

The relation between therapy quality, therapy processes and outcomes and identifying for whom therapy quality matters in CBT and IPT for depression

Citation for published version (APA):

Brujniks, S. J. E., DeRubeis, R. J., Lemmens, L. H. J. M., Peeters, F. P. M. L., Cuijpers, P., & Huibers, M. J. H. (2021). The relation between therapy quality, therapy processes and outcomes and identifying for whom therapy quality matters in CBT and IPT for depression. *Behaviour Research and Therapy*, 139, 1-11. Article 103815. <https://doi.org/10.1016/j.brat.2021.103815>

Document status and date:

Published: 01/04/2021

DOI:

[10.1016/j.brat.2021.103815](https://doi.org/10.1016/j.brat.2021.103815)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 25 Apr. 2024



The relation between therapy quality, therapy processes and outcomes and identifying for whom therapy quality matters in CBT and IPT for depression

Sanne J.E. Bruijniks^{a,b,*}, Robert J. DeRubeis^c, Lotte H.J.M. Lemmens^d, Frenk P.M.L. Peeters^d, Pim Cuijpers^a, Marcus J.H. Huibers^{a,c}

^a Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, the Netherlands

^b Department of Clinical Psychology and Psychotherapy, University of Freiburg, Germany

^c Department of Psychology, University of Pennsylvania, Philadelphia, USA

^d Department of Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands

ARTICLE INFO

Keywords:

Major depressive disorder (MDD)
Cognitive behavioral therapy (CBT)
Interpersonal psychotherapy (IPT)
Therapy quality
Change in therapy processes

ABSTRACT

Background: Research on which therapy processes are crucial for whom is necessary to optimize treatment outcomes for major depressive disorder (MDD) but may be impeded by a lack of variation in therapy quality and because of individual differences in therapy response. The present study used a dataset with large variation in therapy quality to investigate the impact of therapy quality on change in therapy processes and outcome. Machine-learning techniques were used to explore individual differences in these relationships.

Method: Data come from a multicenter trial that randomized patients with MDD into weekly versus twice weekly sessions of cognitive behavioral therapy (CBT) or interpersonal psychotherapy (IPT). Correlations between quality of therapy, change in therapy processes and depressive symptoms were calculated. Using elastic net regression, a prognostic model was developed that investigated individual differences.

Results: There were no significant correlations between therapy quality and change in therapy process variables or between therapy quality and therapy outcome, except between therapy quality and dysfunctional thinking and IPT skills. Improvement in therapy processes played a larger role in patients whose therapy outcome was poorly predicted by their baseline characteristics.

Conclusions: In this study, in which therapy quality varied considerably, we found that therapy quality was not related to change in therapy processes or outcome. In addition, this study provides a first demonstration of proposed methods to identify individual responses to change in therapy processes. Results suggested that the importance of certain therapy processes might differ between patients. Future research into the relation between therapy quality, change in therapy processes and outcome should take into account the variation of therapy quality, focus on improved measurement of therapy quality and use a combination of machine learning techniques and experimental lab studies to determine to which extent and for whom therapy quality and change in therapy processes matter.

1. Introduction

Major depressive disorder (MDD) is one of worlds' most prevalent and disabling mental disorders (World Health Organization, 2017). Various psychological therapies – such as cognitive behavioral therapy (CBT) and interpersonal psychotherapy (IPT) – have shown to be effective treatments for MDD (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; (Cuijpers, Donker, Weissman, Ravitz, & Cristea, 2016);

Ekeblad, Falkenström, Andersson, Vestberg, & Holmqvist, 2016; Lemmens et al., 2015; Luty et al., 2007), but about half of the patients does not respond to initial treatment (Cuijpers et al., 2014; DeRubeis et al., 2005; Dobson et al., 2008) and about one-third of those who initially respond to treatment relapse within one to two years (Lemmens et al., 2019; Vittengl, Clark, Dunn, & Jarrett, 2007). Finding out how therapies work and for whom has the potential to improve the effectiveness of treatment for depression (Cohen & DeRubeis, 2018; Kazdin, 2009). If we

* Corresponding author. Department of Clinical Psychology and Psychotherapy, Institut für Psychologie Albert-Ludwigs-Universität Freiburg, Engelbergerstr. 41 D-79106, Freiburg im Breisgau, Germany.

E-mail address: sanne.bruijniks@psychologie.uni-freiburg.de (S.J.E. Bruijniks).

<https://doi.org/10.1016/j.brat.2021.103815>

Received 14 April 2020; Received in revised form 21 December 2020; Accepted 19 January 2021

Available online 28 January 2021

0005-7967/© 2021 Elsevier Ltd. All rights reserved.

know which therapy procedures (i.e., the procedures employed by the therapist) or therapy processes (i.e., the changes in the patient that are hypothesized to result from these procedures) are crucial for whom (Bruijniks, DeRubeis, Hollon, & Huibers, 2019), these procedures can be specifically targeted during therapy to optimize treatment outcome for MDD.

Yet, empirical knowledge on how and for whom treatment works is still limited (Cuijpers, Reijnders, & Huibers, 2019; Lemmens, Müller, Arntz, & Huibers, 2016; Lorenzo-Luaces & DeRubeis, 2018). First, the relation between therapy quality and outcome is unclear. Psychotherapy assumes that the (success of the) therapy procedures employed by the therapist are responsible for changes in the therapy processes and the subsequent changes in the outcome. However, studies of the relation between how well these therapy procedures are performed (i.e., also referred to as the quality of therapy or therapy competence) and change in therapy processes are scarce, while studies that investigated the relation between the quality of therapy and therapy outcome have shown mixed results (i.e., pointing to both non-significant and significant relations, Kazantzis et al., 2018; Webb, DeRubeis, & Barber, 2010). Second, mechanism research is limited. Most studies have been conducted in the context of RCTs where the associations between change in the mediator (i.e., the proxy variable that measures the therapy processes) and change in therapy outcome were investigated but a causal relationship could not be demonstrated (Lemmens et al., 2016). Furthermore, it is unclear to what extent (the interplay between) common therapy processes (for example: therapeutic alliance, Wampold, 2015) and/or specific therapy processes (for example: changes in dysfunctional thinking, behavioral activation or therapy-specific skills; Lorenzo-Luaces, Keefe, & DeRubeis, 2015; Strunk, DeRubeis, Chiu, & Alvarez, 2007) are important for changes in therapy outcome.

Adding to an already complex issue, DeRubeis and colleagues (2014) pointed out that process-outcome research may be impeded by a lack of variation in quality of therapy within therapy samples and because of the neglect of individual differences in therapy response. Individual differences in therapy response were defined by a distinction between types of patients who differ in their degree to which the quality of therapy provided to them will affect their outcomes, also called patient response patterns (PRPs). Five types of patients/response patterns were suggested: (1) patients who will always respond to therapy, regardless of its quality ('the spontaneous remitters'), (2) patients who will experience improvement without therapy but even better with good therapy ('the easy responders'), (3) patients whose improvement varies as a function of therapy quality ('the pliant responders'), (4) patients who only show improvement if quality is at least good ('challenging patients') and, (5) patients who never improve, no matter how low or high the quality of therapy provided ('the intractable patients'). Besides these individual differences in therapy response, sufficient variation in the quality of the performed therapy procedures seems necessary to investigate the relation between therapy and outcome in a meaningful way. In a simulation study, DeRubeis and colleagues (2014) showed that the correlation between quality of therapy and therapy outcome in a study depends on the distribution of both quality of therapy and PRPs. More specific, the correlation between quality of therapy and therapy outcome will be largest if variation in quality of therapy in the sample is uniform and the sample mostly consists of pliant responders (the patients whose improvement varies as a function of therapy quality). For example, if in a study all therapists provide high quality of therapy (low level of variation, skewed distribution), the correlation between quality of therapy and therapy outcome will be small, even though therapy quality might be important for therapy success. In case a study mostly consists of intractable patients, this correlation will be also small because these patients will not improve no matter how well the therapy was delivered.

One study already suggested that the variation in adherence to therapy procedures or variation in therapy processes (i.e., note that whereas therapy quality (or competence) is defined as how well therapy

procedures are performed, therapy adherence is defined as whether, or to what degree, therapy procedures are performed) might indeed be positively related to finding a relation with therapy outcome. In this comparison of the relation between CT techniques (therapy procedure), working alliance (therapy process) and therapy outcome in two patient samples that were treated in different centers (Webb et al., 2012), CT techniques, but not working alliance, were associated with outcome in one sample, while working alliance, but not CT techniques played a role in the other sample. One explanation for why working alliance and not CT techniques were related to outcome in one of the samples is that this sample had a large number of weak working alliance ratings but that there was less variability in CT techniques compared to the other sample (Webb et al., 2012).

Multiple studies have pointed to individual differences in response to therapy procedures or therapy processes. For example, cognitive methods are associated with more improvement in women compared to men and might be more relevant in severe depression, while behavioral methods and homework procedures are related to more improvement in patients who have high pre-therapy anxiety and few prior episodes (Sasso et al., 2015; Webb et al., 2012). Other studies have focused on the role of the therapy processes (Lorenzo-Luaces, DeRubeis, & Webb, 2014; Lorenzo-Luaces et al., 2017; Zilcha-Mano, Lipsitz, & Errázuriz, 2018; Zilcha-Mano, Muran, Eubanks, Safran, & Winston, 2017) and, for example, showed that the relation of therapeutic alliance with outcome was stronger in patients with no recurrent depression who received CBT (Lorenzo-Luaces et al., 2014, 2017). One study provided a preliminary test of differences in process-outcome relations for patients with a different prognosis (Forand, Huibers, & DeRubeis, 2017). In order to compute individual prognosis, a prognostic model was developed that combined multiple baseline characteristics to predict therapy outcome. The advantage of a multivariable model is that by combining different pieces of patient information it is possible to obtain a complete picture of who is likely to respond to therapy. After developing the prognostic model, it was tested whether the relation between 'patient engagement' with unguided computerized CBT (cCBT) procedures (estimated based on number of sessions, number of individual logins, number of times using the mood diary, and number of times using the activity schedule) and outcome was different for patients with different prognoses. As expected according to the theory of DeRubeis and colleagues (2014), engagement in the cCBT procedures was related to outcome in the group with medium prognosis (earlier referred to as pliant patients), but not in the patients with poor prognosis (also called the intractable patients). However, the patients who were expected to improve even without therapy (also called the easy responders) also showed a relation between engagement in the cCBT procedures and outcome. However, this method also seems highly dependent on the distribution of patients in the specific sample: if a certain sample would include more easy or challenging patients, the 'middle' part of the sample might be wrongly identified as the 'pliant' patients. Thus, although these results support the hypothesis that the importance of therapy procedures is different for different patients, it also illustrates that it is still difficult to identify the group of patients that is most responsive to therapy.

The present study focused on the relation between therapy quality, change in therapy processes and therapy outcome and explored individual differences in these relations. Data was used from a recently conducted Dutch multicenter randomized trial called the FreqMech study (Bruijniks et al., 2015). The FreqMech study investigated the role of session frequency on outcomes in CBT and IPT for depression. Results showed that twice-weekly sessions of IPT and CBT for depression led to better outcomes compared to weekly sessions while the number of sessions was equal between groups (Bruijniks et al., 2020). So far the relation between therapy quality and therapy outcome has been investigated in highly controlled research settings that ensured high quality of care (Kuyken & Tsivrikos, 2009; Shaw et al., 1999; Strunk, Brotman, DeRubeis, & Hollon, 2010; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004). Because the FreqMech dataset showed a large variation in the

quality of therapy and included measurements of both common and specific factors, this dataset was well-suited to an investigation of the relations between therapy quality, changes in therapy processes and outcome.

The aims of the present study were twofold. First, to investigate the correlations between therapy quality, change in therapy processes and therapy outcome. We hypothesized that better quality of therapy would be related to better change in therapy processes and to improvement in depressive symptoms, and that better changes in therapy processes would be related to improvement in depressive symptoms. Second, to explore individual differences in the correlations between therapy quality, change in therapy processes and therapy outcome. Because our dataset was too small to investigate individual differences in process-outcome relations for patients with different prognoses, we conducted a simplified exploratory analysis in which we used patients' baseline characteristics to predict who would be responsive to therapy quality and changes in therapy processes. We built a prognostic model that combined multiple baseline characteristics to predict therapy outcome and used the success of this model in predicting the outcome to define individual differences in responsiveness to therapy quality and changes in therapy processes. Individual differences in the success of the prognostic model were defined as the difference between the predicted outcome and the actual outcome, also called the residual. Residuals thus represented the degree of variance that was unexplained by the prognostic model that used baseline characteristics to predict therapy outcome, i.e., it represented how well patient baseline characteristics predicted depressive outcome. We expected that for the patients where the prognostic model already explained the outcome very well, therapy quality and therapy process change would not matter so much. On the other hand, in the patients where the prognostic model did a poor job, therapy quality and therapy process change might be the reason why prediction based on baseline characteristics only was difficult. In these patients, therapy quality and changes in therapy processes might play a role in predicting their outcome. We expected these patients to be responsive to therapy quality and changes in therapy processes.

2. Method

2.1. Design

This study used data from the FreqMech study (Bruijniks et al., 2015, 2020). In the FreqMech study, 200 patients with major depression were randomized into four groups: CBT weekly, CBT twice weekly, IPT weekly and IPT twice weekly. Participants who were randomized to twice-weekly sessions received 16 sessions during the first 8 weeks of therapy, and 4 sessions during the last 8 weeks (up to 20 sessions over the course of 16 weeks). Patients who were randomized to once-weekly sessions received 16 sessions during the first 16 weeks and 4 sessions during the last 8 weeks (up to 20 sessions in 24 weeks). The study included observer ratings of therapy quality based on videotapes of the sessions and patient self-report ratings of different common and specific therapy processes (therapeutic alliance, dysfunctional thinking, behavioral activation, CBT skills, IPT skills) alongside the measurement of therapy outcome (depression). All procedures were approved by the Medical Ethical Committee of VU Medical Centre Amsterdam (registration number 2014.337). Informed consent was obtained from all participants prior to participation in the study.

2.2. Participants

Participants fulfilled the following inclusion and exclusion criteria: (1) a primary diagnosis of DSM-IV or DSM-5 based major depressive disorder (including chronic depression) or DSM-5 based persistent depressive disorder, (2) $18 \geq \text{age} < 65$, (3) sufficient knowledge of the Dutch language, (4) pre-therapy score ≥ 20 on the Beck Depression Inventory II (BDI-II) (Beck, Steer, & Brown, 1996), (5) access to internet

facilities, (6) not starting antidepressants or no dosage change in the past three months, (7) no acute suicide risk, (8) absence of DSM-IV or 5 diagnosis of an abuse disorders, (9) absence of DSM-IV or 5 diagnosis of a cluster A or B personality disorder, and (10) not having received more than five sessions of adequate CBT or IPT in the previous year (clinician-evaluated at intake). Patients were allowed to participate in the study if they did not provide consent for video-taping the sessions. The present study focused on the subgroup of patients who gave consent for videotaping the sessions ($n = 126$, 63%, $n = 60$ for CBT and $n = 66$ for IPT). There were no differences between patients whose sessions were videotaped compared to the patients whose sessions were not videotaped in gender, age, having a partner, active employment, BDI-II baseline scores or number of depressive episodes. Patients who did not give consent for the videotaping their sessions were more likely to have at least one parent born outside of the Netherlands (no videotaping: $M = 0.47$ ($SD = 0.50$), video-taping: $M = 0.69$ ($SD = 0.46$), $t(143.40) = -3.03$, $p < .01$), and had higher scores on the BDI-II at month 6 (no videotaping: $M = 26.47$ ($SD = 16.50$), videotaping: $M = 20.12$ ($SD = 13.53$), $t(143) = 2.42$, $p < .05$).

2.3. Interventions

CBT for depression was based on the manual by Beck and colleagues (Beck, Rush, Shaw, & Emery, 1979) and IPT was based on the manual by Klerman and colleagues (Klerman, Weissman, Rounsaville, & Chevron, 1984). Patients received a minimum of 12 and maximum of 20 sessions, depending on their progress (total number of sessions CBT ($n = 60$): $M = 17.51$, $SD = 3.61$; IPT ($n = 66$): $M = 17.48$, $SD = 3.73$). Therapies were conducted by 60 different therapists in nine mental health centers across the Netherlands. Therapists conducted therapies in both session frequencies but all but six of the therapists employed only one of the therapy modalities in the study to prevent contamination in the treatment modalities. For eight patients, therapy was partly conducted by two different therapists.

2.4. Quality of therapeutic procedures

Quality of therapy was measured with the Cognitive Therapy Scale (Dobson, Shaw, & Vallis, 1985) (CTS; range = 11–77) for CBT and the IPT Adherence and Quality Scale (Stuart, 2015) (IPT-AQS; range = 9–125) for IPT. The CTS has shown reliability and validity and a score of 40 has been suggested as a cut-off for CBT competence (Kazantzis et al., 2018; Vallis, Shaw, & Dobson, 1986). We used the short version of the IPT-AQS, only using the items that apply to 'All sessions' and focused on measuring the quality of IPT (i.e., not adherence). There is no research on the psychometric properties of the IPT-AQS. Besides providing a total score on the CTS or IPT-AQS, for each tape, raters answered the following question on a 7-point Likert scale: 'How would you rate the clinician overall in this session, as a cognitive/interpersonal therapist?' (1 = Poor, 2 = Barely adequate, 3 = Mediocre, 4 = Satisfactory, 5 = Good, 6 = Very good, 7 = Excellent). Videos between session 6 and 12 were randomly drawn and rated by both experts (three psychotherapists and one psychologist) and master level clinical graduate students specifically trained in one of the modalities (three students per modality). In total, 129 unique CBT videos ($n = 116$ by students, $n = 13$ by experts, average number of video's per patient: 2.34 ($SD = 0.57$)) and 142 unique IPT videos ($n = 131$ by students, $n = 11$ by experts, average number of video's per patient: 2.29 ($SD = 0.58$)) were rated. For each session that was rated by multiple raters, mean scores for that specific session were computed (for CBT, 28 sessions were double rated by students, 11 sessions by one student and one expert and 3 sessions by both students and the expert; for IPT, 47 sessions were double rated by students, 13 sessions by one student and one expert and 9 sessions by both students and the expert). The reason that more videos were double-rated in IPT compared to CBT ($n = 69$ versus $n = 42$) is that more student raters were available in the IPT condition. Mean therapist competence scores were

subsequently computed per patient by calculating the mean of the competence scores for the rated sessions per patient. ICC's between the different raters were good (>0.60) for the CTS and IPT-AQS total scores, except for the IPT-AQS scores between the experts and students. ICC's of the item questions were poor. Details on the training of the raters, video ratings and inter-rater reliability coefficients (ICC's) can be found in data supplement 1. Information about the therapist characteristics for each therapy modality can be found in Table 1.

2.5. Process measures

2.5.1. Common therapy process: therapeutic alliance

Therapeutic alliance was with the short form of the Working Alliance Inventory (WAI patient version; [Busseri & Tyler, 2003](#)). The WAI measures tasks (e.g., behaviors and cognitions that form the therapeutic process), bonds (e.g., positive personal attachments between patient and therapist) and goals (e.g., therapist and patient mutually endorsing and valuing the goals) as components of the therapeutic alliance. The questionnaire consists of 12 items rated on a 5-point Likert scale and was completed online by the patient outside of the session, 1 and 6 months after the start of therapy. Note that the WAI was not measured at baseline because working alliance is not rated at baseline. Reliability (i.e., Cronbach's alpha between .82 and .85) and construct validity of the Dutch WAI-patient version has been supported ([Stinckens, Ulburghs, & Claes, 2009](#)).

2.5.2. Specific therapy processes: cognitive change, behavioral activation, therapy specific skills

Three specific therapy processes were measured with patient self-report questionnaires. First, cognitive change was measured using the Cognition Checklist (CCL; [Steer, Beck, Clark, & Beck, 1994](#)). The CCL measures automatic thoughts and cognitions related to anxiety and depression in 26 items rated on a 5-point Likert scale. Reliability (i.e., Cronbach's alpha between .89 and .92; test-retest correlations between 0.58 and .68) and construct validity of the CCL has been supported ([Steer et al., 1994](#); [Taylor, Koch, Woody, & McLean, 1997](#)). Second, behavioral activation was measured with the Behavioral Activation Depression Scale (BADs; [Raes, Hoes, Van Gucht, Kanter, & Hermans, 2010](#)) measuring avoidance behavior and activity level during the past week in 25 items rated on a 7-point Likert scale. Reliability (i.e., Cronbach's alpha = .88) and structural and construct validity of the Dutch BADs has been supported ([Raes et al., 2010](#)). Third, therapy skills were measured by the Cognitive Competencies of Cognitive Therapy Scale-Self Report (CCTS-SR; [Strunk, Hollars, Adler, Goldstein, & Braun, 2014](#)) in CBT and the IPT Skill Scale-Self Report (IPSS-SR; [Bruijniks, DeRubeis, Hollon, & Huijbers, 2019](#)) for IPT. The CCTS-SR consists of 29 items focused on the patient's ability to use behavioral activation and identify and challenge automatic thoughts and core beliefs on a 7-point Likert scale. The IPSS-SR measures the patients' ability to link interpersonal events to depressive symptoms, to deal with grief, dispute, major life changes and make use of social skills and consists of 25 items on a 7-point Likert scale. These specific factors were measured at baseline and 6 months after start of therapy. Initial reliability (i.e. Cronbach's alpha between

Table 1
Therapist characteristics per therapy modality.

	CBT ($n = 60$)	IPT ($n = 66$)
Number of therapists involved in the study, n	41	25
Clinical experience in general in years, M (SD)	9.36 (8.27)	12.55 (8.77)
Clinical experience in therapy modality in years, M (SD)	7.72 (7.16)	5.5 (7.72)

Note. For clinical experience general, data for $n = 23$ (CBT) and $n = 20$ (IPT) therapists were available. For clinical experience in therapy modality, data for $n = 27$ (CBT) and $n = 18$ (IPT) therapists were available.

.91 and .97; [Bruijniks et al., 2019](#); [Strunk et al., 2014](#)) and construct validity ([Strunk et al., 2014](#)) for the CCTS-SR were supported. Note that the IPSS-SR and the CCTS-SR were administered in all treatment conditions (CBT and IPT) of the sample.

2.6. Therapy outcome

Depressive symptoms during the past two weeks were measured with the Beck Depression Inventory II (BDI-II; [Beck et al., 1996](#)). The BDI-II is a patient self-report questionnaire that consists of 21 items that are rated from 0 to 3. A score 0–13 indicates minimal depression, 14–19 mild depression, 20–28 moderate depression and 29–63 severe depression. Depression was measured at baseline and 6 months after start of therapy. Reliability (Cronbach's alpha = .9, retest reliability between 0.73 and 0.96) and construct validity of the BDI-II has been supported ([Wang & Gorenstein, 2013](#)).

2.7. Data analyses

2.7.1. Aim 1: overall correlations between therapy quality, change in therapy processes and therapy outcome

Pearson correlations between the quality of therapy, change in therapy processes and change in depressive symptoms (outcome) were calculated. Change scores (i.e., change from baseline to six months after start of therapy) were coded so that positive scores indicated an improvement in depression and pointed to an improvement in that process (i.e. increase in therapeutic alliance, behavioral activation, therapy skills and a decrease in dysfunctional thinking). Process- and outcome change scores were adjusted for baseline values by predicting the change score (dependent variable) with the score at baseline (independent variable) (for the WAI the score at month 1) and adding the residual (i.e., the difference between the predicted and observed change score) resulting from this model to the observed change score, resulting in the adjusted change score. Correlations of 0.10, 0.30 and 0.50 are considered as small, moderate and large correlations, respectively ([Cohen, 1988](#)).

2.7.2. Aim 2: exploring individual differences in the correlations between therapy quality, change in therapy processes and therapy outcome

2.7.2.1. Model building.

First, we focused on building the best possible prediction model. Following the method of [Kim et al. \(2019\)](#), we compared model fit of three different models to arrive at the best possible model within this dataset. First, we used elastic net regression ([Zou & Hastie, 2005](#)) to develop a prognostic model that used baseline characteristics ($n = 50$) to predict therapy outcome six months after start of therapy. Elastic net regression is a regression method that combines different shrinkage methods (i.e., Lasso and Ridge regression) to identify the most important variables predicting outcome in a dataset. While Lasso regression enables deletion of variables from the model when they are not essential to the model, in Ridge regression variables are not deleted from the model but shrunk maximally if not essential to the model. Elastic net regression outperforms both forms of regression for multiple reasons, for example because it is able to cope with the model when the numbers of predictors is larger than the numbers of observations (and Lasso regression is not) and it is capable of producing a parsimonious model (which Ridge regression fails to do as it keeps all variables in the model) ([Zou & Hastie, 2005](#)). In this way, these methods aim to increase the power of the analyses (i.e., by reducing underfitting) while they also increase generalizability (i.e., by reducing overfitting, a concept that states that models perform well on training data but have high error rates when the model is performed on another dataset; [Bleeker et al., 2003](#); [Cohen & DeRubeis, 2018](#)). Also, the use of internal cross-validation helps to decrease the degree of overfitting ([Efron, 1983](#); [Efron & Tibshirani, 1993, 1997](#); [Bradley; Steyerberg et al., 2001](#)).

Despite this, models still tend to overfit and have even better results in data on which the model was developed than in new data. To prevent overfitting and investigate how the model built in one part of the dataset would generalize to another part of the dataset, the data was divided in a training set and a test set. Prognostic models were tested for CBT and IPT separately. In model building for CBT and IPT separately, the participants of whom data on the quality of received CBT and IPT was available formed the test set ($n = 60$ for CBT and $n = 66$ for IPT), while the rest of the data formed the training set ($n = 140$ for CBT, $n = 134$ for IPT).

Subsequently, out-sample predictions of the elastic net regression were compared to the out-sample predictions of a baseline model (including BDI-II baseline scores as the only predictor) and a theoretical (including predictors based on theory, CBT: BDI-II baseline, gender, current job, GAF-score, dysfunctional thinking, behavioral activation, CBT skills; IPT: BDI baseline, gender, current job, GAF-score, marital status, IPT skills; see also data supplement 2 for an exact definition of each construct). In the elastic net regression the following parameters were estimated in subsequent order: (1) alpha indicating the balance between Lasso ($\alpha = 1$) and Ridge regression ($\alpha = 0$), (2) lambda, reflecting the amount of shrinkage for the chosen alpha. The optimal alpha was chosen by estimating the cross-validated error for each potential .05 alpha between 0 and 1 (i.e., 0, 0.05, 0.10 etc.). For each possible alpha 10-fold cross validation was conducted 25 times (i.e., 25 iterations) and the cross-validated error computed. The alpha with the minimal mean cross validated error was chosen. Subsequently, the lambda with the lowest mean cross-validated error (minimal lambda) was computed using 10-fold cross validation with the optimal alpha included in the model and repeating this process 1000 times (i.e., 1000 iterations). The success of prediction (i.e., prediction accuracy) for the minimal lambda was defined as the adjusted R-squared (explained variance corrected for the number of predictors). The theoretical and baseline models were trained using 10-fold cross-validation using the caret package in R.

Because our training and test sample was smaller compared to the sample of Kim et al. (2019) (i.e., training sample $n = 140$ and test sample $n = 60$ in our study versus $n = 193$ in the training set and $n = 48$ in the test set in the study of Kim and colleagues) and might therefore be too small for adequate out-of-sample prediction, we decided to also report the within-sample predictions for each model (i.e. model was trained and tested in the same sample). Further details (i.e., pre-processing of the variables, imputation of missing data, in-sample and out-sample predictions) can be found in data supplement 2.

2.7.2.2. Individual differences in predicting therapy outcome. The model with the strongest predictions (i.e., highest explained variance) was used to investigate individual differences in predicting therapy outcome. The individual predictive value of the prognostic model for each individual was computed by subtracting the predicted outcome from the observed outcome, resulting in the residual. A small residual indicated that the predicted outcome was close to the observed outcome: the smaller the residual, the better the prognostic model was in predicting therapy outcome. In contrast, a large residual indicated a large difference between predicted and observed outcome. Positive residuals indicated that patients had better therapy outcomes than expected and negative residuals indicated that patients had worse therapy outcomes than expected. Individual differences in the importance of therapy quality and therapy processes were subsequently investigated in two steps. First, we tested the correlation between the residual score and both therapy quality (using z-scores) and change in therapy processes (i.e., where all changes in therapy processes were combined by using the mean z-score). Second, to enable testing whether patients with different residual scores respond differently to therapy quality and the different therapy processes, we split the group into two groups based on the median of the size of the residual (i.e., hypothesized non-responsive patients whose baseline characteristics had a strong predictive value in therapy

outcome: mean residual size: 3.54, $SD = 2.19$, $n = 30$ (CBT) and 2.63, $SD = 1.57$, $n = 33$ (IPT) versus the hypothesized responsive patients whose baseline characteristics had a weak predictive value in therapy outcome: mean residual size: 14.56, $SD = 4.54$, $n = 30$ (CBT) and 10.67, $SD = 5.04$, $n = 33$ (IPT)). More details on the definition of these groups can also be found in data supplement 2. Differences between groups in correlations between therapy quality and therapy outcome, and changes in therapy processes and therapy outcome were tested using a z-test (Preacher, 2002). In case of a relation between therapy quality or changes in therapy processes with patient responsiveness, we explored which baseline characteristics distinguish the responsive from the non-responsive patients using $p < .01$.

3. Results

3.1. Patient characteristics

The sample ($n = 126$) mostly consisted of females (58.7%), mean age was 38 years ($SD = 12.29$), 36.5% had a partner, 50% was employed and 44.8% experienced a recurrent depressive episode. The uniform distributions of quality of treatment are shown in Fig. 2. Demographics and mean scores on the quality of therapy, therapy processes and therapy outcome per therapy modality can be found in Table 2.

3.2. Overall correlations between therapy quality, change in therapy processes and therapy outcome

There were no significant correlations between therapy quality and change in depressive symptoms (CBT: $r = -0.13$, $p = .35$; IPT: $r = -0.05$, $p = .73$) or between therapy quality and changes in therapy processes (CBT: BADS $r = -0.24$, $p = .09$; CCTS-SR $r = -0.10$, $p = .48$; WAI $r = -0.26$, $p = .07$; IPT: BADS $r = -0.07$, $p = .56$; IPSS-SR $r = -0.10$, $p = .49$; CCTS-SR $r = -0.11$, $p = .45$; CCL $r = -0.05$, $p = .69$; WAI $r = -0.24$, $p = .10$), except for the correlation between therapy quality and dysfunctional thinking ($r = -0.31$, $p < .05$) and IPT skills ($r = -0.35$, $p < .05$) in CBT. Here, higher CBT quality was related to less improvement in dysfunctional thinking and a decrease of IPT skills. Improvement on therapy processes, on the other hand, was significantly correlated with improvement in depressive symptoms in both CBT and IPT: improvement in depression was correlated with an improvement in behavioral activation (CBT: $r = 0.69$, $p < .01$; IPT: $r = 0.61$, $p < .01$), IPT skills (CBT: $r = 0.67$, $p < .01$; IPT: $r = 0.45$, $p < .01$), CBT skills (CBT: $r = 0.61$, $p < .01$; IPT: $r = 0.36$, $p < .05$), therapeutic alliance (CBT: $r = 0.29$, $p < .05$; IPT: $r = 0.35$, $p < .05$) and dysfunctional thinking (CBT: $r = 0.70$, $p < .01$; IPT: $r = 0.75$, $p < .01$).

3.3. Exploring individual differences in correlations between therapy quality, change in therapy processes and therapy outcome

Individual differences in the success of the prognostic model in predicting the outcome was not correlated to quality of therapy ($r = -0.06$, $p = .63$ for CBT and $r = -0.13$, $p = .28$ for IPT, left side Fig. 3a and b). In other words, therapy quality was not differently important for patients for whom it was easy (small residuals) or difficult (large residuals) to predict therapy outcome. However, there was a significant correlation between the residual scores and change in therapy processes ($r = 0.74$, $p < .001$ for CBT and $r = 0.58$, $p < .001$ for IPT, right side Fig. 3a and b): change in therapy processes was differently important for patients with different sizes of residuals. When patients were split into two groups based on the residuals (patients for whom it was difficult to predict therapy outcome (large residuals) versus patients whose baseline characteristics already had a strong predictive value (small residuals)), there were no differences in therapy quality between the groups (also see Table 3), but change in therapy processes was larger in the patients for whom it was difficult to predict therapy outcome. This group seemed more responsive to changes in therapy processes compared to the other

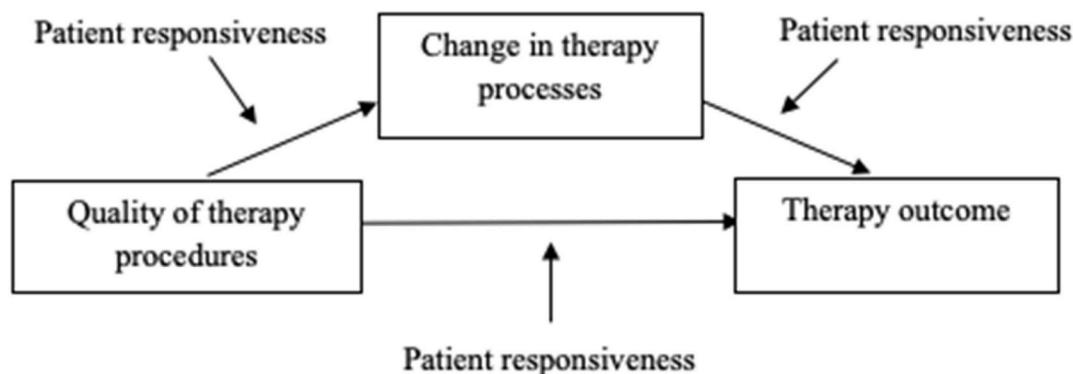


Fig. 1. Framework of the present study, focusing on correlations, and individual differences in correlations (also called patient responsiveness) between therapy quality, change in therapy processes and therapy outcome. Note that we do not intend to analyze these bivariate relations as part of a mediation model.

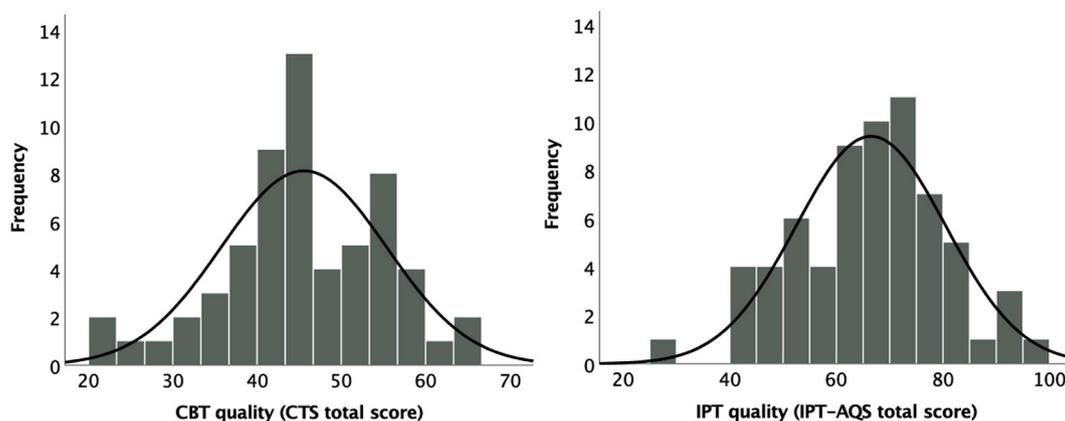


Fig. 2. Distributions of therapy quality for CBT (left; Cognitive Therapy Scale, CTS); range = 11–77) and IPT (right; IPT Adherence and Quality Scale, IPT-AQS); range = 9–125).

group: correlations of dysfunctional thinking, behavioral activation, IPT skills in CBT and of dysfunctional thinking in IPT with therapy outcome were significantly higher compared to the correlations in the other, non-responsive group (also see Table 4).

Analyses on the differences in patient baseline characteristics between responsive and non-responsive groups revealed that in CBT the responsive patients used more medication for anxiety or sleep problems ($p < .01$). There were no significant differences between responsive and non-responsive patients in IPT.

4. Discussion

This study investigated the relation between quality of therapy, change in therapy processes and depressive outcome. In a dataset with a large variation in therapy quality, no significant correlations were found between therapy quality and change in common and specific therapy processes or between therapy quality and therapy outcome, except for a correlation between therapy quality and dysfunctional thinking and IPT skills in CBT: higher CBT quality was related to less improvement in dysfunctional thinking and a decrease of IPT skills. Improvement in therapy processes was significantly associated with improvement in depressive symptoms. In addition, patients whose baseline characteristics explained little in therapy outcome, improvement in therapy processes played a larger role in the improvement of depressive symptoms compared to those whose baseline characteristics already explained substantial variance in therapy outcome.

The common assumption in the field of psychotherapy that the quality of therapeutic procedures determine change in therapy processes and therapy outcome was not supported by the present results. One

likely explanation for the lack of a clear relation between quality of therapy and therapy outcome is that, although variation in therapy quality was large and uniformly distributed, the overall therapy quality was too low to find a relation between therapy quality, changes in therapy processes and therapy outcome. Indeed, only 43–78% and 53% of our sample received adequate CBT or IPT respectively, according to our quality check. This may indicate that the optimal sample to study the role of therapy quality not only needs to be uniformly distributed and contain large variation, but also requires that the mean of therapy quality is high enough to detect these relationships. Alternative explanations for an absence of a therapy quality–outcome relation might be that therapy quality is only related to symptom change in the beginning of treatment (Kazantzis et al., 2018) or that the relation between therapy quality and outcome can be only captured when investigated on a session-to-session basis (Strunk et al., 2010; Webb et al., 2010). One possible explanation for why better therapy quality was related to worse improvement in dysfunctional thinking and IPT skills is related to the concept of therapist responsiveness (Stiles, Honos-Webb, & Surko, 1998; Webb et al., 2010), the fact that therapist deliver their behavior according to the context. For example, especially in the case of patients with very severe depression, therapists will do their utmost to perform the treatment as good as possible (with high therapy quality as a result). Because these might be the patients who do not improve, or even deteriorate, the presence of these patients in the sample may lead to zero or negative correlations between therapy quality and subsequent changes in therapy processes and outcome.

Table 2
Patient characteristics per therapy modality.

	CBT (n = 60)	n	IPT (n = 66)	n
Demographics				
Female, n (%)	36 (60)	60	38 (57.6)	66
Age in years, <i>M</i> (<i>SD</i>)	37.81 (39.50)	60	38.18 (12.13)	66
Partner, n (%)	23 (38.3)	60	23 (34.8)	66
Active employment, n (%)	31 (51.7)	60	32 (48.5)	66
Recurrent depression, n (%)	21 (42)	50	26 (47.3)	55
Number randomized to twice-weekly sessions, n (%)	29 (48.33)	60	33 (50)	66
Quality of therapy procedures				
CBT: CTS total score, <i>M</i> (<i>SD</i>)	45.94 (9.37)	60	–	–
IPT: IPT-AQS total score, <i>M</i> (<i>SD</i>)	–	–	66.48 (14.02)	66
Quality item score, <i>M</i> (<i>SD</i>)	3.73 (1.07)	60	3.64 (1.02)	66
CTS total score ≥ 40 , %	78.3	60		
% adequate therapy	43.33	60	53.03	66
Therapy processes				
<i>Therapeutic alliance</i>				
WAI month 1, <i>M</i> (<i>SD</i>)	41.21 (10.30)	52	41.89 (8.23)	59
WAI month 6, <i>M</i> (<i>SD</i>)	43.31 (11.10)	54	44.53 (8.92)	49
<i>Dysfunctional thinking</i>				
CCL baseline, <i>M</i> (<i>SD</i>)	44.91 (15.34)	59	42.24 (17.24)	66
CCL month 6, <i>M</i> (<i>SD</i>)	35.92 (21.77)	51	32.53 (8.2)	49
<i>Behavioral activation</i>				
BADS baseline, <i>M</i> (<i>SD</i>)	68.71 (19.56)	59	68.66 (20.29)	66
BADS month 6, <i>M</i> (<i>SD</i>)	88 (25.46)	49	91.67 (27.23)	46
<i>CBT skills</i>				
CCTS-SR baseline, <i>M</i> (<i>SD</i>)	79.16 (19.25)	59	83.07 (23.37)	66
CCTS-SR month 6, <i>M</i> (<i>SD</i>)	108.28 (28.68)	52	112.20 (28.65)	49
<i>IPT skills</i>				
IPSS-SR baseline, <i>M</i> (<i>SD</i>)	91.71 (19.21)	59	92.81 (17.71)	66
IPSS-SR month 6, <i>M</i> (<i>SD</i>)	103.33 (21.51)	48	112.20 (28.65)	45
Therapy outcome				
BDI-II baseline score, <i>M</i> (<i>SD</i>)	35.60 (8.88)	60	33 (10.25)	66
BDI-II score at month 6, <i>M</i> (<i>SD</i>)	21.40 (13.93)	52	18.77 (13.09)	49

Note. BADS = Behavioral Activation Depression Scale, BDI-II = Beck Depression Inventory II, CBT = Cognitive Behavioral Therapy; CCL = Cognition Checklist; CCTS-SR = Cognitive Competencies of Cognitive Therapy Scale-Self Report; CTS = Cognitive Therapy Scale (range = 11–77); IPSS-SR = IPT Skill Scale-Self Report; IPT = Interpersonal Psychotherapy; IPT-ASQ = IPT Adherence and Quality Scale (range = 9–125); *M* = mean; *SD* = standard deviation; WAI = Working Alliance Inventory. CTS and IPT-AQS item scores were rated by the following question on a 7-point Likert scale: ‘How would you rate the clinician overall in this session, as a cognitive/interpersonal therapist?’ (1 = Poor, 2 = Barely adequate, 3 = Mediocre, 4 = Satisfactory, 5 = Good, 6 = Very good, 7 = Excellent). Percentage adequate therapy was defined as a score ≥ 4 . Earlier studies reported a mean score of 3.76 (*SD* = 0.76) on the WAI in a general clinical population (i.e., note that a mean score of 3.76 relates to a mean total score of 45.12; Stinckens et al., 2009) and a mean of 72.63 (*SD* = 20.03) on the BADS (Raes et al., 2010), 93.97 (*SD* = 26.36) on the CCTS-SR (Strunk et al., 2014) and 24.04 (*SD* = 10.96) and 15.25 (*SD* = 8.93) on the two subscales of the CCL (Steer et al., 1994) in depressed populations.

In line with previous studies, our explorative analyses illustrate that there is a group of patients for whom changes in therapy processes (in this study: changes in dysfunctional thinking, behavioral activation, and IPT skills) are more important than for others. Results additionally suggested that the patients who use medication related to anxiety and sleep problems at baseline seem to be more responsive in CBT. Although

explorative, this characteristic adds to previous findings that pointed out that less recurrent depression (Lorenzo-Luaces et al., 2014, 2017), medium prognosis (Forand et al., 2017), higher life satisfaction (Zilcha-Mano, Lipsitz, & Errázuriz, 2017) and interpersonal characteristics (Zilcha-Mano, Muran, et al., 2017) interact with therapy procedures or change in therapy processes in the effect on CBT outcome and should be included in future studies.

One of the major advantages of the present study was the extensive measurement of therapy quality by both trained student raters and experts that informed on the complexity and interpretation of these measures. First, our results indicated that, at least in this study, established rating scales yield a more reliable measure of therapy quality compared to the use of single items. Second, our findings indicate that although consistency between students and experts on the CTS was sufficient, consistency between students and experts on the IPT-AQS was low. These findings are in line with the fact that psychometric properties on the IPT-AQS are lacking, while these have been found to be sufficient for the CTS. Possible explanations for the discrepancy between experts and students in the IPT quality ratings could thus be deficits in the psychometric properties on the IPT-AQS, but could also indicate the gap in IPT knowledge and expertise between the students and the experts. In contrast to CBT, IPT is not part of the standard Master psychology program in the Netherlands. Students might therefore be less familiar with IPT, and a 4-h training might not be enough the teach them the underlying principles. By showing that interrater reliability may differ between the use of ratings instruments or single items, and between students and experts in IPT, our study underlines the complexity of making comparisons between studies that use video-ratings to rate therapy quality. Other major strengths of the present study were the large variation in therapy quality and the presence of the measurement of both common and specific process variables. In addition, the present study used advanced statistical machine learning models to select the most reliable and generalizable prediction model that was possible in the present sample size.

However, the present study also has limitations. First, although the advanced machine-learning statistics helped us to overcome issues with overfitting to a certain degree, sample size and power in the present study was still limited (albeit still relatively large in comparison with many other clinical trials in the field). Results should therefore be replicated before final conclusions on individual differences in the importance of therapy process change can be drawn. In addition, we were not able to use the out-of-sample predictions used to identify individual differences in responsiveness to therapy quality and therapy processes. Thus, the prognostic models used to identify individual differences in responsiveness to therapy quality and therapy processes might have been overfit in the present sample (Cohen & DeRubeis, 2018; Luedtke, Sadikova, & Kessler, 2019). The results of the present sample might therefore be limited and specific to the population in the present sample, that relative to earlier CBT-IPT trials includes fewer patients with a low educational background and fewer patients with recurrent depression (Bruijniks et al., 2020). Also, patients in the test sample (i.e., the patients whose therapy sessions were video rated) were less likely to have a parent born outside the Netherlands and had lower scores on the BDI-II compared to the patients in the training sample (i.e., including patients with and without video-ratings), which might have led to some degree of selection bias. Despite these limitations, we believe that this study provides a valuable first demonstration of a method to investigate individual differences in process-outcome relations that already indicated that the importance of therapy processes might vary between patients.

Second, although our study included elaborate measurements of therapy quality, interpretation of these measures remains limited. In regard to the CTS, research already showed that different raters might have different ideas about which therapist behaviors are consistent with high quality and that the characteristics of the patient or student-rater may influence the ratings (Ezawa, Bartels, Pfeifer, & Strunk, 2020;

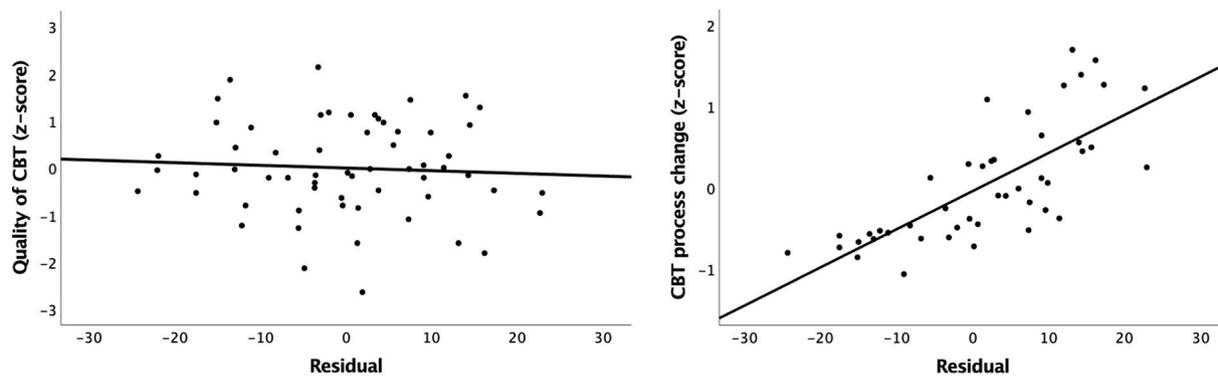


Fig. 3a. The relation between the residual (difference between observed and predicted therapy outcome, also reflecting the success of the prognostic model in predicting the outcome) and therapy quality (left), and between the residual and change in therapy processes (right), for CBT. Note that a residual of 0 would indicate that the prognostic model perfectly predicts therapy outcome, that positive residuals indicated that the patients had much better therapy outcomes than expected based on the prognostic model, and negative residuals indicated that the patients had much worse therapy outcomes than expected based on the prognostic model.

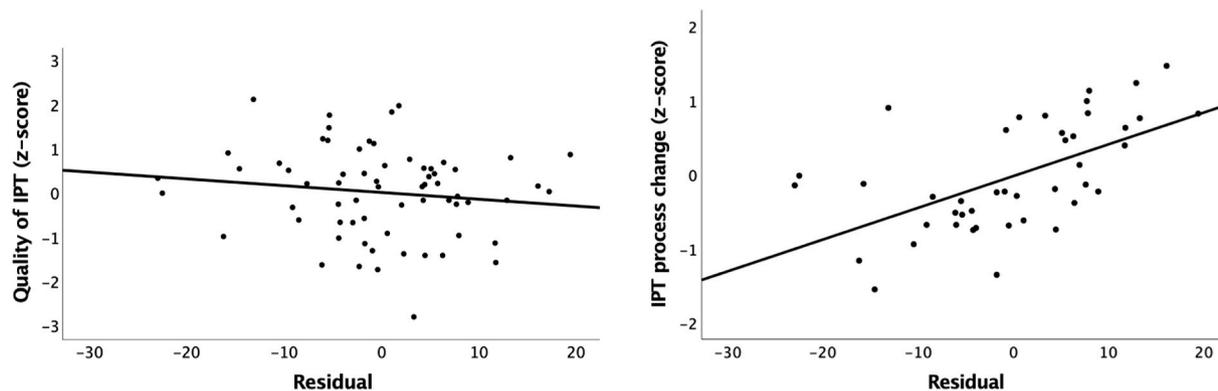


Fig. 3b. The relation between the residual (difference between observed and predicted therapy outcome, also reflecting the predictive value of the prognostic model) and therapy quality (left), and between the residual and changes in therapy processes (right), for IPT. Note that a residual of 0 would indicate that the prognostic model perfectly predicts therapy outcome, that positive residuals indicated that the patients had much better therapy outcomes than expected based on the prognostic model, and negative residuals indicated that the patients had much worse therapy outcomes than expected based on the prognostic model.

Table 3

Correlations between therapy quality and therapy outcome in relation to patient responsiveness.

	Non- responsive	n	Responsive	n
Quality of CBT	-.13	22	-.14	30
Quality of IPT	.07	19	-.12	30

Note. Quality of therapy were measured with the Cognitive Therapy Scale (CTS) for CBT and the IPT Adherence and Quality Scale (IPT-AQS) for IPT. Note that there were no significant differences in correlations between the groups.

Schmidt, Strunk, DeRubeis, Conklin, & Braun, 2018), while the IPT-QAS has not even been validated yet and the congruence between student and expert ratings on the IPT-QAS was low. A potential lack of valid and reliable measurement on IPT quality could also be a potential explanation for not finding a relation between IPT quality and therapy process or outcome change. Also, methods to rate quality of therapy differ between studies (i.e., different number of raters, different percentage of videotapes rated, different expertise of the raters, different therapist and patient samples to be rated). It is therefore very difficult to compare quality ratings between studies and it is also not very clear which therapy quality score could be considered as good therapy quality.

Future research on the relation between therapy quality, therapy processes and outcome might be improved in the several ways. First, in the context of RCTs, researchers should take the variation and mean levels of therapy quality in account when making interpretations of their findings on the relation between quality of therapy, therapy processes

Table 4

Individual differences in the correlation between therapy processes and therapy outcome.

	Non- responsive	n	Responsive	n
<i>CBT</i>				
Alliance (WAI)	.14	21	.41*	27
Dysfunctional thinking (CCL) ^a	.35	22	.77**	29
Behavioral activation (BADs) ^b	.28	20	.82**	29
CBT skills (CCTS-SR)	.54**	22	.65**	30
IPT skills (IPSS-SR) ^b	.09	19	.76**	29
<i>IPT</i>				
Alliance (WAI)	.34	19	.37	28
Dysfunctional thinking (CCL) ^b	.24	19	.84**	30
Behavioral activation (BADs)	.47	16	.70**	30
CBT skills (CCTS-SR)	.40	19	.36	30
IPT skills (IPSS-SR)	.22	15	.50**	30

Note. BADs = Behavioral Activation Depression Scale; CBT = Cognitive Behavioral Therapy; CCL = Cognition Checklist; CCTS-SR = Competencies of Cognitive Therapy Scale – Self Report; IPT = Interpersonal Psychotherapy; IPSS-SR = Interpersonal Psychotherapy Skills Scale – Self Report; WAI = Working Alliance Inventory. * significant correlation at $p < .05$, ** significant correlation at $p < .01$. ^a significant difference in correlation between the responsive and non-responsive group at $p < .05$, ^b significant difference in correlation between the responsive and non-responsive group at $p < .01$.

and therapy outcome. In case variation in therapy quality or the overall therapy quality mean is low, or therapy quality is non-uniformly distributed, interpretations about the relation between therapy

quality, therapy processes and therapy outcome might be limited. Second, the measurement of therapy quality should be improved. New developments in improving the ratings of the quality of cognitive therapy are described by Schmidt et al. (2018) and Boyle et al. (2019) and should be used in new studies that include ratings of therapy quality. Based on our findings, future studies might consider a longer and more extensive training to increase the interrater reliability of the ratings. Third, more research is necessary to develop a complete picture of the therapy processes that are important for specific types of patients to alleviate depression. So far studies mostly focused on individual differences in response to therapeutic alliance and engagement in CBT, while the results in the present study suggest that there might also be individual differences in the role of changes in dysfunctional thinking, behavioral activation and IPT skills in the improvement of depression. Future studies should focus on both common and specific therapy processes, distinguish between treatment modalities (for example: CBT versus IPT) and recognize that within therapy modalities the relation between patient characteristics and process change might be different for different levels of therapy quality. Machine learning approaches (i.e. random forest modeling, Bootstep AIC; Garge, Bobashev, & Eggleston, 2013; Rizopoulos & Rizopoulos, 2009) may be helpful to increase insight into whether different patient characteristics interact with changes in different therapy processes to reduce depression (for a recent example see: Rubel, Zilcha-Mano, Giesemann, Prinz, & Lutz, 2020). Algorithms that predict which patient is most responsive to which therapy processes might help to optimize the delivery of treatment packages. For example, a patient who is predicted to be most responsive to changes in dysfunctional thinking to improve should probably start with cognitive change procedures to alleviate depression, while for the patient where change in behavioral activation is predicted to be most important, it might be best to start with behavioral activation. Subsequent tests should point out whether these personalized treatment paths improve outcomes in psychotherapy for depression. Besides the use of large datasets to develop algorithms for personalizing treatment, experimental lab studies could inform on what therapeutic procedure is helpful for whom. For example, the utility of algorithms that have been developed in large observational samples could be tested in (sub) clinical lab studies before conducting a large randomized trial. Another way to increase insight in the role of therapy quality and individual differences in therapy outcomes that allows for more control on the variation of therapy quality might be by ranking clinically trained and untrained study therapists on competence by using the Standardized Competence Assessment Rating Scale and subsequently randomize (sub) clinical participants to different levels of therapy quality (Schmidt et al., 2018). Last, future studies should pay attention to the external validation of models that are used to personalize treatments (Van Bronswijk, Bruijniks et al., 2020). Although data from randomized controlled trials can provide valuable insight and recommendations for future research, results cannot be automatically be generalized to other settings or samples. In the context of the present study would it therefore be interesting to see whether the present results can be repeated in another dataset.

The present study investigated the relation between the quality of therapy, change in therapy processes and therapy outcome and explored individual differences in the importance of therapy quality and therapy processes. In a sample with large variation, but overall low levels of therapy quality, there was no correlation between therapy quality and therapy outcome or between therapy process change and therapy outcome, except for a correlation between therapy quality and dysfunctional thinking and IPT skills in CBT, while improvements in both common and specific therapy processes were related to improvements in depressive symptoms. The importance of the role of certain therapy processes in the improvement of depression differed between patients. Future research should replicate the proposed method and findings, take into account the variation of therapy quality, focus on improved measurement of therapy quality and use a combination of machine learning techniques and experimental lab studies to

disentangle to which extent, and for whom, therapy quality and therapy processes play a role in the reduction of depression.

Funding

The FreqMech study was funded by ZonMw and Stichting tot Steun VCVGZ. The funders were not involved in the study design; the collection, analysis and interpretation of data; the writing of the report or the decision to submit the article for publication.

CRedit authorship contribution statement

Sanne J.E. Bruijniks: designed the study, Formal analysis, conducted the analyses and, Writing - original draft, wrote the manuscript, All authors contributed to the final version of the manuscript. **Robert J. DeRubeis:** designed the study, All authors contributed to the final version of the manuscript. **Lotte H.J.M. Lemmens:** All authors contributed to the final version of the manuscript. **Frenk P.M.L. Peeters:** All authors contributed to the final version of the manuscript. **Pim Cuijpers:** All authors contributed to the final version of the manuscript. **Marcus J.H. Huibers:** designed the study, All authors contributed to the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

Acknowledgements

We acknowledge the contribution of participating patients and therapists. Without their cooperation the trial would have been impossible to conduct. Furthermore, we would like to thank Carolien Christ, Kosse Jonker, Ina Leeuw and Dina Snippe for their contributions to the expert competence ratings and Dominique Selviyan, Ferestha Qayumi, Denise Richter, Ruth Horsthuis, Yesim Karaslan, Melin Kaya for their contributions to the student competence ratings.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.brat.2021.103815>.

References

- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II: Manual*. Boston: Harcourt Brace.
- Bleeker, S. E., Moll, H. A., Steyerberg, E. W., Donders, A. R. T., Derksen-Lubsen, G., Grobbee, D. E., et al. (2003). External validation is necessary in prediction research: A clinical example. *Journal of Clinical Epidemiology*, 56(9), 826–832. [https://doi.org/10.1016/S0895-4356\(03\)00207-5](https://doi.org/10.1016/S0895-4356(03)00207-5).
- Boyle, K., Deisenhofer, A., Rubel, J. A., Bennemann, B., Weinmann-lutz, B., & Lutz, W. (2019). Assessing treatment integrity in personalized CBT: The inventory of therapeutic interventions and skills assessing treatment integrity in personalized CBT: The inventory of therapeutic interventions and skills. *Cognitive Behaviour Therapy*, 1–18. <https://doi.org/10.1080/16506073.2019.1625945>.
- Bruijniks, S. J. E., Bosmans, J., Peeters, F. P. M. L., Hollon, S. D., van Oppen, P., van den Boogaard, M., et al. (2015). Frequency and change mechanisms of psychotherapy among depressed patients: Study protocol for a multicenter randomized trial comparing twice-weekly versus once-weekly sessions of CBT and IPT. *BMC Psychiatry*, 15(1). <https://doi.org/10.1186/s12888-015-0532-8>.
- Bruijniks, S. J. E., DeRubeis, R. J., Hollon, S. D., & Huibers, M. J. H. (2019). The potential role of learning capacity in cognitive behavior therapy for depression: A systematic review of the evidence and future directions for improving therapeutic learning. *Clinical Psychological Science*, 7(4), 668–692. <https://doi.org/10.1177/2167702619830391>.
- Bruijniks, S. J. E., Lemmens, L. H. J. M., Hollon, S. D., Peeters, F. P. M. L., Cuijpers, P., Arntz, A., et al. (2020). Seeing depressed patients twice weekly improves psychotherapy outcomes in depressed patients. *British Journal of Psychiatry* (in press).

- Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the working alliance inventory and working alliance inventory, short form. *Psychological Assessment*, 15(2), 193–197. <https://doi.org/10.1037/1040-3590.15.2.193>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: NY: Academic.
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review Of Clinical Psychology*, 14, 209–236. <https://doi.org/10.1146/annurev-clinpsy-050817-084746>.
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, 15(3), 245–258. <https://doi.org/10.1002/wps.20346>.
- Cuijpers, P., Donker, T., Weissman, M. M., Ravitz, P., & Cristea, I. A. (2016). Interpersonal psychotherapy for mental health problems: A comprehensive meta-analysis. *American Journal of Psychiatry*, 173(7), 680–687. <https://doi.org/10.1176/appi.ajp.2015.15091141>.
- Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders*, 159, 118–126. <https://doi.org/10.1016/j.jad.2014.02.026>.
- Cuijpers, P., Reijnders, M., & Huibers, M. J. H. (2019). The role of common factors in psychotherapy outcomes. *Annual Review of Clinical Psychology*, 15(1), 207–231. <https://doi.org/10.1146/annurev-clinpsy-050718-095424>.
- DeRubeis, R. J., Hollon, S. D., Amsterdam, J., Shelton, R. C., Young, P. R., Salomon, R. M., et al. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, 62(4), 409–416.
- Dobson, K. S., Hollon, S. D., Dimidjian, S., Schmalzing, K. B., Kohlenberg, R. J., Gallop, R. J., et al. (2008). Randomised trial of behavioural activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology*, 76(3), 468–477. <https://doi.org/10.1037/0022-006X.76.3.468>.
- Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology*, 24(4), 295–300. <https://doi.org/10.1111/j.2044-8260.1985.tb00662.x>.
- Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78(382), 316–331.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Monographs on statistics and applied probability. New York: Chapman & Hall.
- Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438), 548–560.
- Ekeblad, A., Falkenström, F., Andersson, G., Vestberg, R., & Holmqvist, R. (2016). Randomized trial of interpersonal psychotherapy and cognitive behavioral therapy for major depressive disorder in a community-based psychiatric outpatient clinic. *Depression and Anxiety*, 33(12), 1090–1098. <https://doi.org/10.1002/da.22495>.
- Ezawa, I. D., Bartels, G. C., Pfeifer, B. J., & Strunk, D. R. (2020). Pay no attention to the man behind the curtain: The overlooked issue of rater selection in psychotherapy observer ratings. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 1–9. <https://doi.org/10.1080/10503307.2020.1780510>, 0(0).
- Forand, N. R., Huibers, M. J. H., & DeRubeis, R. J. (2017). Prognosis moderates the engagement–outcome relationship in unguided CBT for depression: A proof of concept for the prognosis moderation hypothesis. *Journal of Consulting and Clinical Psychology*, 85(5), 471–483. <https://doi.org/10.1037/ccp0000182>.
- Garge, N. R., Bobashev, G., & Eggleston, B. (2013). *Random forest methodology for model-based recursive partitioning: The mobForest package for R*.
- Kazantzis, N., Clayton, C., Cronin, T., Farchione, D., Limburg, K., & Dobson, K. S. (2018). The cognitive therapy scale and cognitive therapy scale-revised as measures of therapist competence in cognitive behavior therapy for depression: Relations with short and long term outcome. *Cognitive Therapy and Research*, 42(4), 385–397.
- Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 19(4–5), 418–428. <https://doi.org/10.1080/10503300802448899>.
- Kim, T. T., Dufour, S., Xu, C., Cohen, Z. D., Sylvia, L., Deckersbach, T., et al. (2019). Predictive modeling for response to lithium and quetiapine in bipolar disorder. *Bipolar Disorders*, 1–9. <https://doi.org/10.1111/bdi.12752>.
- Klerman, G. L., Weissman, M. M., Rounsaville, B., & Chevron, E. (1984). *Interpersonal therapy of depression (IPT)*. New York: Basic Books.
- Kuyken, W., & Tsvirikos, D. (2009). *Therapist competence, comorbidity and cognitive-behavioral therapy for depression* (pp. 42–48). <https://doi.org/10.1159/000172619>.
- Lemmens, L. H. J. M., Arntz, A., Peeters, F., Hollon, S. D., Roefs, A., & Huibers, M. J. H. (2015). Clinical effectiveness of cognitive therapy v. interpersonal psychotherapy for depression: Results of a randomized controlled trial. *Psychological Medicine*, 45(10), 2095–2110. <https://doi.org/10.1017/S0033291715000033>.
- Lemmens, L. H. J. M., Müller, V. N. L. S., Arntz, A., & Huibers, M. J. H. (2016). Mechanisms of change in psychotherapy for depression: An empirical update and evaluation of research aimed at identifying psychological mediators. *Clinical Psychology Review*, 50, 95–107. <https://doi.org/10.1016/j.cpr.2016.09.004>.
- Lemmens, L., Van Bronswijk, S., Peeters, F., Arntz, A., Hollon, S., & Huibers, M. (2019). Long-term outcomes of acute treatment with cognitive therapy v. interpersonal psychotherapy for adult depression: Follow-up of a randomized controlled trial frequency and change mechanisms of psychotherapy among depressed patients: A multicenter randomize. *Psychological Medicine*, 49(3), 465–473. <https://doi.org/10.1017/S0033291718001083>.
- Lorenzo-Luaces, L., & DeRubeis, R. J. (2018). Miles to go before we sleep: Advancing the understanding of psychotherapy by modeling complex processes. *Cognitive Therapy and Research*, 42(2), 212–217. <https://doi.org/10.1007/s10608-018-9893-x>.
- Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 82(2), 368–373. <https://doi.org/10.1037/a0035994>.
- Lorenzo-Luaces, L., Driessen, E., DeRubeis, R. J., Van, H. L., Keefe, J. R., Hendriksen, M., et al. (2017). Moderation of the alliance–outcome association by prior depressive episodes: Differential effects in cognitive-behavioral therapy and short-term psychodynamic supportive psychotherapy. *Behavior Therapy*, 48(5), 581–595. <https://doi.org/10.1016/j.beth.2016.11.011>.
- Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2015). Cognitive-behavioral therapy: Nature and relation to non-cognitive behavioral therapy. *Behavior Therapy*. <https://doi.org/10.1016/j.beth.2016.02.012>.
- Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science*. <https://doi.org/10.1177/2167702618815466>, 2167702618815466.
- Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M. A., Mulder, R. T., et al. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry*, 190, 496–502. <https://doi.org/10.1192/bjp.bp.106.024729>. JUNE.
- Preacher, K. J. (2002). Calculation for the test of the difference between two independent correlation coefficients. Retrieved from <http://www.quantpsy.org/corrttest/corrttest.htm>.
- Raes, F., Hoes, D., Van Gucht, D., Kanter, J. W., & Hermans, D. (2010). The Dutch version of the behavioral activation for depression scale (BADS): Psychometric properties and factor structure. *Journal of Behavior Therapy and Experimental Psychiatry*, 41(3), 246–250. <https://doi.org/10.1016/j.jbtep.2010.02.001>.
- Rizopoulos, D., & Rizopoulos, M. D. (2009). *Package 'bootStepAIC'*.
- Rubel, J. A., Zilcha-Mano, S., Giesemann, J., Prinz, J., & Lutz, W. (2020). Predicting personalized process–outcome associations in psychotherapy using machine learning approaches—a demonstration. *Psychotherapy Research*, 30(3), 300–309. <https://doi.org/10.1080/10503307.2019.1597994>.
- Sasso, K. E., Strunk, D. R., Braun, J. D., Derubeis, R. J., Brotman, M. A., & Program, A. D. (2015). Identifying moderators of the adherence–outcome relation in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 83(5), 976–984. <https://doi.org/10.1037/ccp0000045> (Identifying).
- Schmidt, I. D., Strunk, D. R., DeRubeis, R. J., Conklin, L. R., & Braun, J. D. (2018). Revisiting how we assess therapist competence in cognitive therapy. *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-018-9908-7>, 0(0), 0.
- Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Lowery, A., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive. *Therapy of Depression*, 67(6), 837–846.
- Steer, R. A., Beck, A. T., Clark, D. A., & Beck, J. S. (1994). Psychometric properties of the Cognition Checklist with psychiatric outpatients and university students. *Psychological Assessment*, 6(1), 67–70. <https://doi.org/10.1037/1040-3590.6.1.67>.
- Steyerberg, E. W., Harrell, F. E., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54, 774–781.
- Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, 5(4), 439–458. <https://doi.org/10.1111/j.1468-2850.1998.tb00166.x>.
- Stinckens, N., Ulburghs, A., & Claes, L. (2009). De werkkaliente als sleutellement in het therapiegebeuren Meting met behulp van de WAV-12, de Nederlandstalige verkorte versie van de Working Alliance Inventory. *Tijdschrift Klinische Psychologie*, 1, 44–60.
- Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*, 78(3), 429–437. <https://doi.org/10.1037/a0019631>.
- Strunk, D. R., DeRubeis, R. J., Chiu, A. W., & Alvarez, J. (2007). Patients' competence in and performance of cognitive therapy skills: Relation to the reduction of relapse risk following treatment for depression. *Journal of Consulting and Clinical Psychology*, 75(4), 523–530. <https://doi.org/10.1037/0022-006X.75.4.523>.
- Strunk, D. R., Hollars, S. N., Adler, A. D., Goldstein, L. A., & Braun, J. D. (2014). Assessing patients' cognitive therapy skills: Initial evaluation of the Competencies of cognitive therapy scale. *Cognitive Therapy and Research*, 38(5), 559–569. <https://doi.org/10.1007/s10608-014-9617-9>.
- Stuart, S. (2015). *IPT adherence and quality scale*. Iowa: Interpersonal Psychotherapy Institute (Unpublished). Retrieved from <https://iptinstitute.com/wp-content/uploads/2015/09/IPT-Quality-Adherence-Scale.pdf>.
- Taylor, S., Koch, W. J., Woody, S., & McLean, P. (1997). Reliability and validity of the Cognition Checklist with psychiatric outpatients. *Assessment*, 4(1), 9–16.
- Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E., & Barkham, M. (2004). Therapist competence and outcome of cognitive therapy for depression. *Cognitive Therapy and Research*, 28(2), 143–157. <https://doi.org/10.1023/B:COTR.0000021536.39173.66>.
- Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 54(3), 381–385. <https://doi.org/10.1037/0022-006X.54.3.381>.
- Van Bronswijk, S. C., Bruijniks, S. J. E., Lorenzo-Luaces, L., DeRubeis, R. J., Lemmens, L. H. J. M., Peeters, F. P. M. L., et al. (2020). Cross-trial prediction in psychotherapy: External validation of the Personalized Advantage Index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression. *Psychotherapy Research*, 1–14. <https://doi.org/10.1080/10503307.2020.1823029>, 0(0).
- Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (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, 475–488. <https://doi.org/10.1037/0022-006X.75.3.475>. June.
- Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, 14(3), 270–277. <https://doi.org/10.1002/wps.20238>.
- Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck depression inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, 35(4), 416–431. <https://doi.org/10.1590/1516-4446-2012-1048>.
- Webb, C. A., DeRubeis, R., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200–211. <https://doi.org/10.1037/a0018912>. Therapist.
- Webb, C. A., DeRubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology*, 80(3), 373–381. <https://doi.org/10.1037/a0027663>.
- World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. CC BY-NC-SA 3.0 IGO (pp. 1–24). Licence: Geneva: World Health Organization. Retrieved from https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/.
- Zilcha-Mano, S., Lipsitz, I., & Errázuriz, P. (2017). When is it effective to focus on the alliance? Analysis of a within-client moderator. *Cognitive Therapy and Research*, 41(1), 1–13. <https://doi.org/10.1007/s10608-017-9867-4>, 0(0).
- Zilcha-Mano, S., Lipsitz, I., & Errázuriz, P. (2018). When is it effective to focus on the alliance? Analysis of a within-client moderator. *Cognitive Therapy and Research*, 42(2), 159–171. <https://doi.org/10.1007/s10608-017-9867-4>.
- Zilcha-Mano, S., Muran, J. C., Eubanks, C. F., Safran, J. D., & Winston, A. (2017). Not just a non-specific factor: Moderators of the effect of within- and between-clients alliance on outcome in CBT. *Cognitive Therapy and Research*, 41(1), 1–13. <https://doi.org/10.1007/s10608-017-9866-5>, 0(0).
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.