

# A Propensity Score Analysis of Homework Adherence-Outcome Relations in Cognitive Behavioral Therapy for Depression

**Judith A. Callan**

University of Pittsburgh

**Nikolaos Kazantzis**

Monash University

**Seo Young Park**

**Charity G. Moore**

University of Pittsburgh

**Michael E. Thase**

University of Pennsylvania

**Abu Minhajuddin**

UT Southwestern Medical Center

**Sander Kornblith**

Allegheny Mental Health Associates, Pittsburgh

**Greg J. Siegle**

University of Pittsburgh

The authors would like to thank Chetachi Emeremni, Ph.D., for statistical support in the early stage of this manuscript. This research was supported by the Sequenced Treatment Alternatives to Relieve Depression (STAR-D) study ([ClinicalTrials.gov](https://clinicaltrials.gov): NCT 00021528; Rush et al., 2004); the Continuation Phase Cognitive Therapy Relapse Prevention (C-CT-RP) trial ([ClinicalTrials.gov](https://clinicaltrials.gov): NCT00118404, NCT00183664, and NCT00218764; Jarrett & Thase, 2010) and by the National Center For Advancing Translational Sciences of the NIH under Award Number KL2TR000146.

The University of Pittsburgh, Institutional Review Board approved our secondary data analysis of Sequenced Treatment Alternatives to Relieve Depression (STAR-D) and the Continuation Phase Cognitive Therapy Relapse Prevention (C-CT-RP).

Address correspondence to Judith A. Callan, Ph.D., R.N., Health and Community Systems, University of Pittsburgh, 3500 Victoria St., Pittsburgh, PA. 15261; e-mail: [callanja@pitt.edu](mailto:callanja@pitt.edu)

---

Little is known about whether or not a consistently high level of homework adherence over the course of therapy benefits patients. This question was examined in two samples of patients who were receiving individual Cognitive Behavioral Therapy (CBT) for depression ( $N_s = 128$  [Sequenced Treatment Alternatives to Relieve Depression: STAR-D] and 183 [Continuation Phase Cognitive Therapy Relapse Prevention: C-CT-RP]). Logistic and linear regression and propensity score models were used to identify whether or not clinician assessments of homework adherence differentiated symptom reduction and remission, as assessed by the Hamilton Depression Rating Scale-17 (HDRS-17), the Quick Inventory of Depressive Symptomatology–Self-Reported Scale (QIDS-SR), and the QIDS–Clinician Scale (QIDS-C). CBT-related response and remission were equally likely between both high and low homework adherers in both studies and in all models. But in propensity adjusted models

that adjusted for session attendance, for both the STAR-D and C-CT-RP samples, greater homework adherence was significantly associated with greater response and remission from depression in the first and last 8 sessions of CBT. Our results suggest that homework adherence can account for response and remission early and late in treatment, with adequate session attendance.

---

*Keywords:* homework adherence; propensity score analysis; cognitive behavioral therapy; depression; outcome

COGNITIVE BEHAVIORAL THERAPY (CBT), as defined by Beck, Rush, Shaw, and Emery (1979), is the most thoroughly evaluated psychological treatment for major depression (Cuijpers et al., 2017; Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). However, data suggests that CBT is associated with recovery or remission in only 40%–60% of protocol-treated patients with unipolar major depression (DeRubeis et al., 2005). There is a need for research to improve understanding of process predictors of treatment response in both clinical trials (Kazantzis, 2018; Kazantzis et al., in press; Lorenzo-Luaces & DeRubeis, 2018) and implementation contexts (Clark et al., 2018).

Early response to CBT (i.e., during the first 4 weeks) and the dose of CBT (i.e., number of sessions) are associated with improvements in depression (Hollon et al., 2005; Jarrett et al., 2001) though response across time varies (Busch, Kanter, Landes, & Kohlenberg, 2006; Rush, Kovacs, Beck, Weissenburger, & Hollon, 1981). Sudden reductions in symptom improvement are also observed as occurring within Sessions 6–8 of depression treatment (i.e., stable 25% reduction from the pregain level in Tang & DeRubeis, 1999). However, a large body of research has now examined sudden gains (see review Aderka, Nickerson, Bøe, & Hofmann, 2012), and there is considerable variability in when the gain occurs (Bohn, Aderka, Schreiber, Stangier, & Hofmann, 2013; Wucherpfennig, Rubel, Hollon, & Lutz, 2017). There is also a need to better understand later sudden gains (i.e., Sessions 8–16). Homework adherence has been proposed as a possible mechanism of rapid and sudden symptom changes in CBT (Doane, Feeny, & Zoellner, 2010), but this lacks research support.

Homework is widely accepted as the vehicle of skill generalization and maintenance in CBT for depression (Beck et al., 1979; Thase & Callan, 2006) and is important for meta-analytic aggregations of CBT's effects (Spielmans & Flückiger, 2018). Even modest improvements in homework adherence can enhance outcomes for those with

mood and sleep disorders (Dong, Soehner, Bélanger, Morin, & Harvey, 2018; Neimeyer, Kazantzis, Kessler, Baker, & Fletcher, 2008). Meta-analyses of homework adherence-outcome relations have estimated small effects ( $r = .22$  in Kazantzis, Deane, & Ronan, 2000;  $r = .26$  in Mausbach, Moore, Roesch, Cardenas, & Patterson, 2010), whereas trials contrasting therapy conditions with and without homework have shown medium to large effects ( $d = .77$  in Kazantzis et al., 2000;  $d = .63$  in Kazantzis, Whittington, & Dattilio, 2010). This suggests that the therapist's decision to include homework in therapy may be more strongly related to outcome than the patient's level of homework adherence. When taken together with unsuccessful attempts to enhance homework adherence through treatment augmentation, the case for further study of particularly high vs. low levels of adherence is even more compelling.

In the majority of studies, homework adherence has been studied as a predictor of overall outcome (i.e., symptom change from pre- to posttreatment) without considering its relationship with symptom reduction or remission occurring early or later in therapy. Fennell and Teasdale (1987) found preliminary evidence that patients ( $N = 34$ ) experienced major reductions in depressive symptoms following Session 3 when there was a change in homework assignments (i.e., from bibliotherapy to behavioral activities). Session attendance and homework adherence are both indicators of treatment engagement, but we are aware of only two prior studies examining these variables within the same investigation. Burns and Nolen-Hoeksema (1992) reported that lower levels of homework adherence were evident among dropouts receiving CBT for depression ( $N = 185$ ). More recently, Glenn et al. (2013) found that session attendance and patient engagement (homework adherence and commitment) were stable and robust predictors of greater reductions in symptoms. Thus, it remains unclear whether homework adherence is a marker of early or later treatment response for those who remain engaged in treatment.

The present study extends previous research on homework in CBT for depression. We considered the role of patient homework adherence profiles (i.e., high adherers vs. low adherers) using propensity score analyses. We were particularly interested in whether homework adherence could differentiate large reductions in symptoms (response and remission) from major depressive disorder throughout the course of therapy. The evidence shows that sudden gains and rapid response can occur during the first 8 sessions, or in the last 8 sessions, of a standard 16–20 session protocol (Aderka et al.,

2012). Thus, the present study sought to determine whether high vs. low homework adherence was significantly associated with an individual's treatment response in the first or last 8 sessions. We adjusted for session attendance in all our models because of the findings reported in Glenn et al. (2013).

### Propensity Score Analysis

In the case of CBT, randomly assigning some patients to homework could help determine the role of homework in clinical outcome. While meta-analytic aggregations of such component analyses have found medium to large effect size differences between “homework” and “no homework” conditions (i.e., Kazantzis et al., 2000), the validity of the control conditions in these studies is questionable because of the crucial role given to homework in CBT (Beck et al., 1979; Kazantzis et al., 2010).

Propensity score analysis has emerged as a valid means of evaluating observational data to estimate the effect of nonrandomized treatment factors and the outcomes of interest (Glenn et al., 2013). In these studies, analysts use blocking and matching of clinically relevant covariates other than the covariate of interest (Rubin, 2007) and thereby manage confounding variables that lead to bias in observational trials (Haukoos & Lewis, 2015; Trojano, Pellegrini, Paolicelli, Fuiani, & Renzo, 2009).

Given the potential for a range of patient and patient-therapist factors to influence homework adherence in CBT for depression, Burns and Spangler (2000) tested a range of models involving patient motivation and therapeutic alliance. Burns and Spangler concluded that the relationship between homework adherence and symptom change was likely causal, but this conclusion was problematic because other theoretically important factors were not assessed (Kazantzis, Ronan, & Deane, 2001), such as prior depressive episodes (Lorenzo-Luaces, DeRubeis, & Webb, 2014). A broad spectrum of patient characteristics may be important in evaluating relations involving homework adherence. Propensity score analysis enables an examination of client characteristics within a single analysis, which prevents the need for multiple tests of hypotheses and the potential for inflated Type I error (West et al., 2014).

While it has been proposed that patient factors may influence the homework adherence–outcome relationship (DeRubeis, Gelfand, German, Fournier, & Forand, 2014), no previous study specifically examined this possibility. Previous studies included tests of patient characteristics as potential moderators to explain predictors of treatment response (e.g., Sasso, Strunk, Braun, DeRubeis, & Brotman, 2015;

Webb et al., 2012). These have, however, focused on therapist adherence rather than patient homework adherence. It follows that measurement of homework adherence could provide essential information for estimating CBT effects, and also inform researchers how to better study patient factors that determine it. For example, if research could confirm that homework adherence is involved in the remission of depression during CBT, either during the first 8 sessions or later 8 sessions (i.e., before or after the time a sudden gain may occur), then this would suggest that maximizing adherence would provide a solid basis for the examination of other change processes in CBT (Lorenzo-Luaces, German, & DeRubeis, 2015).

### HYPOTHESES

As the present study reports the first propensity score analyses to estimate homework adherence–outcome relations in CBT from datasets, our two hypotheses both replicate and extend previous findings. First, consistent with the meta-analytic findings of Kazantzis et al. (2010) and Mausbach et al. (2010), overall, higher levels of homework adherence during the entire CBT trial for depression was expected to be associated with higher rates of remission and response in both samples (H1).

Second, based on the findings of Fennell and Teasdale (1987) and Burns and Nolen-Hoeksema (1992), we expected that greater adherence to homework (high vs low) would be associated with whether the individual is a treatment responder or has remission in the first 8 sessions, and then later in the last 8 sessions, when adjusting for session attendance (H2). By focusing on whether homework adherence profiles would differentiate treatment responses and remission, we are addressing an identified gap in the literature (see Strunk & Pfeifer, 2015).

### Method

Our propensity analyses draw on two previous trials in which CBT for depression was examined: the Sequenced Treatment Alternatives to Relieve Depression (STAR-D) study (NCT00021528—Rush et al., 2004) and the Continuation Phase Cognitive Therapy Relapse Prevention (C-CT-RP) trial (NCT00118404, NCT00183664, & NCT00218764—Jarrett & Thase, 2010). Data from two trials are included for replication purposes, and to ensure a robust examination of study hypotheses (see also Webb et al., 2012).

#### STUDY I: THE STAR-D STUDY

The rationale and design of the STAR-D study are described elsewhere (Fava et al., 2003; Rush et al.,

2004). Briefly, the STAR-D study featured a progressive, equipoise and stratified, randomization design to assign treatment to a sample of 4,041 outpatients with nonpsychotic major depression. Patients were 18–75 years old and scored 14 or greater on the 17-item Hamilton Depression Rating Scale (HDRS-17; Hamilton, 1960). Exclusion criteria were pregnancy, medical conditions that precluded study medications, hospitalization for substance detoxification, and/or diagnosis of schizophrenia, schizoaffective disorder, bipolar disorder, anorexia, bulimia, or primary obsessive-compulsive disorder.

If depression did not remit with antidepressant therapy in the first stage of the study, patients were offered other treatment possibilities. One of the second-stage treatments was 16 to 20 sessions of CBT, either alone or in combination with antidepressant therapy. Of 147 patients given CBT, complete data on homework adherence was available for 128 patients (see supplement Figure 1 for Consort diagram).

#### STUDY 2: THE C-CT-RP TRIAL

The rationale and design of the C-CT-RP trial are described in detail elsewhere (Jarrett & Thase, 2010). The C-CT-RP sample consisted of 523 outpatients who: (a) had recurrent major depressive disorder with remission between episodes; and (b) had one prior episode of depression with complete inter-episode recovery or had antecedent dysthymic disorder. Patients ranged from 18 to 70 years of age and scored 14 or higher on the HDRS-17 at screening and baseline interviews. Exclusion criteria were pregnancy, poorly controlled medical disorder, significant Axis I psychiatric disorder (e.g., schizophrenia or bipolar disorder), active substance abuse or dependence, and active suicidal risk.

The patients in the trial were studied prospectively during procedurally determined CBT (Beck et al., 1979; see Additional Methodology in Supplement), provided for 12–14 weeks during the acute phase of the trial. Participants received CBT twice a week for 4 weeks. If participants had not experienced a 40% reduction in symptoms, they continued to receive CBT treatment twice a week for 4 additional weeks, followed by once a week during the final month. Otherwise, CBT treatment was continued weekly for 8 more weeks. The 12–14 week acute-phase sessions were followed by a continuation phase, which was not analyzed in our study. Of 187 patients in the acute-phase CBT from one study site, complete data on homework adherence was available for 183 (98%) participants (see Figure 2 Consort diagram in supplement).

#### MEASUREMENTS

##### *Depression Symptoms*

Clinical outcomes were measured through self-report and clinician interview in both studies. Self-reported depressive severity was measured by the Quick Inventory of Depressive Symptomatology–Self-Reported Scale (QIDS-SR; Rush et al., 2003). Clinician-administered measures of depression severity were made with the HDRS-17 and the Quick Inventory of Depressive Symptomatology (QIDS-C; Rush et al., 2003) (STAR-D study only). Extensive psychometric and descriptive analyses for the instruments implemented in the two studies have been previously reported (Jarrett & Thase, 2010; Rush et al., 2004). These instruments are commonly used and all demonstrate acceptable reliability and validity. Some measures vary between the two studies. The analyses are presented as two separate studies. This prevented loss of information through data merging, and allowed for the examination of different covariate domains.

Two types of treatment response variables were defined for each of the clinical outcome measures (i.e., HDRS-17, QIDS-C, and QIDS-SR). Response was defined as 50% reduction from baseline scores in each of the measures. A patient was considered to have remission if (a) HDRS-17 was less than or equal to 7, and (b) QIDS-C/QIDS-SR was less than or equal to 5 (Thase et al., 2007).

##### *Homework Adherence*

A therapist-reported assessment of homework adherence was adopted (The Therapist Compliance Checklist; Kornblith, 2000), where patients' weekly adherence was evaluated on a 4-point Likert scale with specific scale anchors: 1 (*not done*), 2 (*did some of what was planned*), 3 (*did most of what was planned*), and 4 (*did all of what was planned*). Single item assessments are commonplace in research on homework assignments (see review in Kazantzis, Brownfield, Mosely, Usatoff, & Flighty, 2017). The STAR-D trial used the homework adherence measure ranging from 1 to 4. The C-CT-RP trial used a measure ranging from 0 to 3. Although the two studies used different numerical ranges, the descriptive anchors remained the same in both trials. In the STAR-D study, participants were classified as "high adherers" if the average of all observations was greater than or equal to 3.0; and "low adherers" if their average was less than 3.0. In the C-CT-RP trial, participants were coded as "high adherers" if the average of all observations was greater than or equal to 2.25; and "low adherers" for scores less than 2.25. Missing values were excluded from calculating the average adherence score.

Cutoffs varied because of different numerical labeling of the anchors. For ease of comparison, we sought to make the adherence measures identical. Unfortunately, the transformation was not monotonic, so it was abandoned.

#### STATISTICAL ANALYSES

The estimation of homework adherence–outcome relations is potentially confounded by patient characteristics associated with being either a high or low adherer. We addressed this by utilizing propensity score analysis. We first estimated propensity scores, defined as the probability of high homework adherence as a function of baseline covariates (Rosenbaum & Rubin, 1983), and used weighted regression where observations were weighted based on estimated propensity scores. This weighting method does not require decisions on the number of strata and strata boundaries; therefore, bias due to inexact stratification of the propensity score is eliminated (Lunceford & Davidian, 2004). High homework adherence participants received a weight equal to the inverse of the propensity score (ps, i.e.,  $1/ps$ ). Low homework adherence participants received a weight equal to the inverse of 1 minus their estimated propensity score (i.e.,  $1/[1 - ps]$ ; see Robins, Hernan, & Brumback, 2000, for more on this method).

To determine which baseline characteristics might impact homework adherence, the covariates from the STAR-D and C-CT-RP datasets having either clinical relevance or having been shown to affect clinical outcomes in past studies were considered (see Supplement Table 1 and 2). In the STAR-D study, the variables included were the Work and Social Adjustment Scale (WSAS) score, the Quality of Life Enjoyment and Satisfaction Scale (QLESQ) score, the HDRS-17 baseline score, and the QIDS-C score. In the C-CT-RP trial, the variables included were the HDRS-17 baseline score, the QIDS-SR baseline score, the Quality of Life (QOL) overall score, the Social Adjustment Scale (SOC) total score, and the Attributional Style Questionnaire (ASQ) failure global score. If a baseline variable was associated with homework adherence at the level of  $p < .50$ , it was included in a logistic regression model to estimate the propensity score. While homework adherence is a continuous variable, a logistic regression model is typically used for the propensity score model. Specifically, the propensity scores reflect each participant's conditional probability of being exposed to the treatment (in this case of being a high homework adherer or a low homework adherer) given baseline characteristics and is, thus, a dichotomous procedure (Trojano et al., 2009). Unlike other common models,

propensity score models can be built with less restrictive inclusion criteria without danger of overparametrizing (Brookhart et al., 2006). We used  $p < .50$  as the entry criterion because this would both maximize the number of variables having association with high adherence, but also be a manageable number—given the smaller sample size and limited number of high adherers. Since  $p$ -value is the probability of Type I error, by including the variables with  $p < .50$ , we are including the observed variables with any signal that might be predictive of homework adherence.

Adequate propensity score overlap between the two groups is sought and means that the distributions of the propensity score, when visually inspected, are not excessively different. If the estimated propensity score is too different, unstable estimates of the treatment effect are generated (Imai, King, & Stuart, 2008). Finally, to summarize the measures of association of predicted probabilities and observed responses, eight measures (i.e., percent concordant, percent discordant, percent tied, pairs, Somers' D, gamma, tau-a, and C) were calculated for both study samples.

Response and remission were investigated using logistic regression models with homework adherence as the main independent variable. The odds of a positive clinical outcome for high versus low homework adherers was calculated three times: first by the propensity-based weights, second by covariate-adjusting for the number of CBT sessions attended, and finally by both the propensity-based weights and adjusting for the number of CBT sessions attended. This robust 3-step process provides a more stringent estimate that safeguards against the bias that high homework adherence is simply associated with high levels of overall treatment adherence.

The change in scores for the HDRS-17, QIDS-C, and QIDS-SR measures was calculated by subtracting the posttreatment score from the pretreatment score. Therefore, negative changes indicate alleviation of depression symptoms. For each study, a linear regression model, controlling for baseline mood ratings, with propensity-based weights was used to assess differences in the mean of the outcomes for high versus low homework adherers.

Multiple imputation ( $m = 5$  imputations) was used to process nonresponse for any of the outcomes. To perform a sensitivity analysis of each study, all analyses were repeated using only participants who attended the first 8 CBT sessions and the last 8 sessions. Specifically, only adherence scores corresponding to the timeframe of the analyses were used (e.g., for the analyses focused on response/remission during the first 8 session, average adherence ratings

during the first 8 sessions were examined). All analyses were run using SAS version 9.2 (SAS Institute Inc., 2008).

## Results

### STUDY SAMPLES

The STAR-D sample consisted of 128 participants who were offered 16 to 20 sessions of CBT. Most

participants were middle-aged, female, White, and college educated (i.e., 14 to 15 years of education or more). A nearly equal number were (a) employed versus unemployed/retired and (b) married versus single/divorced/widowed (Table 1). The majority ( $n = 107$ ) of the sample had experienced recurrent depression. The mean numbers of sessions attended and standard deviation ( $SD$ ) were 10.67 ( $SD = 4.26$ )

Table 1  
STAR-D Study: Patient Characteristics Predicting Homework Adherence

Characteristic	n (%) All cases (N = 128)	n (%) Homework adherence		p-value	OR (95% CI) Homework adherence	
		Low (n = 69)	High (n = 59)		Unadjusted	Adjusted
Age, years, mean (SD)	42.2 (12.7)	41.3 (11.5)	43.1 (13.9)	0.41	1.01 (0.98–1.04)	1.00 (0.97–1.04)
Gender				0.25		
Female	80 (62.5)	40 (58.0)	40 (67.8)		1.53 (0.74–3.15)	1.11 (0.49–2.52)
Male	48 (37.5)	29 (42.0)	19 (32.2)		Reference	Reference
Race				0.41		
White	101 (79.5)	52 (75.4)	49 (84.5)		Reference	Reference
Black	16 (12.6)	11 (16.0)	5 (8.6)		0.48 (0.16–1.49)	0.69 (0.17–2.79)
Other	10 (7.9)	6 (8.7)	4 (6.9)		0.71 (0.19–2.66)	0.64 (0.13–3.09)
Hispanic ethnicity	13 (10.2)	4 (5.8)	9 (15.3)	0.08	2.92 (0.85–10.05)	1.01 (0.24–4.22)
Employment status				0.47		
Employed	70 (54.7)	39 (56.5)	31 (52.5)		0.98 (0.45–2.07)	1.02 (0.43–2.45)
Unemployed	47 (36.7)	26 (37.7)	21 (35.6)		Reference	Reference
Retired	11 (8.6)	4 (5.8)	7 (11.9)		2.17 (0.56–8.41)	1.06 (0.23–4.82)
Health insurance type				0.48		
Private	67 (52.3)	38 (55.1)	29 (49.2)		0.53 (0.15–1.85)	0.82 (0.22–3.00)
Public	13 (10.2)	5 (7.3)	8 (13.6)		Reference	Reference
None	48 (37.5)	26 (37.7)	22 (37.3)		0.48 (0.14–1.61)	0.79 (0.20–3.15)
Marital status				0.77		
Single	37 (28.9)	21 (30.4)	16 (27.1)		Reference	Reference
Married/cohabiting	57 (44.5)	31 (44.9)	26 (44.1)		1.10 (0.48–2.53)	0.92 (0.35–2.44)
Divorced/separated	28 (21.9)	13 (18.8)	15 (25.4)		1.50 (0.57–4.06)	1.31 (0.42–4.08)
Widowed	6 (4.7)	4 (5.8)	2 (3.4)		0.66 (0.11–4.04)	0.70 (0.09–5.68)
Recurrent depression	107 (83.6)	57 (82.6)	50 (84.8)	0.75	1.17 (0.46–3.01)	0.96 (0.31–2.96)
Education, years, mean (SD)	14.5 (3.4)	14.6 (3.1)	14.5 (3.7)	0.90	0.99 (0.90–1.10)	1.02 (0.90–1.16)
Total persons in household, mean (SD)	2.6 (1.5)	2.8 (1.5)	2.5 (1.4)	0.15	0.83 (0.65–1.07)	0.98 (0.74–1.30)
Number of depression episodes, mean (SD)	7.5 (14.1)	7.8 (14.3)	7.3 (14.1)	0.86	1.00 (0.97–1.03)	1.00 (0.97–1.03)
Duration of current depression episode, months, mean (SD)	25.2 (42.0)	23.9(36.0)	26.8(46.5)	0.70	1.00 (0.99–1.01)	1.00 (0.99–1.01)
CIRS number of comorbidities, mean (SD)	3.3 (2.2)	3.2 (2.3)	3.3 (2.1)	0.82	1.02 (0.87–1.19)	1.02 (0.85–1.22)
CIRS severity of comorbidities, mean (SD)	4.2 (3.1)	4.1 (3.3)	4.4 (2.9)	0.63	1.03 (0.92–1.15)	1.02 (0.89–1.16)
Axis I, number of diagnoses, mean (SD)	0.3 (0.7)	0.4 (0.8)	0.2 (0.5)	0.15	0.66 (0.37–1.17)	0.83 (0.44–1.59)
Axis II, number of diagnoses, mean (SD)	0.02 (0.1)	0.02 (0.1)	0.02 (0.1)	0.98	1.04 (0.06–7.02)	1.20 (0.07–20.33)
PCS12 score, mean (SD)	48.7 (11.9)	48.0 (11.8)	49.4 (12.1)	0.51	1.01 (0.98–1.04)	1.00 (0.97–1.04)
MCS12 score, mean (SD)	25.4 (7.8)	25.3 (6.8)	25.4 (8.9)	0.92	1.00 (0.96–1.05)	0.99 (0.94–1.05)
WSAS score, mean (SD)	26.3 (8.1)	27.3 (7.9)	25.3 (8.2)	0.17	0.97 (0.93–1.01)	1.00 (0.95–1.05)
QLESQ score, mean (SD)	37.7 (12.8)	36.0 (12.7)	39.7 (12.6)	0.10	1.03 (1.00–1.05)	1.00 (0.97–1.04)
HDRS-17 score, mean (SD)	22.8 (5.0)	23.5 (5.2)	22.0 (4.8)	0.10	0.94 (0.88–1.01)	1.00 (0.92–1.08)
QIDS-C score, mean (SD)	16.4 (3.2)	16.8 (3.1)	15.8 (3.3)	0.09	0.91 (0.81–1.02)	1.00 (0.88–1.13)
QIDS-SR score, mean (SD)	15.9 (4.0)	16.0 (4.4)	15.9 (3.7)	0.89	0.99 (0.91–1.08)	1.01 (0.91–1.11)

Note. Abbreviations: CI = confidence interval; CIRS = Cumulative Illness Rating Scale; HDRS-17 = 17-item Hamilton Depression Rating Scale; MCS12 = Mental Components Summary Scale Score (SF-12); OR = odds ratio; PCS12 = Physical Components Summary Scale Score (SF-12); QIDS-C = Quick Inventory of Depressive Symptomatology–Clinician Scale; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Reported Scale; QLESQ = Quality of Life Enjoyment and Satisfaction Scale; SD = standard deviation; STAR-D = Sequenced Treatment Alternatives to Relieve Depression; WSAS = Work and Social Adjustment Scale.

for the total sample, 10.66 ( $SD = 4.49$ ) for high adherers, and 10.68 ( $SD = 4.49$ ) for low adherers. Nearly 20% of the sample missed more than nine sessions. Of the 128 participants, 69 (54%) were low adherers to homework, and 59 (46%) were high adherers. Moreover, a histogram of the propensity scores for these two adherence groups showed adequate overlap (Figure 1). Adequate overlap and, thereby, balance, are manifest by similar distributions. On the 1–4 scale used in this sample, the mean scores and  $SD$  for homework completion were 2.79 ( $SD = 0.71$ ) for the total sample, 3.40 ( $SD = 0.33$ ) for the high adherers, and 2.26 ( $SD = 0.49$ ) for the low adherers.

The C-CT-RP sample consisted of 183 participants who had completed 16 to 20 CBT sessions for 12–14 weeks. Most participants were middle-aged, female, White, college educated, and employed. More were single/divorced/widowed rather than married (Table 2). The mean numbers and  $SD$  of sessions attended were 17.08 ( $SD = 2.75$ ) for the total sample, 16.90 ( $SD = 2.70$ ) for high adherers, and 17.26 ( $SD = 2.76$ ) for low adherers. Of the 183 participants, 94 (51%) were low adherers to homework, and 89 (49%) were high adherers. On the 0–3 scale used in this sample, the mean scores and  $SD$  for homework completion were 2.12 ( $SD = 0.56$ ) for the total sample, 2.59 ( $SD = 0.22$ ) for the high adherers, and 1.68 ( $SD = 0.40$ ) for the low adherers.

PROPSENSITY SCORE ADJUSTMENT

For the STAR-D analyses, the propensity models included the following demographic indicators: age, gender, race, Hispanic ethnicity, employment status, health insurance type, number of persons in

the household, Q-LESQ score, WSAS score, HDRS-17 baseline score, and additional Axis I disorders. For the C-CT-RP analyses, the propensity models included age, gender, race, age at disease onset, education, HDRS-17 baseline score, length of current depression episode, QIDS-SR baseline score, QOL overall score, SOC total score, and ASQ failure global score. After propensity adjustment, there was a marked improvement in the balance between the high and low homework adherence groups, with odds ratios (OR) closer to 1.00 (null), and 95% confidence intervals (CI) capturing the null value (see Tables 1 and 2).

The C statistic was used to test the discrimination of the propensity score models in the STAR-D study and the C-CT-RP trial. In binary outcomes, the C statistic is a measure of concordance and identical to the area under the receiver operating characteristic (ROC) curve (Hanley & McNeil, 1982). Sensible models range from 0.5 to 1.0 (Westreich, Cole, Funk, Brookhart, & Sturmer, 2011). When there is no discriminating power at all, C statistic becomes 0.5. C statistic = 1.0 represents the perfect discrimination. Given our C statistic of 0.76 in both propensity models, the models showed reasonable discrimination between the high and low homework adherers. Additional measures of the discrimination of the model (e.g. percent concordant, percent discordant, percent tied, pairs, Somers’ D, gamma, tau-a are acceptable (see Table 3 in the supplement).

CBT TRIALS CLINICAL OUTCOME AND LEVELS OF CBT HOMEWORK ADHERENCE

Across the two CBT trials, when *response* was examined, the HDRS-17 and QIDS-SR scores in

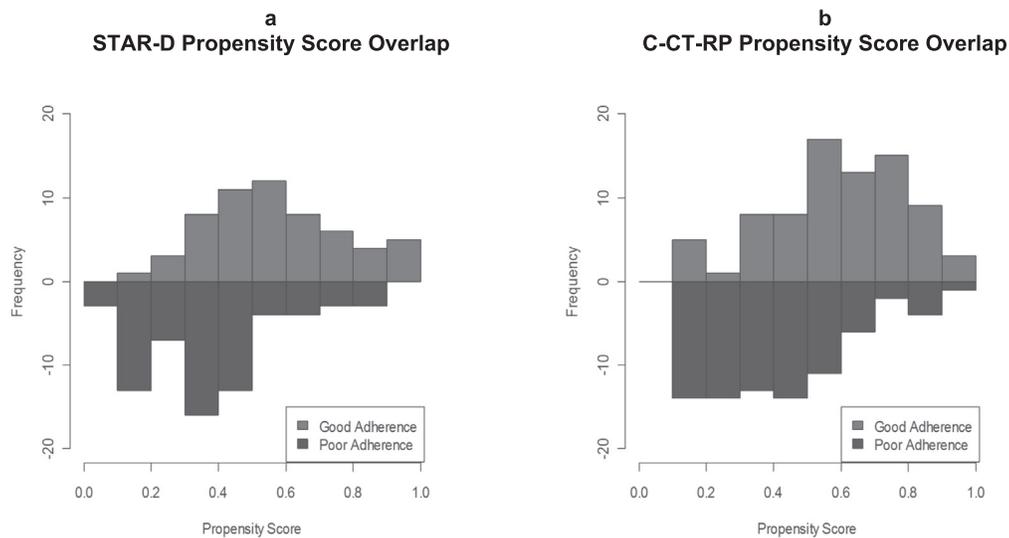


FIGURE 1 Histograms of propensity scores for the Sequenced Treatment Alternative to Relieve Depression (STAR-D) study and the Continuation Phase Cognitive Therapy Relapse Prevention (C-CT-RP) trial.

Table 2  
C-CT-RP Trial: Patient Characteristics Predicting Homework Adherence

Characteristic	n (%) All cases (N = 183)	n (%)		p-value	OR (95% CI)	
		Homework adherence			Homework adherence	
		Low (n = 94)	High (n = 89)		Unadjusted	Adjusted
Age, years, mean (SD)	43.33 (11.77)	42.34 (11.02)	44.38 (12.49)	0.24	1.02 (0.99–1.04)	1.001 (0.97–1.03)
Gender				0.008		
Female	116 (63.39)	43 (45.74)	24 (26.97)		Reference	Reference
Male	67 (36.61)	51 (54.26)	65 (73.03)		0.43 (0.23–0.80)	1.15 (0.53–2.53)
Race/ethnicity				0.42		
White	170 (92.9)	87 (92.55)	83 (93.26)		Reference	Reference
Black/African American	10 (5.46)	4 (4.26)	6 (6.74)		1.57 (0.43–5.71)	1.33 (0.25–7.12)
Asian/Pacific Islander	1 (0.55)	1 (1.06)	0 (0.00)		—	—
Hispanic	—	—	—		—	—
Native American/Alaskan	—	—	—		—	—
Other	2 (1.09)	2 (2.13)	0 (0.00)		—	—
Marital status				0.85		
Single	62 (33.88)	32 (34.04)	30 (33.71)		Reference	Reference
Married	59 (32.24)	30 (31.91)	29 (32.58)		1.03 (0.50–2.12)	0.911 (0.36–2.34)
Separated	10 (5.46)	7 (7.45)	3 (3.37)		0.44 (0.11–1.88)	0.70 (0.113–4.32)
Widowed	5 (2.73)	2 (2.13)	3 (3.37)		1.55 (0.24–9.94)	1.34 (0.19–9.32)
Divorced	37 (20.22)	19 (20.21)	18 (20.22)		1.03 (0.45–2.36)	1.61 (0.60–4.33)
Living together	10 (5.46)	4 (4.26)	6 (6.64)		1.55 (0.40–6.05)	6.11 (1.03–36.35)
Age at disease onset, mean (SD)	23.17 (11.54)	21.11 (9.73)	25.35 (12.89)	0.01	1.03 (1.01–1.06)	0.99 (0.96–1.02)
Education, years, mean (SD)	15.07 (3.02)	14.56 (2.71)	15.61 (3.24)	0.02	1.14 (1.02–1.27)	1.01 (0.89–1.15)
Employment				0.71		
Full-time	84 (45.9)	40 (42.55)	44 (49.44)		1.35(0.65–2.82)	1.21 (0.48–3.08)
Part-time	23 (12.57)	12 (12.77)	11 (12.36)		1.10 (0.40–3.02)	0.52 (0.14–1.94)
Home care	12 (6.56)	8 (8.51)	4 (4.49)		0.60 (0.16–2.29)	0.51 (0.11–2.36)
Student	9 (4.92)	6 (6.38)	3 (3.37)		0.90 (0.18–4.50)	0.21 (0.02–2.25)
Retired	4 (2.19)	1 (1.06)	3 (3.37)		3.60 (0.35–37.36)	1.80 (0.14–23.83)
Unemployed	44 (24.04)	24 (25.53)	20 (22.47)		Reference	Reference
Other	7 (3.83)	3 (3.19)	4 (4.49)		1.60 (0.32–8.01)	3.17 (0.43–23.68)
HDRS-17 baseline score, mean (SD)	21.34 (3.99)	21.06 (4.03)	21.63 (3.95)	0.34	1.04 (0.96–1.12)	0.98 (0.89–1.07)
Length current depression episode in months, mean (SD)	15.18 (24.33)	16.90 (25.10)	13.36 (23.48)	0.33	0.99 (0.98–1.01)	1.00 (0.99–1.02)
Number of previous depression episodes, mean (SD)	18.11 (32.89)	19.20 (34.25)	16.97 (31.54)	0.65	1.00 (0.99–1.01)	1.01 (1.00–1.02)
QIDS-SR baseline score, mean (SD)	14.64 (4.57)	14.91 (4.76)	14.34 (4.37)	0.40	0.97 (0.91–1.04)	0.99 (0.92–1.08)
BHS total score, mean (SD)	11.92 (5.45)	12.05 (5.49)	11.78 (5.43)	0.75	0.99 (0.93–1.05)	1.04 (0.97–1.11)
Q-LES-Q overall score, mean (SD)	2.33 (0.93)	2.24 (0.91)	2.42 (0.96)	0.23	1.23 (0.88–1.70)	0.99 (0.66–1.47)
SCS total score, mean (SD)	- 5.21 (25.14)	- 7.08 (26.63)	- 3.26 (23.45)	0.33	1.01 (0.99–1.02)	10.00 (0.98–1.01)
SOC total score, mean (SD)	2.51 (0.43)	2.55 (0.44)	2.46 (0.40)	0.14	0.61 (0.30–1.24)	1.06 (0.45–2.52)
ASQ failure stable score, mean (SD)	1.07 (0.93)	1.10 (0.99)	1.03 (0.87)	0.63	0.92 (0.66–1.28)	1.10 (0.73–1.64)
ASQ failure global score, mean (SD)	0.99 (1.08)	1.06 (1.07)	0.92 (1.08)	0.37	0.88 (0.66–1.17)	1.09 (0.75–1.57)
DAS total score, mean (SD)	145.38 (35.81)	145.1 (37.1)	145.6 (34.5)	0.93	1.00 (0.99–1.01)	1.01 (1.00–1.02)
DYS total score, mean (SD)	86.98 (23.54)	85.43 (23.65)	88.79 (23.53)	0.47	1.01 (0.99–1.03)	1.00 (0.97–1.02)
IIP score, mean (SD)	1.63 (0.56)	1.62 (0.61)	1.64 (0.51)	0.81	1.08 (0.64–1.83)	2.07 (1.07–4.00)

Note. Abbreviations: ASQ = Attributional Style Questionnaire; BHS = Beck Hopelessness Scale; C-CT-RP = Continuation Phase Cognitive Therapy Relapse Prevention; CI = confidence interval; DAS = Dysfunctional Attitudes Scale; DYS = Dyadic Adjustment Scale; HDRS-17 = 17-item Hamilton Depression Rating Scale; IIP = Inventory or Interpersonal Problems; OR = odds ratio; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Reported Scale; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Scale; SCS = Self-Control Schedule; SD = standard deviation; SOC = Social Adjustment Scale.

both samples and the QIDS-C scores in the STAR-D sample showed that a response to CBT was not significantly different among high and low homework

adherers—the C-CT-RP sample did not include QIDS-C scores. Across the entire CBT trials, when remission was examined, the remission patterns were

nearly identical in the two samples (i.e., CBT remission was not significantly different among high and low homework adherers) (Table 3).

In the STAR-D sample, the correlations between homework adherence and clinical outcomes at follow-up were as follows:  $r = -.26, p = .01$  (HDRS-17),  $r = -.21, p = .02$  (QIDS-C), and  $r = -.22, p = .01$  (QIDS-SR). In the C-CT-RP sample, the correlations were as follows:  $r = -.29, p < .001$  (HDRS-17) and  $r = -.21, p = .004$  (QIDS-SR). Because baseline mood ratings had a much stronger correlation with high adherence in the STAR-D sample than in the C-CT-RP sample, we performed sensitivity analyses for the STAR-D sample. Upon removing the baseline mood ratings from the model and then performing a propensity-adjusted analysis for HDRS-17 remission, we found two predictors of homework adherence: gender (OR = 2.66, 95% CI [1.10, 6.45]) and Hispanic ethnicity (OR = 13.49, 95% CI [1.42, 128.54]). We also found that high adherers were more likely than low adherers to achieve remission (OR = 2.00, 95% CI [1.12, 3.56]). In the STAR-D sample, there was a large drop in the HRDS-17 score if adherence was high. However, our analyses revealed that this drop and

other changes in scores for the STAR-D and C-CT-RP samples were not statistically significant (Table 4; H1).

REMISSION AND RESPONSE DURING FIRST 8 SESSIONS

In the C-CT-RP sample, high adherence differentiated the achievement of *response* in patients during the first 8 CBT sessions, as measured by the HDRS-17 in the propensity-adjusted model that included propensity and the number of sessions attended (OR = 2.70, 95% CI [1.01, 7.24],  $p < .05$ ). In addition, high adherence had been associated with the achievement of *remission* during the first 8 sessions (at trend level of  $p = .06$ ), as measured by the QIDS-SR in the propensity adjusted model with number of sessions (OR = 1.97, 95% CI [0.96, 4.05]; H2).

REMISSION AND RESPONSE DURING LAST 8 SESSIONS

In the C-CT-RP sample, high adherence differentiated the achievement of *response* in patients during the last 8 CBT sessions, as measured by the HDRS-17 in the propensity-adjusted model that included

Table 3  
Logistic Regression for Adherence Effects on Response and Remission

	<b>Low Adherence n (%) of response</b>	<b>High Adherence n (%) of response</b>	<b>Unadjusted OR (95% CI) for outcome</b>	<b>Propensity Weighted Regression OR (95% CI) for outcome</b>	<b>Covariate- adjustment for Number of Sessions OR (95% CI) for outcome</b>	<b>Propensity Weighted Regression and Covariate Adjustment for Number of Sessions OR (95% CI) for outcome</b>
<i>STAR-D Trial</i>						
HDRS-17 score response (50% reduction from baseline)	36 (52.2)	31 (52.5)	1.01 (0.49–2.07)	1.05 (0.46–2.43)	1.01 (0.49–2.12)	1.05 (0.46–2.40)
QIDS-C score response (50% reduction from baseline)	22 (31.9)	20 (33.9)	1.08 (0.41–2.79)	1.18 (0.46–3.01)	1.08 (0.41–2.79)	1.17 (0.47–2.91)
QIDS-SR score response (50% reduction from baseline)	36 (52.2)	36 (61.0)	1.43 (0.68-3.02)	1.55 (0.47-2.82)	1.43 (0.68-3.02)	1.55 (0.86-2.80)
HDRS-17 remission (score < 7)	19 (27.5)	27 (45.8)	2.17 (1.02-4.71)	1.91 (0.75-4.85)	2.19 (1.01-4.76)	1.90 (0.75-4.81)
QIDS-C remission (score < 5)	22 (31.9)	24 (40.7)	1.45 (0.60-3.51)	1.64 (0.60-4.53)	1.45 (0.59-3.54)	1.64 (0.60-4.46)
QIDS-SR remission (score < 5)	22 (31.9)	20 (33.9)	1.09 (0.47-2.51)	1.47 (0.55-3.90)	1.09 (0.47-2.54)	1.47 (0.56-3.86)
<i>C-CT-RP Trial</i>						
HDRS-17 response score (50% reduction from baseline)	55 (60.44)	64 (71.91)	1.68 (0.90–3.12)	1.00 (0.47–2.15)	1.64 (0.88–3.08)	0.98 (0.45–2.10)
QIDS-SR response score (50% reduction from baseline)	60 (65.93)	68 (76.40)	1.67 (0.87-3.22)	1.79 (0.76-4.19)	1.64 (0.85-3.16)	1.70 (0.72-4.04)
HDRS-17 remission (score < 7)	32 (35.16)	45 (50.56)	1.89 (1.04-3.43)	1.92 (0.89-4.14)	1.84 (1.00- 3.41)	1.81 (0.82-4.00)
QIDS-SR remission (score < 5)	45 (49.45)	56 (62.92)	1.74 (1.00-3.15)	1.44 (0.67-3.08)	1.69 (0.92-3.09)	1.33 (0.62-2.84)

Note. Abbreviations: CI = confidence interval; HDRS-17 = 17-item Hamilton Depression Rating Scale; OR = odds ratio; QIDS-C = Quick Inventory of Depressive Symptomatology–Clinician Scale; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Reported Scale.

Table 4  
Linear Regression of Change Scores for Low and High Adherence Groups

		<i>Adjusted for Baseline Score</i>		<i>Adjusted for Baseline score and Propensity Weighted</i>		<i>Adjusted for Baseline score and number of sessions</i>		<i>Adjusted for Baseline score, number of sessions and Propensity Weighted</i>	
		<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>
<i>STAR-D Study</i>									
<i>HRDS-17 Score</i>	Mean score at exit (SE)	13.11 (1.03)	11.41 (1.12)	12.97 (1.09)	11.04 (1.72)	13.11 (1.04)	11.40 (1.12)	12.98 (1.11)	11.04 (1.68)
	Mean difference (SE), p-value	-1.70 (1.58), 0.28		-1.93 (2.07), 0.36		-1.71 (1.58), 0.28		-1.95 (2.00), 0.34	
<i>QID-C Score</i>	Mean score at exit (SE)	7.87 (0.66)	7.24 (0.76)	7.83 (0.75)	6.96 (1.00)	7.87 (0.66)	7.24 (0.77)	7.82 (0.76)	6.96 (0.98)
	Mean difference (SE), p-value	-0.63 (1.11), 0.57		-0.87 (1.34), 0.52		-0.63 (1.11), 0.57		-0.86 (1.30), 0.51	
<i>QID-SR Score</i>	Mean score at exit (SE)	8.02 (0.65)	7.14 (0.77)	7.93 (0.73)	6.83 (1.13)	8.01 (0.66)	7.14 (0.77)	7.93 (0.73)	6.83 (1.10)
	Mean difference (SE), p-value	-0.88 (1.05), 0.41		-1.10 (1.43), 0.45		-0.88 (1.05), 0.41		-1.10 (1.38), 0.43	
<i>C-CT-RP-Trial</i>									
<i>HRDS-17 Score</i>	Mean score at exit (SE)	10.48 (0.67)	8.16 (0.65)	9.84 (0.69)	8.99 (0.68)	10.38 (0.65)	8.26 (0.65)	9.68 (0.68)	8.98 (0.68)
	Mean difference (SE), p-value	-2.33 (0.94), 0.01		-0.85 (0.95), 0.38		-2.13 (0.92), 0.02		-0.70 (0.95), 0.46	
<i>QID-SR Score</i>	Mean score at exit (SE)	6.19 (0.46)	5.11 (0.40)	5.86 (0.58)	5.25 (0.50)	6.16 (0.44)	5.14 (0.41)	5.72 (0.57)	5.26 (0.48)
	Mean difference (SE), p-value	-1.08 (0.61), 0.08		-0.61 (0.76), 0.42		-1.02 (0.60), 0.09		-0.46 (0.74), 0.54	

*Note.* Abbreviations: HDRS-17 = 17-item Hamilton Depression Rating Scale; QIDS-C = Quick Inventory of Depressive Symptomatology–Clinician Scale; QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Reported Scale; SE = standard error; STAR-D = Sequenced Treatment Alternatives to Relieve Depression. C-CT-RP = Continuation Phase Cognitive Therapy Relapse Prevention

the number of sessions attended (OR = 2.29, 95% CI [1.08, 4.83],  $p = .03$ ) and as measured by the QIDS-SR (OR = 2.13, 95% CI [1.02, 4.47],  $p < .05$ ) and the QIDS-SR when propensity and number of sessions were analyzed (OR = 3.06, 95% CI [1.23, 7.58],  $p = .02$ ). Moreover, high adherence differentiated the achievement of *remission* in patients in the last 8 CBT sessions, as measured by the HDRS-17 in the propensity-adjusted model adjusted for number of sessions attended (OR = 2.20, 95% CI [1.07, 4.54],  $p = .03$ ).

In the STAR-D sample, high adherence had a trend level relationship with the achievement of *remission* during the last 8 sessions ( $p = .07$  and  $.08$  respectively), as measured by the HDRS-17 and QIDS-SR in the propensity adjusted models with number of sessions (OR = 2.83, 95% CI [0.92, 8.75] and OR = 2.84, 95% CI [0.89, 9.07]).

#### SUPPLEMENTARY ANALYSES

Using a chi-square test, we examined whether dropout (Y/N) from treatment was related to homework adherence (low/high) in the STAR-D study. We could not include the sample from C-CT-RP in this analysis as only acute phase therapy completers were in their final dataset. We defined dropouts as those who attended less than 12 sessions as STAR-D had 16 to 20 sessions in the CBT trial. Given this definition, 37.50% of this sample were dropouts ( $n = 48$ ). The resulting  $p$  value was 0.96. Thus, we did not find evidence indicating high or low homework adherence was related to therapy dropout.

Finally, to examine the role of therapists in clinical outcome and homework adherence, we investigated whether there was a systematic effect of therapists from baseline measures to posttreatment. Using ICC, in the STAR-D study (therapist ID's unavailable in the C-CT-RP dataset) we examined WSAS, QLESQ, QIDS-SR, QIDS-C, and HDRS at baseline and then posttreatment. All ICC values were low (i.e.,  $< 0.1$ ) or negative, suggesting no systematic therapist effect from baseline to post-treatment on these measures.

#### Discussion

Identifying patient factors that predict treatment response can enable the adaptation of empirically supported interventions in clinical practice (Norcross & Lambert, 2014). Here, we work to advance this direction with methodologies capable of examining the complex associations among patient qualities, processes, and outcomes within the same study. Previous analyses of homework adherence–outcome relations have generally not taken patient characteristics into account (see Burns

& Nolen-Hoeksema, 1992, for an exception), thereby missing the opportunity to identify different patterns of response among the samples studied (Kazantzis et al., 2010; Mausbach et al., 2010). In this study, we examined “low” and “high” homework adherers in terms of their individual differences and treatment response/remission within CBT for depression.

Our findings partially supported our hypotheses. With respect to our first hypothesis that homework adherence significantly differentiates patients according to their reduction in depressive symptoms at the end of treatment, our propensity analyses, which accounted for patient characteristics, reveal that overall response and remission were equally likely among high and low adherers in both samples. However, consistent with our second hypothesis that greater homework adherence would be associated with greater response and remission from depression in the first 8 sessions and last 8 sessions, we found that homework adherence did differentiate response and remission. Specifically, models of subsets of participants in the first eight sessions who demonstrated high adherence had significantly greater response (C-CT-RP sample). Models from subsets of participants in the last eight sessions who demonstrated high adherence had significantly greater response and remission (C-CT-RP sample) in propensity-adjusted models that include number of sessions attended. We also found trend level results for high adherence as a differentiator of remission during the first 8 sessions (C-CT-RP sample) and the last 8 sessions (STAR-D).

These findings are the first to show that homework adherence (high vs. low) is significantly associated with treatment response in the first 8 sessions, and then later in the last 8 sessions of CBT for depression, while taking into account various patient characteristics. There is a growing evidence for rapid and sudden symptomatic responses early in CBT (Schibbye et al., 2014; Wucherpfennig et al., 2017), which have been explained as being due to therapeutic work during sessions and factors external to therapy (e.g., alleviation of significant psychosocial stressors; Lorenzo-Luaces et al., 2015). However, the findings from the present study now point to the need for a broadened focus to homework adherence. These findings are consistent with the evidence for homework adherence–outcome relations in depression treatment (Kazantzis et al., 2016) and emerging evidence for links between homework adherence, session attendance, and dropout (Glenn et al., 2013).

The finding that session attendance was nearly identical across those with low adherence and high adherence is noteworthy. In fact, in the C-CT-RP

sample, the low adherence group had a slightly higher session attendance. This is especially intriguing, given that high homework adherence was positively associated with response and remission in nearly all of the analyses of the first and last 8 sessions. Our findings are consistent with Glenn et al. (2013) with regards to homework adherence, but differ from their report that session attendance (i.e., high treatment dose) predicted better symptom and functional outcomes. Our finding also contrasts with a large study of UK mental health services that found number of treatment sessions was positively associated with treatment outcome and number of missed sessions was negatively associated with outcome (Clark et al., 2018). The data presented from the STAR-D and C-CT-RP studies may indicate that the more salient interpretation was that when sessions are held constant, level of homework adherence made the difference in treatment outcome. However, more research is required to fully explore this possibility, including the examination of homework as a moderator within meta-analyses of CBT (Spielmanns & Flückiger, 2018).

We obtained small yet reliable adherence-outcome relations that were remarkably consistent with the broader literature. Specifically, the effect size estimations in this literature using the  $r$  index have generally found adherence-outcome relations of  $r = .22$  (Kazantzis et al., 2010) or  $r = .26$  (Mausbach et al., 2010). Our results suggest that higher homework adherence appears to be a significant variable in the response and remission from depression in the early and later treatment phases of CBT, when adjusted for number of sessions attended.

Although homework adherence was a significant predictor of treatment responses in our study, therapists should not rely solely on homework adherence as the only determinant of change. Our propensity analyses, which took into account a wide range of patient characteristics, revealed that greater homework adherence differentiated patients in terms of both earlier and later treatment responses. Homework adherence has been theorized to depend on other factors, such as the therapists' skill and the patients appraisals of the homework (Addis & Jacobson, 2000). While our study did not find evidence for therapist effects as a supplementary analysis, further assessment of therapist skill is indicated given preliminary findings in the treatment of anxiety (Weck, Richtberg, Esch, Hofling, & Stangier, 2013) and depression (Bryant, Simons, & Thase, 1999).

The current work has several limitations. First, the measures of homework adherence were exclusively rated by therapists. The potential limitation

of relying on a single source of adherence data has been underscored by research highlighting different adherence-outcome relations, when data have been gathered by client, therapist, and independent sources (Kazantzis et al., 2016; Mausbach et al., 2010). Future research working to replicate and advance the current results may benefit from adopting a multi-modal approach to assessing adherence (i.e., using evaluations of logs, including differentiation between quantity and quality of homework, see Neimeyer et al., 2008; client self-report, therapist assessment, and where possible, objective data such as those from technology based completion of therapeutic tasks). In addition, assessment of theoretically meaningful determinants of homework adherence is increasingly commonplace (Hara et al., 2015). Second, as only one form of psychotherapy (CBT) was assessed in the current study, it is unclear whether the current results would extend to other forms of psychotherapy. Specifically, disparate types of CBT may involve a different configuration of homework tasks. Other unmeasured therapeutic processes would also serve as mediators (e.g., alliance; Sasso et al., 2015), and other client attributes, such as prior episodes, may serve as moderators. The trial data also relied on pre-post symptom assessments, which preclude examination of temporal effects. A model predicting week-to-week symptom improvement, rather than averaging adherence ratings across weeks/months of therapy and correlating these ratings with response/remission may provide a more fine-tuned understanding of the process of therapeutic improvement in relation to homework adherence (Braun, Strunk, Sasso & Cooper, 2015; Strunk, Cooper, Ryan, DeRubeis, & Hollon, 2012).

Finally, given our sample sizes, we may have been inadequately powered to attain distributional balance of observed covariates (Luellen, Shadish, & Clark, 2005). Although propensity score researchers have not specified what sample size is adequate for propensity analyses, they typically use large samples to attain distributional balance of observed covariates (Luellen et al., 2005). In our analyses, a moderate sample size was used. Propensity modeling studies generally find many significant predictors of the primary test of hypotheses, but this was not the case in this study, even at alpha levels higher than .50. This may indicate that our propensity models were not comprehensive. Covariates were chosen based on clinical relevancy to outcome and adherence, yet unknown or missing covariates may have reduced the efficiency of the analyses and affected the results.

Limitations aside, this study investigated the propensity of patients who have greater levels of homework adherence to attain desirable treatment

outcomes. We found initial evidence that greater adherence to homework was associated with treatment response (reduced symptoms of depression and remission from depression) in patients within the first and last 8 sessions of CBT, when we adjusted for number of sessions attended. This highlights two important considerations. First, it is clear from both the current results and the existing literature that homework adherence is an important indicator of treatment outcome, although the influence of other process and patient factors requires further study. Second, the results underscore the existence of process and patient factors as not being distinct, but interrelated in the determination of homework-outcome relations.

A more comprehensive assessment of homework adherence and the processes directly influencing adherence, such as therapist skill, is clearly indicated. A recent systematic review identified that the vast majority of studies still depend on self-report and clinician assessment (Kazantzis et al., 2017), and thus the promise of objective assessments pioneered in early behavior therapy research (e.g., Kornblith, Rehm, O'Hara, & Lamparski, 1983) have yet to be fully realized. Possible strategies for therapists to increase patient adherence to homework include a more thorough linking of the cognitive case conceptualization in the stages of homework review (i.e., task-specific appraisals, such as perceived difficulty and obstacles), design (i.e., a rationale for the homework beliefs about problems and coping), and assign (i.e., perceived importance, readiness, and confidence following collaboration of the specifics of the task). Focused strategies have been subject to empirical study (Addis & Jacobson, 2000; Hara et al., 2015), but comprehensive treatment improvement protocols (see Kazantzis, MacEwan, & Dattilio, 2005) should be examined in prospective trials to demonstrate their effects on homework adherence.

#### Conflict of Interest Statement

Dr. N. Kazantzis receives royalties from Guilford Publications, Inc., Springer Publishing, and Routledge Press (which includes royalties from books on the content of this research). Consulting and honoraria during the past several years have come from the Australian Psychological Society, Cektos (Denmark), the International Association of Cognitive Psychotherapy, and the World Congress of Behavior and Cognitive Therapies.

Michael Thase reports no potential conflicts of interests associated with this research. He does report earning income as a consultant from the following entities during the past 36 months: Akili Inc., Alkermes, Allergan, AstraZeneca, Bristol-Myers Squibb Company, Cerecor, Fabre-Kramer Pharmaceuticals, Inc., Gerson Lehman Group, Guidepoint Global, Janssen, H. Lundbeck A/S, MedAvante, Merck and Co., Moksha8, Neuronetics, Inc., Otsuka, Pamlab LLC (Nestle), Pfizer, Roche, Shire US, Inc., Sunovion Pharmaceuticals, Inc., and

Takeda. He also receives royalties from American Psychiatric Foundation, Guilford Publications, Herald House, W.W. Norton & Company, Inc., LLC. He has received grant funding from the Acadia, Agency for Healthcare Research and Quality, Allergan, AssureRx, Avanir, Axome, Forest Pharmaceuticals, Janssen, Intracellular, National Institutes of Health, Patient Centered Outcome Research Institute, and Takeda.

#### Appendix A. Supplementary Data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.beth.2018.05.010>.

#### References

- Addis, M. E., & Jacobson, N. S. (2000). A closer look at the treatment rationale and homework compliance in cognitive-behavioral therapy for depression. *Cognitive Therapy and Research*, 24(3), 313–326. <https://doi.org/10.1023/A:1005563304265>
- Aderka, I. M., Nickerson, A., Boe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: a meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
- Beck, A. T., Rush, A. J., Shaw, B., & Emery, G. (1979). *Cognitive therapy of depression*. New York: The Guilford Press.
- Bohn, C., Aderka, I. M., Schreiber, F., Stangier, U., & Hofmann, S. G. (2013). Sudden gains in cognitive therapy and interpersonal therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology*, 81(1), 177–182. <https://doi.org/10.1037/a0031198>
- Braun, J. D., Strunk, D. R., Sasso, K. E., & Cooper, A. A. (2015). Therapist use of Socratic questioning predicts session-to-session symptom change in cognitive therapy for depression. *Behaviour Research and Therapy*, 70, 32–37. <https://doi.org/10.1016/j.brat.2015.05.004>
- Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Sturmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12), 1149–1156. <https://doi.org/10.1093/aje/kwj149>
- Bryant, M. J., Simons, A. D., & Thase, M. E. (1999). Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research. *Cognitive Therapy and Research*, 23(4), 381–399. <https://doi.org/10.1023/A:1018703901116>
- Burns, D. D., & Nolen-Hoeksema, S. (1992). Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: a structural equation model. *Journal of Consulting and Clinical Psychology*, 60(3), 441–449. <https://doi.org/10.1037/0022-006X.60.3.441>
- Burns, D. D., & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? *Journal of Consulting and Clinical Psychology*, 68(1), 46–56. <https://doi.org/10.1037//0022-006X.68.1.46>
- Busch, A. M., Kanter, J. W., Landes, S. J., & Kohlenberg, R. J. (2006). Sudden gains and outcome: A broader temporal analysis of cognitive therapy for depression. *Behavior Therapy*, 37(1), 61–68. <https://doi.org/10.1016/j.beth.2005.04.002>
- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): An analysis of public data. *Lancet (London, England)*, 391(10121), 679–686. [https://doi.org/10.1016/S0140-6736\(17\)32133-5](https://doi.org/10.1016/S0140-6736(17)32133-5)

- Cuijpers, P., Weitz, E., Lamers, F., Penninx, B. W., Twisk, J., DeRubeis, R. J., . . . Hollon, S. D. (2017). Melancholic and atypical depression as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression. *Depression and Anxiety*, 34(3), 246–256. <https://doi.org/10.1002/da.22580>
- DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others—and some groups of therapists less—about what matters in psychotherapy. *Psychotherapy Research*, 24(3), 419–428. <https://doi.org/10.1080/10503307.2013.838654>
- DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, 62(4), 409–416. <https://doi.org/10.1001/archpsyc.62.4.409>
- Doane, L. S., Feeny, N. C., & Zoellner, L. A. (2010). A preliminary investigation of sudden gains in exposure therapy for PTSD. *Behaviour Research and Therapy*, 48(6), 555–560. <https://doi.org/10.1016/j.brat.2010.02.002>
- Dong, L., Soehner, A. M., Bélanger, L., Morin, C. M., & Harvey, A. G. (2018). Treatment agreement, adherence, and outcome in cognitive behavioral treatments for insomnia. *Journal of Consulting and Clinical Psychology*. <https://doi.org/10.1037/ccp0000269>
- Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase, M. E., Sackeim, H. A., . . . Kupfer, D. J. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR\*D) study. *Psychiatric Clinics of North America*, 26(2), 457–494. [https://doi.org/10.1016/S0197-2456\(03\)00112-0](https://doi.org/10.1016/S0197-2456(03)00112-0)
- Fennell, M. J. V., & Teasdale, J. D. (1987). Cognitive therapy for depression: Individual differences and the process of change. *Cognitive Therapy and Research*, 11(2), 253–271. <https://doi.org/10.1007/BF01183269>
- Glenn, D., Golinelli, D., Rose, R. D., Roy-Byrne, P., Stein, M. B., Sullivan, G., . . . Craske, M. G. (2013). Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. *Journal of Consulting and Clinical Psychology*, 81(4), 639–649. <https://doi.org/10.1037/a0033403>
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23(1), 56–62.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.
- Haukoos, J. S., & Lewis, R. J. (2015). The Propensity Score. *JAMA*, 314(15), 1637–1638. <https://doi.org/10.1001/jama.2015.13480>
- Hara, K. M., Westra, H. A., Aviram, A., Button, M. L., Constantino, M. J., & Antony, M. M. (2015). Therapist awareness of client resistance in cognitive-behavioral therapy for generalized anxiety disorder. *Cognitive Behaviour Therapy*, 44(2), 162–174. <https://doi.org/10.1080/16506073.2014.998705>
- Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 36(5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>
- Hollon, S. D., Jarrett, R. B., Nierenberg, A. A., Thase, M. E., Trivedi, M., & Rush, A. J. (2005). Psychotherapy and medication in the treatment of adult and geriatric depression. *The Journal of Clinical Psychiatry*, 66(4), 455–468.
- Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2), 481–502. <https://doi.org/10.1111/j.1467-985X.2007.00527.x>
- Jarrett, R. B., Kraft, D., Doyle, J., Foster, B. M., Eaves, G. G., & Silver, P. C. (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase: A randomized clinical trial. *Archives of General Psychiatry*, 58(4), 381–388. <https://doi.org/10.1001/archpsyc.58.4.381>
- Jarrett, R. B., & Thase, M. E. (2010). Comparative efficacy and durability of continuation phase cognitive therapy for preventing recurrent depression: Design of a double-blinded, fluoxetine-and pill placebo-controlled, randomized trial with 2-year follow-up. *Contemporary Clinical Trials*, 31(4), 355–377. <https://doi.org/10.1016/j.cct.2010.04.004>
- Kazantzis, N. (2018). Introduction to the special issue on processes of cognitive behavioral therapy: Does “necessary, but not sufficient” still capture it? *Cognitive Therapy & Research*, 42(2), 115–120. <https://doi.org/10.1007/s10608-018-9891-z>
- Kazantzis, N., Brownfield, N., Mosely, L., Usatoff, A., & Flighty, A. (2017). Homework in CBT: A systematic review of adherence assessment in anxiety and depression treatment (2011–2016). *Psychiatric Clinics of North America*, 40, 625–639. <https://doi.org/10.1016/j.psc.2017.08.001>
- Kazantzis, N., Deane, F. P., & Ronan, K. R. (2000). Homework assignments in cognitive and behavioral therapy: A Meta-analysis. *Clinical Psychology: Science and Practice*, 7(2), 189–202. <https://doi.org/10.1093/clipsy.7.2.189>
- Kazantzis, N., Luong, H. K., Usatoff, A. S., Impala, T., Yew, R. Y., & Hofmann, S. G. (in press). The processes of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy & Research*.
- Kazantzis, N., MacEwan, J., & Dattilio, F. M. (2005). A guiding model for practice. In N. Kazantzis, F. P. Deane, K. R. Ronan, & L. L'Abate (Eds.), *Using homework assignments in cognitive behavior therapy* (pp. 359–407). New York: Routledge.
- Kazantzis, N., Ronan, K. R., & Deane, F. P. (2001). Concluding causation from correlation: Comment on Burns and Spangler (2000). *Journal of Consulting and Clinical Psychology*, 69(6), 1079–1083. <https://doi.org/10.1037/0022-006X.69.6.1079>
- Kazantzis, N., Whittington, C., & Dattilio, F. (2010). Meta-analysis of homework effects in cognitive and behavioral therapy: A replication and extension. *Clinical Psychology: Science and Practice*, 17(2), 144–156. <https://doi.org/10.1111/j.1468-2850.2010.01204.x>
- Kazantzis, N., Whittington, C. J., Zelencich, L., Norton, P. J., Kyrios, M., & Hofmann, S. G. (2016). Quantity and quality of homework compliance: A meta-analysis of relations with outcome in cognitive behavior therapy. *Behavior Therapy*, 47(5), 755–772. <https://doi.org/10.1016/j.beth.2016.05.002>
- Kornblith, S. J. (2000). *The therapist compliance checklist*. Unpublished material. University of Pittsburgh, Pittsburgh, PA.
- Kornblith, S. J., Rehm, L. P., O'Hara, M. W., & Lamparski, D. M. (1983). The contribution of self-reinforcement training and behavioral assignments to the efficacy of self-control therapy for depression. *Cognitive Therapy and Research*, 7, 499–528.
- 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*. <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., German, R. E., & DeRubeis, R. J. (2015). It's complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review, 41*, 3–15. <https://doi.org/10.1016/j.cpr.2014.12.003>
- Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores. *Evaluation Review, 29*(6), 530–558. <https://doi.org/10.1177/0193841x05275596>
- Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine, 23*(19), 2937–2960. <https://doi.org/10.1002/sim.1903>
- Mausbach, B. T., Moore, R., Roesch, S., Cardenas, V., & Patterson, T. L. (2010). The relationship between homework compliance and therapy outcomes: An updated meta-analysis. *Cognitive Therapy and Research, 34*(5), 429–438. <https://doi.org/10.1007/s10608-010-9297-z>
- Neimeyer, R. A., Kazantzis, N., Kassler, D. M., Baker, K. D., & Fletcher, R. (2008). Group cognitive behavioural therapy for depression outcomes predicted by willingness to engage in homework, compliance with homework, and cognitive restructuring skill acquisition. *Cognitive Behaviour Therapy, 37*(4), 199–215. <https://doi.org/10.1080/16506070801981240>
- Norcross, J. C., & Lambert, M. J. (2014). Relationship science and practice in psychotherapy: Closing commentary. *Psychotherapy, 51*(3), 398–403. <https://doi.org/10.1037/a0037418>
- Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*(5), 550–560.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine, 26*(1), 20–36. <https://doi.org/10.1002/sim.2739>
- Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., . . . Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR\*D): Rationale and design. *Controlled Clinical Trials, 25*(1), 119–142. [https://doi.org/10.1016/S0197-2456\(03\)00112-0](https://doi.org/10.1016/S0197-2456(03)00112-0)
- Rush, A. J., Kovacs, M., Beck, A. T., Weissenburger, J., & Hollon, S. D. (1981). Differential effects of cognitive therapy and pharmacotherapy on depressive symptoms. *Journal of Affective Disorders, 3*(3), 221–229. [https://doi.org/10.1016/0165-0327\(81\)90024-0](https://doi.org/10.1016/0165-0327(81)90024-0)
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T., Arnow, B. A., Klein, D. N., . . . Keller, M. B. (2003). The 16-item quick inventory of depressive symptomatology (QIDS) clinician rating (QIDS-C) and self-report (QIDS-SR). *Biological Psychiatry, 54*(5), 573–583. [https://doi.org/10.1016/S0006-3223\(02\)01866-8](https://doi.org/10.1016/S0006-3223(02)01866-8)
- Sasso, K. E., Strunk, D. R., Braun, J. D., DeRubeis, R. J., & Brotman, M. A. (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>
- Schibbye, P., Ghaderi, A., Ljótsson, B., Hedman, E., Lindfors, N., Rück, C., & Kaldø, V. (2014). Using early change to predict outcome in cognitive behaviour therapy: Exploring timeframe, calculation method, and differences of disorder-specific versus general measures. *PLOS ONE, 9*(6), 1–10. <https://doi.org/10.1371/journal.pone.0100614>
- Spielmann, G. I., & Flückiger, C. (2018). Moderators in psychotherapy meta-analysis. *Psychotherapy Research, 1–14*. <https://doi.org/10.1080/10503307.2017.1422214>
- Strunk, D. R., Cooper, A. A., Ryan, E. T., DeRubeis, R. J., & Hollon, S. D. (2012). The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intersession symptom gains. *Journal of Consulting and Clinical Psychology, 80*(5), 730–738. <https://doi.org/10.1037/a0029281>
- Strunk, D. R., & Pfeifer, B. J. (2015). The value of research methods in psychotherapy research: A reply to the commentary of Dr. Norton. *Australian Psychologist, 50*(6), 416–417. <https://doi.org/10.1111/ap.12150>
- Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904.
- Thase, M. E., & Callan, J. A. (2006). The role of homework in cognitive behavior therapy of depression. *Journal of Psychotherapy Integration, 16*(2), 162–177. <https://doi.org/10.1037/1053-0479.16.2.162>
- Thase, M. E., Friedman, E. S., Biggs, M. M., Wisniewski, S. R., Trivedi, M. H., Luther, J. F., . . . Rush, A. J. (2007). Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: A STAR\*D report. *American Journal of Psychiatry, 164*(5), 739–752. <https://doi.org/10.1176/ajp.2007.164.5.739>
- Trojano, M., Pellegrini, F., Paolicelli, D., Fuiani, A., & Renzo, V. D. (2009). Observational studies: Propensity score analysis of non-randomized data. *International MS Journal, 16*(3), 90–97.
- 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. *Journal of Consulting and Clinical Psychology, 80*(3), 373–381. <https://doi.org/10.1037/a0027663>
- Weck, F., Richtberg, S., Esch, S., Hofling, V., & Stangier, U. (2013). The relationship between therapist competence and homework compliance in maintenance cognitive therapy for recurrent depression: Secondary analysis of a randomized trial. *Behavior Therapy, 44*(1), 162–172. <https://doi.org/10.1016/j.beth.2012.09.004>
- West, S. G., Cham, H., Thoemmes, F., Renneberg, B., Schulze, J., & Weiler, M. (2014). Propensity scores as a basis for equating groups. *Journal of Consulting and Clinical Psychology, 82*(5), 906–919. <https://doi.org/10.1037/a0036387>
- Westreich, D., Cole, S. R., Funk, M. J., Brookhart, M. A., & Sturmer, T. (2011). The role of the c-statistic in variable selection for propensity score models. *Pharmacoepidemiology and Drug Safety, 20*(3), 317–320. <https://doi.org/10.1002/pds.207>
- Wucherpfennig, F., Rubel, J. A., Hollon, S. D., & Lutz, W. (2017). Sudden gains in routine care cognitive behavioral therapy for depression: A replication with extensions. *Behaviour Research and Therapy, 89*, 24–32. <https://doi.org/10.1016/j.brat.2016.11.003>

RECEIVED: October 16, 2017

ACCEPTED: May 30, 2018

AVAILABLE ONLINE: 5 June 2018