

# Predicting response to cognitive therapy and interpersonal therapy, with or without antidepressant medication, for major depression: A pragmatic trial in routine practice

Citation for published version (APA):

Huibers, M. J. H., van Breukelen, G., Roelofs, J., Hollon, S. D., Markowitz, J. C., van Os, J., Arntz, A., & Peeters, F. (2014). Predicting response to cognitive therapy and interpersonal therapy, with or without antidepressant medication, for major depression: A pragmatic trial in routine practice. *Journal of Affective Disorders*, 152-154, 146-154. <https://doi.org/10.1016/j.jad.2013.08.027>

## Document status and date:

Published: 01/01/2014

## DOI:

[10.1016/j.jad.2013.08.027](https://doi.org/10.1016/j.jad.2013.08.027)

## 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](http://www.umlib.nl/taverne-license)

## Take down policy

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

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 02 Jun. 2023



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Affective Disorders

journal homepage: [www.elsevier.com/locate/jad](http://www.elsevier.com/locate/jad)

## Research report

# Predicting response to cognitive therapy and interpersonal therapy, with or without antidepressant medication, for major depression: A pragmatic trial in routine practice

Marcus J.H. Huibers<sup>a,b,c,\*</sup>, Gerard van Breukelen<sup>d</sup>, Jeffrey Roelofs<sup>a</sup>, Steven D. Hollon<sup>e</sup>, John C. Markowitz<sup>f,g</sup>, Jim van Os<sup>h,i</sup>, Arnoud Arntz<sup>b,c</sup>, Frenk Peeters<sup>c,h</sup><sup>a</sup> Department of Clinical Psychological Science, Research Institute Experimental Psychology, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands<sup>b</sup> Department of Clinical Psychology, VU University Amsterdam, The Netherlands<sup>c</sup> Academic RIAGG Maastricht, Maastricht, The Netherlands<sup>d</sup> Department of Methodology and Statistics, Research Institute Caphri, Maastricht University, The Netherlands<sup>e</sup> Department of Psychology, Vanderbilt University, Nashville, TN, USA<sup>f</sup> New York State Psychiatric Institute, NYC, USA<sup>g</sup> Columbia University College of Physicians & Surgeons, NYC, USA<sup>h</sup> Department of Psychiatry and Neuropsychology, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht University, Maastricht, The Netherlands<sup>i</sup> King's College London, King's Health Partners, Institute of Psychiatry, London, UK

## ARTICLE INFO

## Article history:

Received 20 March 2013

Received in revised form

22 August 2013

Accepted 27 August 2013

Available online 3 September 2013

## Keywords:

Depression

Evidence-based treatments

Psychotherapy

Antidepressant medication

Prediction of response

## ABSTRACT

**Background:** Identifying patient characteristics that predict response within treatments (prognostic) or between treatments (prescriptive) can inform clinical decision-making. In this study, we sought to identify predictors of response to evidence-based treatments in a sample of depressed patients seeking help in routine practice.

**Methods:** Data come from a pragmatic trial of 174 patients with major depression who received an evidence-based treatment of their own choice: cognitive therapy (CT), interpersonal therapy (IPT), antidepressant medication (ADM) alone or in combination with either of the two psychotherapies. Patient characteristics measured at baseline were examined to see if they predicted subsequent response as measured with the Beck Depression Inventory (BDI) over the course of 26 weeks of treatment, using mixed regression modeling.

**Results:** Higher agoraphobia scores at baseline predicted more change in depression scores across treatments, irrespective of the treatment received. Physical functioning moderated the response to treatment: patients with high physical functioning fared better in combined treatment than patients with low physical functioning, whereas physical functioning did not predict a differential response in the psychotherapy group. Moreover, the lowest levels of physical functioning predicted an increase of depressive symptoms in combined treatment.

**Limitations:** A relatively small sample size, and selection of several predictors that were less theory-driven, which hampers the translation to clinical practice.

**Conclusions:** If replicated, the prognostic and prescriptive indices identified in this study could guide decision-making in routine practice. Development of more uniform requirements for the analysis and reporting of prediction studies is recommended.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Numerous trials have demonstrated that cognitive therapy (CT) and interpersonal therapy (IPT) are effective, well-established treatments for unipolar depression (Anderson, 2001; Cuijpers

et al., 2008; Gibbons et al., 2012), and the same can be said for antidepressant medications (ADM) (Anderson, 2001; Gibbons et al., 2012). Psychotherapy and antidepressants are frequently delivered concurrently, and when combined have somewhat greater effectiveness than either modality alone (Cuijpers et al., 2009a, 2009b).

Predictors of treatment outcome are valuable, as they can guide decision-making in routine practice. Predictors come in two types, are generated by different types of designs, and can have different uses (Fournier et al., 2009). Prognostic variables derive from designs that hold treatment constant (or ignore differences in modality) and seek

\* Corresponding author at: Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. Tel.: +31 20 598 8768.

E-mail address: [m.j.h.huibers@vu.nl](mailto:m.j.h.huibers@vu.nl) (M.J.H. Huibers).

to determine whether individual differences measured at baseline predict subsequent variation in response (Kraemer et al., 2002). Prognostic indices tell us which patients benefit most from a given treatment, but not which treatment is best for a given patient. They identify which patients require special attention, such as different types or doses (i.e., more medication or sessions) of treatment.

Prescriptive variables, also known as moderators, derive from comparative treatment designs and seek to determine whether individual baseline differences predict subsequent variation in response as a function of treatment type (Kraemer et al., 2002). Prescriptive indices can determine the optimal treatment for a given patient. As Kazdin (2007) has noted, moderation always implies differential mediation, as differential effects imply differential causal pathways to change. Understanding how treatments work can inform improving existing treatments and developing new ones. Thus, identifying predictive indices (both prognostic and prescriptive) has both clinical and scientific importance.

In the past decades, many prediction models have been reported, often from randomized treatment studies. Reviewing the vast literature on predictors of treatment outcome reveals mixed and sometimes even conflicting results (Papakostas and Fava, 2008; Nilsen et al., 2012). Mood disorders and most other psychopathology domains lack a consistent evidence base for predictors of treatment outcome. To draw conclusions from the extant literature is virtually impossible given the heterogeneity in reported study designs, patient samples, treatments, outcome measures, and statistical approaches. Prediction models are particularly prone to arbitrary findings, statistical errors, and misinterpretation, as they typically allow a large variety of modeling options in combination with large numbers of variables to study within each other's context. Moreover, as Fournier et al. (2009) point out, most prediction studies have derived from a single treatment modality (prognostic only), whereas prediction analyses encompassing more than one treatment condition might yield more information (prescriptive), especially in identifying moderators.

Rather than providing another exhaustive overview of the existing literature, we focus on three eminent randomized treatment prediction studies of greatest relevance to the present study because of the treatments they compare: Sotsky et al. (1991), Fournier et al. (2009), and Carter et al. (2011).

The National Institute of Mental Health Treatment of Depression Collaborative Research Project (TDCRP) compared the effects of 16 weeks of randomly assigned CT, IPT, ADM (imipramine) with clinical management, and placebo with clinical management. Investigators found no overall outcome differences among the groups (Elkin et al., 1989), although a reanalysis found CT to be less efficacious than ADM in more severely depressed patients, an instance of moderation (Elkin et al., 1995). Sotsky et al. used the TDCRP data to investigate other potential prognostic and prescriptive predictors of outcome (1991). Controlling for baseline depressive severity and marital status, they found six prognostic indices of depressive severity at the end of the 16-week trial: social dysfunction; cognitive dysfunction; (low) expectation of improvement; “endogenous” depression; double depression; and duration of current episode. They also identified three prescriptive indices (moderators) of treatment outcome relative to pill-placebo: patients with low social dysfunction showed a better (specific) response to IPT; patients with low cognitive dysfunction showed a better (specific) response to CT or ADM; and patients with high work dysfunction showed a better (specific) response to ADM. None of these indices predicted differential response among the active treatments.

As questions have been raised about the adequacy of the implementation of CT in the TDCRP (Jacobson and Hollon, 1996b, 1996a), a more recent treatment study was designed to investigate whether ADM truly outperforms CT in treating more severely depressed patients when both are adequately implemented

(DeRubeis et al., 2005; Hollon et al., 2005). In that trial, both CT and ADM (paroxetine) were superior to pill-placebo at 8 weeks in terms of categorical response and virtually identical to one another at week 16. Using an elegant, multivariate data analytic strategy, Fournier et al. (2009) found three prognostic indices that predicted less favorable outcome at 16 weeks regardless of treatment condition: chronic depression, older age, and lower intelligence. They also found three prescriptive indices (moderators) of outcome, indicating subsets of patients who did better in CT than in ADM: patients who were married, unemployed, or had a greater number of recent life events. They concluded that these prognostic indices might usefully identify subgroups of patients who required more or different treatments, whereas the prescriptive indices identified might define subgroups of patients who might particularly benefit from CT relative to ADM. Two previous analyses of the same sample that focused on single moderators found: (1) depressed patients with comorbid personality disorders responded better to ADM than to CT, whereas patients without comorbid personality disorders responded better to CT than to ADM (Fournier et al., 2008); and (2) patients who had previously taken ADMs responded better to CT than to ADM, whereas patients without a medication history responded similarly to either (Leykin et al., 2007a).

Finally, the Christchurch Psychotherapy for Depression Study, comparing the effectiveness of randomly assigned IPT or CT for major depression, found no differences in outcome between the two psychotherapies in the full sample (Luty et al., 2007). However, separate analyses found that patients with severe depression (Luty et al., 2007) or with a comorbid personality disorder (Joyce et al., 2007) fared better in CT than in IPT, suggesting that each was a prescriptive index that moderated response to differential treatment. A subsequent multiple prediction analysis identified three prognostic and one prescriptive indices: a single episode of depression (versus recurrent depression), a higher perceived logicalness of therapy, and a moderate belief that childhood factors caused the depression were all associated with better overall outcomes post-treatment (all prognostic), whereas patients with more comorbid personality disorder symptoms did better in CT than in IPT (prescriptive) (Carter et al., 2011).

Despite the evident relevance of these findings, the results from RCTs do not necessarily translate to the routine practice they ultimately intend to inform. It can be argued that patient samples in RCTs do not represent populations in general practice (Zimmerman et al., 2005), as randomization by definition excludes patients unwilling to be randomized and general practice often delivers less optimal treatment than in controlled trials (Westen et al., 2004). Moreover, patient preference is an essential variable in general practice that most RCTs ignore or even contravene. The transportability of evidence-based treatments to regular treatment settings and the perceived gap between science and practice in the (mental) health field has received much discussion (e.g. Chalkidou et al., 2012; Leykin et al., 2007b; Shafran et al., 2009). Nevertheless, it remains an empirical question whether the predictors and moderators identified in RCTs translate to routine practice, where choice of treatment more often reflects limited treatment availability and access than a true choice of options.

We present data from a sample of depressed patients who received an evidence-based treatment of their own choice in a naturalistic setting, using a controlled but non-randomized study design. Patients could choose between cognitive therapy (CT) or interpersonal psychotherapy (IPT), with or without antidepressant medication (ADM), or ADM alone. We previously reported on the relative effectiveness of the interventions and lack of outcome differences between them, aside from a time  $\times$  treatment interaction wherein patients who received CT improved faster across the first 16 weeks in monotherapy than combined treatment, although outcomes

were similar by 26 weeks (Peeters et al., 2012). Our goal in the current study was to identify predictors of response (both prognostic and prescriptive) to evidence-based treatment in routine care, using the analytic approach developed by Fournier et al. (2009). We believe we are the first to do so.

## 2. Methods

### 2.1. Design of the study

We conducted a controlled but non-randomized study of up to 26 weeks of treatment conducted in a naturalistic setting. Patient preference determined treatment allocation.

### 2.2. Participants

Participants were depressed patients seeking treatment at the mood disorders treatment program of an outpatient mental health care centre (RIAGG Maastricht) in Maastricht, the Netherlands. This secondary care facility treats individuals aged 18–65 with varied psychiatric disorders referred by other health professionals (e.g., general practitioners and social workers). Health insurance companies cover the entire cost of treatments at the centre. Initial screening triages patients to specialized treatment programs for further diagnostic interviewing and treatment.

The only inclusion criterion was a primary diagnosis of non-delusional major depressive disorder (MDD) determined with the Structured Clinical Interview for DSM-IV Axis I (SCID-I): (First et al., 1996). Trained masters- or doctoral-level psychologists, psychotherapists, psychiatrists, and senior psychiatric residents (supervised by psychiatrists with a minimum of 5 years clinical experience) administered the SCID-I as a routine part of the diagnostic procedure in the mood disorders program. Inter-rater reliability of the SCID-I was not assessed. The only study exclusion criteria were primary diagnoses other than MDD (e.g., bipolar disorder, psychotic disorder, or substance abuse), high acute suicide risk, and insufficient fluency in Dutch. Comorbid secondary Axis I (e.g. panic disorder) and Axis II diagnoses (e.g. avoidant personality disorder) were permitted.

The Ethics Committee of Maastricht University approved the study. All participants provided written informed consent, after which they were invited to the clinic for the baseline assessment. Treatment usually started a week after the baseline assessment. Of all eligible individuals referred to the mood disorders treatment program, 65% agreed to study participation. No information about baseline characteristics of non-participants is available, which prevents comparing them with study participants.

### 2.3. Treatments

In the mood disorders program, depressed individuals preferentially receive treatment with CT, IPT, ADM, or combined CT/ADM or IPT/ADM. Patients already receiving ADM from their general practitioner at study entry were free to continue or discontinue, with or without psychotherapy.

For statistical reasons, we omitted the ADM-only group from the analysis: the subsample was small ( $n=16$ ), and as we had to build regression models that contained many predictors and interactions with treatment, including a fifth treatment group would have made prediction analysis virtually impossible which already entailed many higher order interactions (see statistical section). We chose to leave out the ADM-only group because we felt this to be the least representative group: most patients come to the mental health centre seeking some sort of psychotherapeutic help.

Length of acute treatment depended on achieving remission, which varied between 6 and 26 weeks. Patients whose treatments lasted longer than 26 weeks due to non-response were included in the analysis up until that point. CT and IPT typically ranged between 12 and 20 sessions.

During the diagnostic work-up, consisting of an open interview and the SCID-I, patients received verbal and written information about treatment options in the mood disorders program. Patients were explicitly informed that CT and IPT are time-limited, empirically validated therapies lasting a maximum of 15–20 sessions.

After this work-up, a multidisciplinary team meeting, briefly discussing clinical history and diagnosis, yielded a treatment recommendation. Treatment recommendations were mostly based on evidence-based clinical decision rules (e.g. in case of severe depression, combination treatment is recommended) or clinical intuition (e.g. in case of dominant interpersonal problems, IPT is recommended) in the absence of evidence for a recommendation. As is typical in Dutch treatment settings, patient and therapist agreement determined the final choice of treatment, predominantly guided by patient preference.

#### 2.3.1. Cognitive therapy

Licensed psychologists and psychotherapists provided CT. Their CT experience ranged from 1 to 12 years at study onset. Three therapists were faculty teaching post-graduate courses in CT. All therapists received appropriate training and followed the procedures outlined in standard texts of CT for depression (Beck et al., 1979). Weekly sessions typically lasted 50 min, with the possibility of fortnightly booster sessions in later stages. CT therapists met together weekly for 60 min to review ongoing cases and address difficulties encountered during therapy.

#### 2.3.2. Interpersonal psychotherapy

Licensed psychologists, psychotherapists, and psychiatrists provided IPT. Their IPT experience ranged from 1 to 10 years at the beginning of the study. Three therapists were faculty in post-graduate IPT courses. All therapists received appropriate training and followed the procedures outlined in standard texts of IPT for depression (Klerman et al., 1984). As in CT, sessions were held for 50 min weekly with the possibility of maintenance sessions following acute phase treatment. IPT therapists also met for an hour weekly to review ongoing cases.

#### 2.3.3. Pharmacotherapy

Patients in the combined treatment conditions received additional 15-min ADM sessions. These sessions focused on medication management (biochemical rationale, discussion of adverse events, dosage adjustment) and management of the patients' functioning (assessing functioning in major life spheres). Techniques and strategies specific to CT or IPT were neither encouraged nor strictly prohibited. Most participants received a serotonin reuptake inhibitor (SSRI) following national and international guidelines. In case of SSRI non-response in the current episode, participants were prescribed another SSRI, venlafaxine, or a tricyclic, augmented with lithium when necessary.

### 2.4. Outcome measure

The Beck Depression Inventory Second Edition (BDI-II): (Beck et al., 1996) was the main outcome measure, assessed pre-treatment (before treatment allocation) and at 8, 16, and 26 weeks. The BDI measures depressive severity, with higher scores indicating more severe depression (range 0–63). Its construct validity and reliability have received consistent support in varied samples (Beck et al., 1996).

## 2.5. Predictor variables

There are two general approaches to prediction analysis: (1) associating a selection of clinically meaningful patient characteristics at baseline with prospective outcome or (2) building a statistical model that provides the (statistically) optimal combination of baseline characteristics to predict prospective outcome. We chose the latter, simultaneously investigating as many potential predictors as possible, rather than a (limited) selection of predictors based on theory. Following the recommendations of Fournier et al. (2009), we analyzed potential predictors in separate domains (see statistical analysis). Potential predictors were all measured at baseline (before treatment allocation), using validated instruments or self-report. Continuous variables were centered on the grand mean, while categorical variables were made binary for the ease of interpretation and analysis.

### 2.5.1. Domain 1: depression-related variables

We included the following variables: SCID-I (First et al., 1996) diagnosis of recurrent depression (0=first episode, 1=recurrent depression); number of previous episodes; duration of depression (months); previous depression treatment (0=no, 1=yes); and current antidepressants prescribed by GP (0=no, 1=yes).

### 2.5.2. Domain 2: demographic variables

Variables included: age, gender (0=female, 1=male); education (0=low, high school or basic education, 1=high, professional education or university); partner status (0=no partner, 1=partner); active employment (0=no, 1=yes); monthly income (0=less than €2500, 1=€2500 or more); and negative life events (e.g., death of a loved one, divorce, loss of job, etc.) in the preceding 12 months (0=no, 1=yes).

### 2.5.3. Domain 3: psychological processes and beliefs

Variables included: expectation of treatment outcome (0=learn to deal with depression, 1=become less depressed); rumination, measured with the *Ruminative Response Scale* (RRS; (Nolen-Hoeksema, 1991)); worry, measured with the *Penn State Worry Questionnaire* (PSWQ; (Meyer et al., 1990)); positive metacognitive beliefs about rumination, measured with the *Positive Beliefs about Rumination Scale* (PBRS; (Papageorgiou and Wells, 2001b)); and negative metacognitive beliefs about rumination, measured with the *Negative Beliefs about Rumination Scale* (NBRS; (Papageorgiou and Wells, 2001a)).

### 2.5.4. Domain 4: psychological distress

Psychological distress (overall emotional dysfunction) was measured with the *Symptom Check-List 90* (SCL-90; (Derogatis et al., 1973)), using its nine subscales in the Dutch translation ((Arrindell and Ettema, 1981)): agoraphobia (SCL-AGO); anxiety (SCL-ANX); depression (SCL-DEP); somatic complaints (SCL-SOM); insufficiency of thoughts and actions (SCL-INS); distrust and interpersonal sensitivity (SCL-DIS); hostility (SCL-HOS); and sleep problems (SCL-SLE). Also assessed in this domain was the total number of SCID-I (First et al., 1996) diagnoses, including depression. This variable is essentially a measure of comorbidity, and contains information on the number of comorbid disorders.

### 2.5.5. Domain 5: quality of life

The *RAND-36 Health Survey* (RAND-36; (Hays et al., 1993)) assessed quality of life, covering several health domains in nine distinct subscales: physical functioning (RAND-PHY); social functioning (RAND-SOC); role limitations (physical problem) (RAND-RPH); role limitations (emotional problem) (RAND-REM); vitality (RAND-VIT); mental health (RAND-MEN); pain (RAND-PAI); general health

perception (RAND-PER); and perceived health change during the last year (RAND-CHA).

### 2.5.6. Domain 6: reasons for depression

Patients' conceptualization of why they were depressed was assessed with the nine subscales of the *Reasons for Depression Questionnaire* (RFD; (Addis and Jacobson, 1995)): characterological (RFD-CHA), existential (RFD-EXI), interpersonal conflict (RFD-CON), intimacy (RFD-INT), achievement (RFD-ACH), relationship (RFD-REL), physical (RFD-PHY), childhood (RFD-CHI), and biological (RFD-BIO).

## 2.6. Statistical analysis

Prediction models were tested with mixed linear regression modeling, using maximum likelihood estimation in SPSS (version 18.0). The mixed model had three levels: therapists, patients, and measurements (time at 0, 8, 16, 26 weeks). There were too few patients per therapist to support meaningful analysis of therapist effects, and we therefore omitted the therapist level from the analyses. Patient effects were included by choosing the most general, unstructured covariance matrix for the repeated measures, allowing patient and patient by time effects.

The baseline recording of the BDI was included in the analyses as a repeated measure and not as a predictor, because the latter is more prone to bias in non-randomized studies than the former (Van Breukelen, 2006). Note however, that other baseline measures of depression (see domains 1 and 4) were included as baseline predictors.

The dependent variable was the BDI (measured at 0, 8, 16 and 26 weeks). Intervention was represented by two dummy variables: CT versus IPT (CTvsIPT; CT=1, IPT=0) and psychotherapy alone versus psychotherapy combined with ADM (PSYvsCOM; psychotherapy alone=1, combination treatment=0).

First, we fitted the growth curve for time (0, 8, 16 and 26 weeks, coded in the regression analyses as 0, 1, 2 and 3). It emerged that the best fit was provided not by a linear model (time), but by a polynomial model (time and time<sup>2</sup>) instead. We then tested the *relative effectiveness* of the interventions by building a treatment regression model with the main effects of time, time<sup>2</sup>, CTvsIPT, PSYvsCOM, and all two- and three-way interactions of group with time. Non-significant interaction effects were hierarchically excluded from the model, starting with the highest order term and never excluding a lower term as long as it was involved in a higher order term in the model, until only significant terms remained (using  $\alpha=0.05$ , two-tailed). The previous paper described this treatment regression model (Peeters et al., 2012).

Following Fournier et al. (2009), we subsequently evaluated *potential predictors* within their own domains. Fournier and colleagues evaluated prognostic and prescriptive indices simultaneously, but our complicated treatment regression model precluded doing this in our dataset. Within each domain, main effects of potential predictors and their interactions with time and time<sup>2</sup> (requiring three terms for each predictor) were added to the treatment regression model and hierarchically excluded, retaining those predictors with  $p < 0.10$ . In the next step, the retained predictors from all six domains were then simultaneously entered into the final model. Non-significant predictors were again eliminated hierarchically (starting with the interactions with time<sup>2</sup>) until only significant terms remained in a combined prediction model. This was done stepwise: in the first step, predictors were retained with a two-tailed  $\alpha=0.10$ , in the second step with  $\alpha=0.05$ , and in the final step with  $\alpha=0.01$ , to adjust for multiple testing.

We then analyzed *potential prescriptive indices (moderators)* of treatment outcome. For each significant predictor in the final prediction model, we built a separate model using the significant

terms of the treatment regression model as the base model; the main effects of predictor, time, CTvsIPT, PSYvsCOM; and all two-, three- and four-way interactions of these variables (i.e., 7 interaction terms for each predictor). Since time<sup>2</sup> effects are only temporary, interactions of potential moderators and time<sup>2</sup> were omitted from the model to simplify the analyses. Non-significant moderators were hierarchically excluded from the model until only significant terms remained, starting with the highest order interaction and the significance level set at  $\alpha=0.10$ . Significant moderators from these separate, single predictor models were then simultaneously added to the combined prediction model of the previous paragraph. Non-significant moderators were eliminated backwards from the model (stepwise, at  $\alpha=0.10$ ,  $\alpha=0.05$ , and  $\alpha=0.01$ ) until only significant terms remained in a final (combined) prediction and moderation model.

Moderating effects were illustrated in plots of predicted values using SPSS Graphs.

### 3. Results

#### 3.1. Study population, treatments received and attrition

One hundred seventy-four depressed patients participated in this study: 63 chose CT, 56 IPT, 34 CT combined with ADM, and 21 IPT combined with ADM. Fifty-nine percent were female ( $N=103$ ). Mean age was 42.4 years ( $SD=11.1$ ). Baseline mean BDI was 24.5 ( $SD=9.6$ ); mean duration of the current episode was 8.2 months ( $SD=13.3$ ). In total, 63 patients (36%) suffered from chronic depression (i.e., current episode lasting 2 years or longer) and 92 (53%) reported at least one comorbid Axis I disorder. Comparing the baseline characteristics (including potential predictors) of the four treatment groups (Peeters et al., 2012), we found no significant baseline differences among the groups, except that age was lower in the CT-ADM group (mean 36 years) than other groups (means 41–45 years). Age was therefore included as covariate in all initial analyses, but as adjustment did not alter the findings, we omitted it from the final prediction model.

The percentages of participants available for the prediction analyses (not lost to follow-up) at 8, 16, and 26 weeks were 91%, 83%, and 76% respectively, falling within an acceptable range of attrition. All available data from all 174 patients were included into the mixed regression without imputation of missing data and without dropping any patient.

#### 3.2. Evaluation and reduction of predictors within their own domains

Of the 47 potential predictors under study, 23 achieved significance ( $p < 0.10$ ) within their own predictor domains. Table 1 provides an overview per domain. A significant single predictor term indicates only a main effect of the predictor: the baseline value of the predictor is associated with scores on the BDI at all time points, including at baseline. A significant predictor  $\times$  time term indicates an interaction with the linear time effect: the speed of linear BDI change over time depends on the predictor value. A predictor  $\times$  time<sup>2</sup> term likewise means that the amount of quadratic BDI change depends on the predictor value.

#### 3.3. Final prediction and moderation model

As the total number of significant predictors plus their interactions with time was too large (41 terms) to combine into a single prediction model, we reduced the number of potential predictors by combining the significant predictors of domains 1, 2, 3 and 6 (21 terms) and 4 and 5 (20 terms) into two intermediate models in

**Table 1**

Summary<sup>a</sup> of significant predictors ( $p < 0.10$ ) within the separate predictor domains.

<i>Domain 1: depression-related variables</i>
Previous depression treatment
<i>Domain 2: demographic variables</i>
Age $\times$ time
Gender
Active employment
<i>Domain 3: psychological processes and beliefs</i>
Expectation of treatment outcome $\times$ time <sup>2</sup>
Rumination (RRS) $\times$ time
Negative beliefs about rumination (NBRS) $\times$ time <sup>2</sup>
<i>Domain 4: psychological distress</i>
Agoraphobia (SCL-AGO) $\times$ time
Anxiety (SCL-ANX) $\times$ time <sup>2</sup>
Depression (SCL-DEP) $\times$ time <sup>2</sup>
Somatic complaints (SCL-SOM)
Insufficiency of thoughts and actions (SCL-INS)
Sleep problems (SCL-SLE)
<i>Domain 5: quality of life</i>
Physical functioning (RAND-PHY)
Social functioning (RAND-SOC)
Vitality (RAND-VIT)
Mental Health (RAND-MEN) $\times$ time
Perception of health change (RAND-CHA) $\times$ time <sup>2</sup>
<i>Domain 6: reasons for depression</i>
Existential (RDF-EXI) $\times$ time
Interpersonal conflict (RDF-CON) $\times$ time
Achievement (RDF-ACH) $\times$ time
Childhood (RDF-CHI)
Relationship (RDF-REL)

<sup>a</sup> This is a summary of the predictors in each domain, not of the actual prediction models. For example, the 'age  $\times$  time' term in domain 2 by definition implies that the terms 'age' and 'time' were also part of the model.

which non-significant predictors (at  $\alpha=0.10$ ) were eliminated until only significant predictors remained. Thus, we reduced the number of terms to 32 for the final prediction model combining all domains. This model was then further reduced by hierarchical elimination of non-significant predictor by time and predictor effects, using first  $\alpha=0.10$ , then  $\alpha=0.05$ , and finally  $\alpha=0.01$  two-tailed, in view of multiple testing.

Table 2 presents the final prediction models for both prognostic and prescriptive indices (moderators). The treatment terms serve simply to control for the treatment effects in the prediction model. Five significant predictors emerged: depression, vitality, perception of health change, agoraphobia, and physical functioning. Table 3 summarizes the baseline values of these predictors, demonstrating the absence of significant or clinically meaningful baseline between-group differences.

Depression (measured by SCL-90), vitality, and perception of health emerged as main effect predictors of outcome (no interaction with time), meaning that they were merely associated with depression scores at all time points. Higher SCL-depression at baseline predicted higher depression BDI scores at all time points, while higher vitality and the perception of improved health in the last year predicted lower BDI scores at all time points. Agoraphobia as a predictor on the other hand interacted with time: higher agoraphobia predicted a greater decrease in BDI scores in the course of follow-up.

Only one significant prescriptive index (moderator) was identified. Physical functioning initially emerged as a significant main effect predictor. Evaluation of potential moderators in a subsequent step demonstrated that physical functioning was actually part of a moderator effect. Table 2 illustrates a significant three-way interaction of physical functioning, time, and psychotherapy alone versus combined treatment. As these results are not easily interpretable based on the regression coefficients, we plotted this

moderated effect and did so for CT and IPT separately to demonstrate that it was robust across type of psychotherapy (Fig. 1).

As shown in the figure, we plotted BDI change against physical functioning. In both the CT and IPT groups, higher physical functioning at baseline predicted greater reduction of depression scores (whereas the lowest physical functioning scores actually predicted an increase in depression scores) over time in the combined treatment group only. No such predictive effect appeared with psychotherapy alone.

#### 4. Discussion

##### 4.1. Main findings

Investigating baseline predictors of depression treatment outcome in routine practice settings, we found three main effect predictors, one prognostic predictor and one prescriptive index (moderator). Depressed patients with relatively low baseline depression scores, high vitality, and the perception of improved general health in the last year had on average lower depression scores across the course of treatment, irrespective of which

**Table 2**  
Final prediction and moderation model<sup>a</sup>.

Predictors	$\beta$	S. E.	p
<i>Treatment</i>			
Intercept	23.21	0.69	0.000
Time <sup>b</sup>	-5.37	1.28	0.000
Time <sup>2</sup>	0.46	0.40	0.249
CTvsIPT	1.73	0.96	0.073
PSYvsCOM	2.04	1.33	0.127
CTvsIPT × time	-3.05	1.78	0.089
CTvsIPT × time <sup>2</sup>	0.96	0.55	0.084
PSYvsCOM × time	-0.58	2.43	0.811
PSYvsCOM × time <sup>2</sup>	0.47	0.74	0.522
CTvsIPT × PSYvsCOM	-1.86	1.74	0.285
CTvsIPT × PSYvsCOM × time	5.74	3.19	0.074
CTvsIPT × PSYvsCOM × time <sup>2</sup>	-1.88	0.98	0.057
<i>Predictors</i>			
Depression (SCL-DEP)	0.41	0.03	0.000
Vitality (RAND-VIT)	-0.36	0.13	0.010
Perception of health change (RAND-CHA)	-1.21	0.42	0.005
Agoraphobia (SCL-AGO)	0.19	0.08	0.022
Agoraphobia (SCL-AGO) × time	-0.19	0.04	0.000
<i>Moderators</i>			
Physical functioning (RAND-PHY)	-0.32	0.11	0.002
Physical functioning × time	0.05	0.06	0.380
Physical functioning × PSYvsCOM	0.06	0.18	0.719
Physical functioning × PSYvsCOM × time	-0.28	0.10	0.011

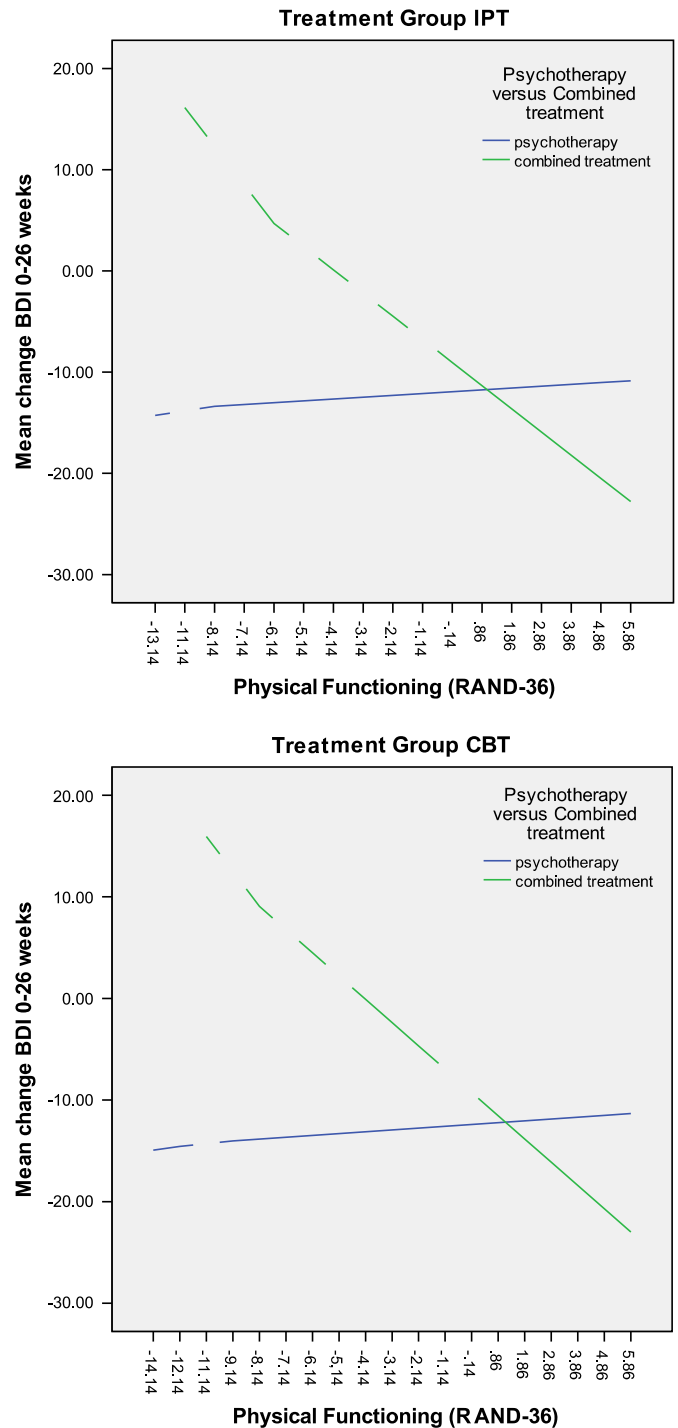
<sup>a</sup> Presented is the final prediction and moderation model (and not a summary of the intermediate models that were developed). Therefore, the results for predictors under 'treatment' differ slightly from the treatment model that was published in Peeters et al. (2012).

<sup>b</sup> Time=0, 8, 16 and 26 weeks (coded in the regression analyses as 0, 1, 2 and 3).

**Table 3**  
Predictor values at baseline, in means (sd).

Predictors	CT	IPT	IPT-ADM	CT-ADM	F (p-value) <sup>a</sup>
Depression (SCL-DEP)	46.9 (13.1)	45.5 (12.4)	50.4 (13.7)	47.8 (11.4)	0.81 (0.48)
Vitality (RAND-VIT)	9.2 (3.1)	9.7 (3.0)	9.7 (3.0)	9.6 (2.9)	0.25 (0.85)
Perception of health change (RAND-CHA)	2.2 (0.9)	2.4 (0.9)	2.2 (1.1)	2.5 (0.9)	0.81 (0.49)
Agoraphobia (SCL-AGO)	12.5 (5.2)	11.2 (4.5)	13.9 (7.3)	13.3 (6.9)	1.71 (0.16)
Physical functioning (RAND-PHY)	23.4 (5.4)	24.5 (4.6)	23.8 (4.9)	25.1 (4.7)	1.06 (0.36)

<sup>a</sup> Mean group comparison from ANOVA.



**Fig. 1.** Moderation of physical functioning, time and treatment (psychotherapy alone versus combined treatment).

empirically validated treatment they received. These predictors point in the same direction: they are all associated with relatively lower levels of depression from start to follow-up. Given the fact that they are already manifest at baseline, these factors generally would not be conceived as predictors of treatment effects.

Agoraphobia was the only predictor that actually predicted change over time: depression scores in patients with high agoraphobia scores decreased more over time than in those with lower levels of agoraphobia. A possible explanation is that suffering from agoraphobic complaints contributes to depression, and that the reduction of agoraphobic symptoms during treatment adds to the reduction of depressive symptoms.

The only moderation we found was an interaction among level of physical functioning, treatment modality, and time. Patients with high physical functioning fared better in combined treatment than patients with low physical functioning, whereas physical functioning did not predict a differential response in the psychotherapy group. Moreover, the lowest levels of physical functioning predicted an increase of depressive symptoms in combined treatment. Pragmatically, this means that patients with high physical functioning might be well advised to add medication to psychotherapy, although this finding requires replication in randomized controlled trials before it can be used to guide clinical practice.

#### 4.2. Interpretation of findings

Why and how level of physical functioning predicts differential response within combined treatment and not in psychotherapy alone remains unclear. In general, depressed patients with higher levels of somatic comorbidity have poorer course trajectories compared to patients with low levels of somatic comorbidity (Angst et al., 2009; Rhebergen et al., 2012). It can be presumed that the low levels of physical functioning in our study are a proxy for somatic comorbidity, and hence predict a less favorable response to treatment. However, this association was found only in patients receiving combined treatment. In those with low levels of physical functioning and receiving combined treatment, we even found an increase in depression – suggesting that the addition of antidepressants to psychological therapy leads to worsening of depression in those with poor physical functioning. In the absence of randomization, it cannot be concluded that low physical functioning is a contraindication for prescribing ADMs, but our findings certainly call for further research.

On the other hand, it is possible that meaningful but unmeasured variables influenced treatment planning (i.e. the pace in which steps in the therapy were taken) by patient and therapist. For example, some depressed patients may engage less actively in psychotherapy when also treated with antidepressants (as some patients report feeling numbed by the medication), and antidepressants are known to be less efficacious in depressed patients with low levels of physical functioning (Leuchter et al., 2009; Lin et al., 2006). A combination of diminished progress in psychotherapy combined with the absence of antidepressant effects from the medication might then explain the effect we found.

A common sense explanation could be that in fact the medication, rather than the combination of psychotherapy and medication, is driving the observed effect. We tried to explore this in an ancillary analysis (data not reported) in which we evaluated the slope of the association between change in depression and physical functioning in a fifth group of patients that only chose antidepressant medication, but this sample was too small to allow meaningful interpretation. These explanations are highly speculative, and the interpretation of physical functioning as moderator and comparison of our predictors with predictors found in other studies requires caution.

First, this is a non-randomized study, and we cannot assume that potential confounds that we did not measure are distributed equally among the groups, as is safer to assume in RCTs. Especially prescriptive indices that imply a differential treatment effect should be interpreted very cautiously, as it cannot be ruled out that a third unknown variable explains the effect. Prescriptive indices derived from an RCT can be used to outline treatment recommendations, but our findings do not allow such causal inferences. However, moderators detected in a non-randomized trial call for testing in a subsequent randomized trial. In that sense, the evaluation of moderators in non-randomized trials certainly has pragmatic value.

Second, our statistical approach that reduced 47 potential predictors to 5 significant predictors entails that our final model was built solely on the predictive power it generates. This may explain why all five predictors were psychometrically strong subscales of two self-report measures (i.e. SCL-90 and RAND-36) that tap into aspects of depression (viz., psychological distress and quality of life). Moreover, three predictors (depression, vitality and perception of health) did not qualify as true prognostic indices, as they were merely associated with depression scores across time. These ‘correlates’ should not come as a surprise, and are therefore less informative. Agoraphobia, the only prognostic predictor, is most likely a general proxy for comorbid anxiety that typically accompanies (or adds to) depression. The strength of our approach is the overall predictive power of the final model, but the predictors individually might have less clinical relevance or meaning.

Third, our study differs in other ways from the prediction studies mentioned in the introduction (Carter et al., 2011; Fournier et al., 2008, 2009; Joyce et al., 2007; Leykin et al., 2007a; Luty et al., 2007; Sotsky et al., 1991). We found little overlap with the predictors and moderators in these earlier studies. One explanation is that a gap indeed exists between science and practice, efficacy and effectiveness, and predictors found in RCTs do not translate to routine clinical settings that lack randomization and highly controlled study conditions. This is unlikely, however, since the aforementioned studies themselves reported differing predictors. If studies differ in design, treatments, predictors evaluated, sampling, and especially statistical approaches, as often occurs in prediction studies, direct comparison of findings loses meaning.

The nature of prediction analysis itself leaves room for the wide variation in reported findings, depending on the analytic approach and the often subjective decisions researchers must make in building their prediction models. We cannot rule out that our models would have generated different results had we been able to do this. It is therefore crucial that uniform data analytic approaches such as Fournier et al.'s (2009) are developed for future prediction studies to limit the interpretation bias that stems from the methodological heterogeneity of extant studies. Actual replication studies in this field would be very helpful.

#### 4.3. Strengths and limitations

This study has several strengths and weaknesses. A strength is its routine clinical setting, the circumstances in which antidepressant treatments are actually delivered. Some might criticize the absence of random allocation to the evidence-based treatments, but others have endorsed the need for well-designed *effectiveness* studies that evaluate best practices and the dissemination of evidence-based treatments to routine settings (McHugh and Barlow, 2010; Shafran et al., 2009). We know of no other studies that have studied several evidence-based antidepressant treatments allocated based on patient preference, directly compared in a pragmatic design.

An obvious limitation is our relatively small sample size. The multilevel analyses we conducted are known to generate increased power relative to less advanced statistical analyses, but



valid power calculations are difficult to make for these kinds of models. Like most predictor studies, we struggled with a large number of potential predictors we wished to analyze together, and we applied Fournier et al. (2008) data analytic approach to deal with this problem. Still, the final prediction model we built contained 21 terms, which runs a certain risk of overfitting the data. Our study may have been underpowered (despite finding significant predictors and moderators), especially for the more complex interactions investigated.

Our present choice of  $\alpha=0.01$  is based on two considerations. First, we had to balance between the risks of a type I and type II error, and it can be shown that if  $\alpha=0.01$ , our present sample size of  $N=174$  gives a power of 80% to detect a correlation of 0.25 between the outcome and a predictor uncorrelated to other predictors, or to detect a correlation of 0.35 between the outcome and a predictor which itself has 50% variance overlap with other predictors (Fox, 1997; Lachin, 1981). Applying a full Bonferroni correction would reduce the power for these correlations to less than 60%. Second, according to Narum (2006) the  $\alpha$  needed to control the so-called false discovery rate (i.e. proportion of false positives among all significant results) when testing 32 predictors, as in our study, is 0.012, which is very close to our  $\alpha$  of 0.01 and would imply that the moderation by physical functioning was significant.

Our selection of potential predictors was partly based on the literature, but some predictors we included were less theory-driven, derived from our standard assessment, and incorporated into our research program for other observational process studies (e.g., Roelofs et al., 2008, 2006, 2007). Assessment of DSM personality disorders would have been informative but was not feasible in our naturalistic setting.

Particular selection and allocation practices at the mental health centre may limit the generalizability of our results to other patient samples and clinics. For example, patients chose their preferred treatment partly based on the treatment recommendations they received, and as this process was not completely protocolized or assessed, it is unclear how this affected sample selection and treatment allocation. Similarly, it is likely that the groups contained both treatment-resistant patients and patients starting treatment (psychotherapy and/or medication) for the first time. Due to the naturalistic nature of our study, we do not have reliable information on prior treatment history, and the proportion of treatment resistant patients may vary among the groups.

Other limitations include the lack of information on the baseline characteristics of patients who entered the clinic but refused to participate, and the absence of a medication adherence measure.

## 5. Conclusion

Investigating potential baseline predictors and moderators of depression treatment outcome in a naturalistic setting, we found that depressive severity, vitality and perceived physical health improvement were associated with depression across time, and that agoraphobia predicted change in depression over the course of treatment. Physical functioning moderated response to combined treatment (as opposed to psychotherapy alone); patients with high physical functioning fared better in combined treatment than patients with low physical functioning, whereas physical functioning did not predict a differential response in the psychotherapy group. Moreover, the lowest levels of physical functioning actually predicted an increase of depressive symptoms in combined treatment.

These results might aid decision-making in routine practice, although the predictors we found require replication first. Differences in design and statistical approach complicate comparison

with previous studies; prediction studies are especially sensitive to small deviations of this kind. Development of more uniform requirements for the analysis and reporting of prediction studies is therefore needed.

### Role of funding source

This study was funded by the Department of Clinical Psychological Science, Maastricht University.

### Conflict of interest

John C. Markowitz is supported in part by grant MH079078 from the National Institute of Mental Health. Steven D. Hollon is supported in part by grant MH01697. Other authors: no conflict of interest.

### Acknowledgments

John C. Markowitz is supported in part by grant MH079078 from the National Institute of Mental Health. Steven D. Hollon is supported in part by grant MH01697.

## References

- Addis, M.E., Jacobson, N.S., 1995. Why do people think they are depressed? The reasons for depression questionnaire. *Psychotherapy* 32, 476–483.
- Anderson, I.M., 2001. Meta-analytical studies on new antidepressants. *British Medical Bulletin* 57, 161–178.
- Angst, J., Gamma, A., Rossler, W., Ajdacic, V., Klein, D.N., 2009. Long-term depression versus episodic major depression: results from the prospective Zurich study of a community sample. *Journal of Affective Disorders* 115, 112–121.
- Arrindell, W.A., Ettema, J.H.M., 1981. Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de Symptom Checklist (SCL-90)/ Dimensional structure, reliability and validity of the Dutch version of the Symptom Checklist (SCL-90). *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden* 36, 77–108.
- Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G., 1979. *Cognitive Therapy of Depression*. The Guilford Press, New York.
- Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *Journal of Personality Assessment* 67, 588–597.
- Carter, J.D., Luty, S.E., Mckenzie, J.M., Mulder, R.T., Frampton, C.M., Joyce, P.R., 2011. Patient predictors of response to cognitive behaviour therapy and interpersonal psychotherapy in a randomised clinical trial for depression. *Journal of Affective Disorders* 128, 252–261.
- Chalkidou, K., Tunis, S., Whicher, D., Fowler, R., Zwarenstein, M., 2012. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clinical Trials* 9, 436–446.
- Cuijpers, P., Dekker, J., Hollon, S.D., Andersson, G., 2009a. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. *Journal of Clinical Psychiatry* 70, 1219–1229.
- Cuijpers, P., Van Straten, A., Andersson, G., Van Oppen, P., 2008. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* 76, 909–922.
- Cuijpers, P., Van Straten, A., Warmerdam, L., Andersson, G., 2009b. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Depression and Anxiety* 26, 279–288.
- Derogatis, L.R., Lipman, R.S., Covi, L., 1973. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacological Bulletin* 9, 13–28.
- DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., Shelton, R.C., Young, P.R., Salomon, R.M., O'reardon, J.P., Lovett, M.L., Gladis, M.M., Brown, L.L., Gallop, R., 2005. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry* 62, 409–416.
- Elkin, I., Gibbons, R.D., Shea, M.T., Sotsky, S.M., Watkins, J.T., Pilkonis, P.A., Hedeker, D., 1995. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology* 63, 841–847.
- Elkin, I., Shea, M.T., Watkins, J.T., Imber, S.D., Sotsky, S.M., Collins, J.F., Glass, D.R., Pilkonis, P.A., Leber, W.R., Docherty, J.P., Al, E., 1989. NIMH treatment of depression collaborative research program. General effectiveness of treatments. *Archives of General Psychiatry* 46, 971–982.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1996. *Structured Clinical Interview for DSM-IV Axis I Disorders*. Biometrics Research Department, New York State Psychiatric Institute, New York.
- Fournier, J.C., DeRubeis, R.J., Shelton, R.C., Gallop, R., Amsterdam, J.D., Hollon, S.D., 2008. Antidepressant medications v. cognitive therapy in people with depression with or without personality disorder. *British Journal of Psychiatry* 192, 124–129.
- Fournier, J.C., DeRubeis, R.J., Shelton, R.C., Hollon, S.D., Amsterdam, J.D., Gallop, R., 2009. Prediction of moderate to severe depression. *Journal of Consulting and Clinical Psychology* 77, 775–787.

- Fox, J., 1997. Applied Regression Analysis, Linear Models, and Related Methods. SAGE, Thousand Oaks (CA).
- Gibbons, R.D., Hur, K., Brown, C.H., Davis, J.M., Mann, J.J., 2012. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Archives of General Psychiatry* 69, 572-579.
- Hays, R.D., Sherbourne, C.D., Mazel, R.M., 1993. The RAND 36-Item Health Survey 1.0. *Health Economics* 2, 217-227.
- Hollon, S.D., DeRubeis, R.J., Shelton, R.C., Amsterdam, J.D., Salomon, R.M., O'reardon, J.P., Lovett, M.L., Young, P.R., Haman, K.L., Freeman, B.B., Gallop, R., 2005. Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Archives of General Psychiatry* 62, 417-422.
- Jacobson, N.S., Hollon, S.D., 1996a. Cognitive-behavior therapy versus pharmacotherapy: now that the jury's returned its verdict, it's time to present the rest of the evidence. *Journal of Consulting and Clinical Psychology* 64, 74-80.
- Jacobson, N.S., Hollon, S.D., 1996b. Prospects for future comparisons between drugs and psychotherapy: lessons from the CBT-versus-pharmacotherapy exchange. *Journal of Consulting and Clinical Psychology* 64, 104-108.
- Joyce, P.R., Mckenzie, J.M., Carter, J.D., Rae, A.M., Luty, S.E., Frampton, C.M., Mulder, R.T., 2007. Temperament, character and personality disorders as predictors of response to interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry* 190, 503-508.
- Kazdin, A.E., 2007. Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology* 3, 1-27.
- Klerman, G., Weissman, M., Rounsaville, B., Chevron, E., 1984. *Interpersonal Psychotherapy of Depression*. Basic Books, New York.
- Kraemer, H.C., Wilson, G.T., Fairburn, C.G., Agras, W.S., 2002. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry* 59, 877-883.
- Lachin, J.M., 1981. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 2, 93-113.
- Leuchter, A.F., Cook, I.A., Marangell, L.B., Gilmer, W.S., Burgoyne, K.S., Howland, R.H., Trivedi, M.H., Zisook, S., Jain, R., Mccracken, J.T., Fava, M., Iosifescu, D., Greenwald, S., 2009. Comparative effectiveness of biomarkers and clinical indicators for predicting outcomes of SSRI treatment in Major Depressive Disorder: results of the BRITE-MD study. *Psychiatry Research* 169, 124-131.
- Leykin, Y., Amsterdam, J.D., DeRubeis, R.J., Gallop, R., Shelton, R.C., Hollon, S.D., 2007a. Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *Journal of Consulting and Clinical Psychology* 75, 267-276.
- Leykin, Y., DeRubeis, R.J., Gallop, R., Amsterdam, J.D., Shelton, R.C., Hollon, S.D., 2007b. The relation of patients' treatment preferences to outcome in a randomized clinical trial. *Behavior Therapy* 38, 209-217.
- Lin, E.H., Tang, L., Katon, W., Hegel, M.T., Sullivan, M.D., Unutzer, J., 2006. Arthritis pain and disability: response to collaborative depression care. *General Hospital Psychiatry* 28, 482-486.
- Luty, S.E., Carter, J.D., Mckenzie, J.M., Rae, A.M., Frampton, C.M., Mulder, R.T., Joyce, P.R., 2007. Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry* 190, 496-502.
- McHugh, R.K., Barlow, D.H., 2010. The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. *American Psychologist* 65, 73-84.
- Meyer, T.J., Miller, M.L., Metzger, R.L., Borkovec, T.D., 1990. Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy* 28, 487-495.
- Narum, S.R., 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics* 7, 783-787.
- Nilssen, T.S., Eisemann, M., Kvernmo, S., 2012. Predictors and moderators of outcome in child and adolescent anxiety and depression: a systematic review of psychological treatment studies. *European Child and Adolescent Psychiatry*.
- Nolen-Hoeksema, S., 1991. Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology* 100, 569-582.
- Papageorgiou, C., Wells, A., 2001a. Metacognitive beliefs about rumination in recurrent major depression. *Cognitive and Behavioral Practice* 8, 160-164.
- Papageorgiou, C., Wells, A., 2001b. Positive beliefs about depressive rumination: development and preliminary validation of a self-report scale. *Behavior Therapy* 32, 13-26.
- Papakostas, G.I., Fava, M., 2008. Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. *Dialogues in Clinical Neuroscience* 10, 439-451.
- Peeters, F., Huibers, M., Roelofs, J., Van Breukelen, G., Hollon, S.D., Markowitz, J.C., Van Os, J., Arntz, A., 2012. The clinical effectiveness of evidence-based interventions for depression: a pragmatic trial in routine practice. *Journal of Affective Disorders* 145, 349-355.
- Rhebergen, D., Lamers, F., Spijker, J., de Graaf, R., Beekman, A.T., Penninx, B.W., 2012. Course trajectories of unipolar depressive disorders identified by latent class growth analysis. *Psychological Medicine* 42, 1383-1396.
- Roelofs, J., Huibers, M., Peeters, F., Arntz, A., Van Os, J., 2008. Rumination and worrying as possible mediators in the relation between neuroticism and symptoms of depression and anxiety in clinically depressed individuals. *Behaviour Research and Therapy* 46, 1283-1289.
- Roelofs, J., Muris, P., Huibers, M., Peeters, F., Arntz, A., 2006. On the measurement of rumination: a psychometric evaluation of the ruminative response scale and the rumination on sadness scale in undergraduates. *Journal of Behavior Therapy and Experimental Psychiatry* 37, 299-313.
- Roelofs, J., Papageorgiou, C., Gerber, R.D., Huibers, M., Peeters, F., Arntz, A., 2007. On the links between self-discrepancies, rumination, metacognitions, and symptoms of depression in undergraduates. *Behaviour Research and Therapy* 45, 1295-1305.
- Shafran, R., Clark, D.M., Fairburn, C.G., Arntz, A., Barlow, D.H., Ehlers, A., Freeston, M., Garety, P.A., Hollon, S.D., Ost, L.G., Salkovskis, P.M., Williams, J.M., Wilson, G.T., 2009. Mind the gap: improving the dissemination of CBT. *Behaviour Research and Therapy* 47, 902-909.
- Sotsky, S.M., Glass, D.R., Shea, M.T., Pilkonis, P.A., Collins, J.F., Elkin, I., Watkins, J.T., Imber, S.D., Leber, W.R., Moyer, J., et al., 1991. Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry* 148, 997-1008.
- Van Breukelen, G.J., 2006. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. *Journal of Clinical Epidemiology* 59, 920-925.
- Westen, D., Novotny, C.M., Thompson-Brenner, H., 2004. The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin* 130, 631-663.
- Zimmerman, M., Chelminski, I., Posternak, M.A., 2005. Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. *American Journal of Psychiatry* 162, 1370-1372.