical practice and hereby increases the generalizability of study findings. In addition, although the LIFE has shown to be a valid instrument to retrospectively assess depression severity up to 1 year, recall biases may have occurred. To conclude, our sample may have been too small to detect (smaller) differential effects between CT and IPT.

In sum, our findings suggest that IPT may have an enduring effect similar to that already established for CT. However, in order to make strong claims about the enduring effect for IPT, more powerful tests are needed. Comparisons to prior ADM are one possible option, as that is how the enduring effect for CT has been established. Furthermore, attention should be paid to the predictors and moderators of relapse, as we plan to do in a

https://doi.org/10.1017/S0033291718001083

follow-up publication. More insight in the (relative) long-term effects of CT and IPT, and associated factors, can provide valuable information about treatment options and prognosis for depressed patients and may assist in the process of treatment selection (DeRubeis *et al.* 2014; Huibers *et al.* 2015), hereby improving everyday health care for depression.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291718001083.

Acknowledgements. The authors wish to acknowledge the contribution of participating patients and therapists at RIAGG Maastricht. Furthermore, we gratefully thank Annie Raven and Annie Hendriks for their practical assistance during the study, and Prof Dr Robert DeRubeis and his laboratory members for their input on the definitions of response and relapse.

This research was funded by the research institute of Experimental Psychopathology (EPP), the Netherlands, and the Academic Community Mental Health Centre (RIAGG) in Maastricht, the Netherlands.

Author contributions. MH, AA, FP, and SH designed the trial. MH obtained funding for the study. LL conducted the trial and carried out recruitment and data collection. MH, AA, FP, and SH supervised throughout the study. All authors had full access to all the data in the study and share responsibility for the decision to submit for publication. LL, SvB, AA, and MH performed the data analysis and interpretation. LL drafted the manuscript in close collaboration with SvB. All authors provided critical revisions and have approved the final version of the manuscript.

Declaration of interests. None.

Notes

[1] Acute phase treatment is aimed at inducing remission.

[2] Survival is defined here as 'no relapse'.

[3] Due to practical reasons, not all interviews could be administred immediately after study termination. In case the interview took place later on, the total period between the 12-month assessment and the date of the interview was covered in the interview. However, only the ratings made between 12 and 24 months were used in the analyses.

[4] The reason for using two methods was purely practical. At the end of the follow-up phase, all patients were invited for a face-to-face interview at the research center. For practical reasons, these interviews had to take place during office hours. Those who indicated that they could not make it to the research center (i.e. because they moved away, or because they were not available during office hours) were offered a telephone interview instead (these interviews could be scheduled in the evenings as well). Note that administering the LIFE by telephone is common practice (e.g. Warshaw *et al.* 2001). Results of those interviewed by telephone (47%; CT = 40%, IPT = 54%) did not differ from those who were interviewed at the research center. In addition, nine patients provided data via a digital version of the interview, which was rated afterwards, because they could not participate in both a face-to-face and a telephone interview.

[5] These assessments covered the period between 9 and 24 months. As a result, information about additional (psychological and pharmacological) support for MDD between the months 7 and 9 is not available.

[6] The cut-off of 20 was chosen because it marks the border between mild and moderate depression on the BDI-II. Patients with mild depression only meet the minimal criteria to make the diagnosis. Clinically, this suggests that even though day-to-day functioning may still be hard work and might feel like a struggle, the depressive symptoms are manageable and result in only minor impairments in social and occupational functioning.

[7] Specific information about missing BDI-II data on each assessment point (8, 9, 10, 11, 12, and 24 months), stratified per condition, can be found in Table 2.

[8] This also includes patients whose treatments lasted longer than the expected 7-month period.

[9] Calculated as: (*n* responders without relapse/all responders) × (all responders/all participants); for CT (32/45) × (45/76) = 42.1%; for IPT (25/40) × (40/75) = 33.3%.

References

Barth J, Munder T, Gerger H, Nuësch E, Trelle S, Znoj H, Jüni P and Cuijpers P (2013) Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Medicine* **10**, e1001454.

Beck AT, Steer R and Brown GK (1996) *Beck Depression Inventory II: Manual*. Boston: Hartcourt Brace.

Beck AT, Rush AJ, Shaw BF and Emery G (1979) *Cognitive Therapy of Depression*. New York: Guilford Press.

Burcusa SL and Iacono WG (2007) Risk for recurrence in depression. *Clinical Psychology Review* **27**, 959–985.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Erlbaum.

Cox DR and Oakes D (1984) *Analysis of Survival Data*. London: Chapman & Hall.

Cuijpers P, Andersson G, Donker T and Van Straten A (2011) Psychological treatment of depression: results of a series of meta-analyses. *Nordic Journal of Psychiatry* **65**, 354–364.

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

Cuijpers P, Donker T, Weissman MM, Ravitz P and Cristea IA (2013*b*) Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open* **3**, e002542.

Cuijpers P, Hollon SD, Van Straten A, Bockting C, Berking M and Andersson G (2013*c*) The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry* **12**, 137–148.

Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT and Reynolds CF (2016) Interpersonal psychotherapy for mental health problems: a comprehensive meta-analysis. *American Journal of Psychiatry* **173**, 680–687.

De Graaf LE, Gerhards S, Evers S, Arntz A, Riper H, Severens J, Widdershoven G, Metsemakers J and Huibers M (2008) Clinical and cost-effectiveness of computerised cognitive behavioural therapy for depression in primary care: design of a randomised trial. *BMC Public Health* **8**, 224.

Derubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA and Lorenzo-Luaces L (2014) The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE* **9**, e83875.

Diggle P, Heagerty P, Liang K and Zeger S (2002) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Dobson KS, Shaw BF and Vallis TM (1985) Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology* **24**, 295–300.

Eaton WW, Shao H, Nestadt G, Lee HB, Bienvenu OJ and Zandi P (2008) Population-based study of first onset and chronicity in major depressive disorder. *Archives of General Psychiatry* **65**, 513–520.

Euroqol-Group (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 and Williams JBW (1997) *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*. New York: Biometrics Research Department New York State Psychiatric Institute.

Hollon SD, Derubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'reardon JP, Lovett ML, Young PR, Haman KL, Freeman BB and Gallop R (1984) *Systems for Rating Therapies for Depression*. Los Angeles, CA: Annual Convention of the American Psychiatric Association.

Hollon SD, Evans MD, Auerbach A, Derubeis RJ, Elkin I, Lowery A, Kriss M, Grove W, Tuason VB and 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 and Lowery HA (2005) Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Archives of General Psychiatry* **62**, 417–422.

Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P and Derubeis RJ (2015) Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLoS ONE* **10**, e0140771.

Jacobson NS and 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 and 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.

Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, Mcdonald-Scott P and Andreasen NC (1987) The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry* **44**, 540–548.

Klerman GL, Weissman MM, Rounsaville BJ and Chevron ES (1984) *Interpersonal Psychotherapy for Depression*. New York: Basis Books.

Lemmens LHJM, Arntz A, Peeters FPML, Hollon SD, Roefs A and 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.

Lemmens LHJM, Arntz A, Peeters FPML, Hollon SD, Roefs A and Huibers MJH (2015) Clinical effectiveness of cognitive therapy vs. interpersonal psychotherapy for depression: results of a randomized controlled trial. *Psychological Medicine* **45**, 2095–2110.

Lemmens LHJM, Galindo-Garre F, Arntz A, Peeters F, Hollon SD, Derubeis RJ and Huibers MJH (2017) Exploring mechanisms of change in cognitive therapy and interpersonal psychotherapy for adult depression. *Behaviour Research and Therapy* **94**, 81–92.

Schafer JL and Graham JW (2002) Missing data: our view of the state of the art. *Psychological Methods* **7**, 147–177.

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, Collins JF, Pilkonis PA, Beckham E, Glass DR and Dolan RT (1992) Course of depressive symptoms over follow-up: findings from the National Institute of Mental Health treatment of depression collaborative research program. *Archives of General Psychiatry* **49**, 782–787.

Singer JD and Willett JB (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Snijders TAB and Bosker RJ (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd Edn. London: Sage.

Solomon DA, Keller MB, Leon AC and Al. E (2000) Multiple recurrences of major depressive disorder. *American Journal of Psychiatry* **157**, 229–233.

Stuart S 2011. *IPT Adherence and Quality Scale*. Iowa: Interpersonal Psychotherapy Institute.

Van Der Does W (2002) *De Nederlandse Versie van de Beck Depression Inventory – Second Edition* (*BDI-II-NL*). Enschede: Ipskamp.

Vittengl JR, Clark LA, Dunn TW and Jarrett RB (2007) Reducing relapse and recurrence in unipolar depression: a comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology* **75**, 457–488.

Warshaw MG, Keller MB and Stout RL (1994) Reliability and validity of the longitudinal interval follow-up evaluation for assessing outcome of anxiety disorders. *Journal of Psychiatric Research* **28**, 531–545.

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

https://doi.org/10.1017/S0033291718001083