Received: 15 June 2016 | Revised: 2 November 2016 | Accepted: 14 November 2016

DOI: 10.1002/da.22590

RESEARCH ARTICLE

WILEY AD/AA

Internet-based cognitive behavior therapy for major depressive disorder: A randomized controlled trial

Isabelle M. Rosso, PhD[1,2] | William D.S. Killgore, PhD[3] | Elizabeth A. Olson, PhD[1,2] | Christian A. Webb, PhD[1,2] | Rena Fukunaga, PhD[1,2] | Randy P. Auerbach, PhD[1,2] | Hannah Gogel, MA[1] | Jennifer L. Buchholz, AB[1] | Scott L. Rauch, MD[1,2]

[1]Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, MA, USA

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[3]Department of Psychology, University of Arizona, Tucson, AZ, USA

Correspondence
Scott L. Rauch, Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, MA, 02478.
Email: srauch@partners.org.

Grant sponsor: US Army Military Operational Medicine Research Program; Contract grant number: W81XWH-12-1-0109.

Background: Prior research has shown that the Sadness Program, a technician-assisted Internet-based cognitive behavioral therapy (iCBT) intervention developed in Australia, is effective for treating major depressive disorder (MDD). The current study aimed to expand this work by adapting the protocol for an American population and testing the Sadness Program with an attention control group.

Methods: In this parallel-group, randomized controlled trial, adult MDD participants (18–45 years) were randomized to a 10-week period of iCBT (n = 37) or monitored attention control (MAC; n = 40). Participants in the iCBT group completed six online therapy lessons, which included access to content summaries and homework assignments. During the 10-week trial, iCBT and MAC participants logged into the web-based system six times to complete self-report symptom scales, and a nonclinician technician contacted participants weekly to provide encouragement and support. The primary outcome was the Hamilton Rating Scale for Depression (HRSD), and the secondary outcomes were the Patient Health Questionnaire-9 and Kessler-10.

Results: Intent-to-treat analyses revealed significantly greater reductions in depressive symptoms in iCBT compared with MAC participants, using both the self-report measures and the clinician-rated HRSD (d = −0.80). Importantly, iCBT participants also showed significantly higher rates of clinical response and remission. Exploratory analyses did not support illness severity as a moderator of treatment outcome.

Conclusions: The Sadness Program led to significant reductions in depression and distress symptoms. With its potential to be delivered in a scalable, cost-efficient manner, iCBT is a promising strategy to enhance access to effective care.

KEYWORDS
cognitive behavioral therapy, Internet, intervention, major depressive disorder, randomized controlled trial (RCT)

1 | INTRODUCTION

Major depressive disorder (MDD) is a common and debilitating disease (Hasin, Goodwin, Stinson, & Grant, 2005; Kessler et al., 2003). Although effective evidence-based therapies exist, the majority of individuals with depression (>70%) do not seek treatment (Andrews, 2001; Tolin, 2010), which may reflect the many barriers to adequate care, including geographic distance, prohibitive cost, unavailability of clinicians, and perceived stigma. To address this issue, Internet-delivered treatments (e.g., Internet-based cognitive behavior therapy [iCBT]) have been developed, and prior research has demonstrated positive short- and long-term effects (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010). Internet-based interventions offer brief, structured treatment that can be accessed by patients from the convenience and privacy of their own homes, relatively anonymously, and according to their own schedules. Moreover, iCBT enables

Abbreviations: GENLIN, generalized linear models; HLM, hierarchical linear modeling; HRSD, Hamilton Rating Scale for Depression; iCBT, Internet-based cognitive behavioral therapy; ITT, intent-to -treat; MAC, monitored attention control; MDD, major depressive disorder; MI, multiple imputation; RA, research assistant; RCT, randomized controlled trial; SCID, Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition; SI, suicidal ideation

treatment delivery to a large number of patients while minimizing costs and clinician time (Donker et al., 2015).

In a meta-analysis of randomized controlled trials (RCTs) for depression using Internet- and computer-delivered CBT, Andersson and Cuijpers (2009) reported a small-to-medium effect size (d = 0.41) for Internet- and computer-delivered CBT versus control groups; however, therapist-guided approaches (d = 0.61) were more effective than self-guided interventions (d = 0.25) (Andersson & Cuijpers, 2009). Despite the surge in Internet-delivered treatments, research testing the efficacy of these interventions has often lacked sufficient methodological rigor. Kiluk and colleagues reviewed the most common and significant methodological weaknesses: (1) lack of standardized diagnostic criteria; (2) lack of comparison condition or use of a wait-list only control; (3) sole reliance on self-report for assessment of outcome, rather than interviews administered by assessors blind to treatment condition; and (4) insufficient attention to issues of internal validity (e.g., no intent-to -treat (ITT) analyses) (Kiluk et al., 2011). Thus, while there is growing support for iCBT as a feasible and efficacious treatment approach for MDD, efficacy studies incorporating these criteria are necessary before scaling to larger populations.

Researchers in Australia developed an iCBT program for MDD, the Sadness Program, which has demonstrated efficacy at reducing depressive symptoms among adults (Perini, Titov, & Andrews, 2008, 2009; Titov et al., 2010). Briefly, this program comprises six online lessons, homework assignments, automatic reminder and notification emails, and weekly supportive contact with a technician over an 8- to 10-week period (Titov et al., 2010). Of note, the CBT treatment period is completed without clinical staff; however, clinicians are available to attend to suicide-related issues. Given its clinical efficacy and reduced demand for clinician time, this iCBT Sadness Program is highly cost effective (Hedman, Ljótsson, & Lindefors, 2012); moreover, completion rates were high (~80%) and effect sizes large (d = 1.20–1.60) for within-subject improvement in depression over a 10-week intervention (Titov et al., 2010). In addition, although this prior study was underpowered to support a claim of non-inferiority, it found that response rates statistically did not differ when the treatment was guided by a clinician or technician (Titov et al., 2010). At the same time, prior studies have not used rigorous methodology to test iCBT efficacy, which may limit our understanding of this promising approach.

In this study, we aimed to test the efficacy of the Australian Sadness Program (www.thiswayup.org.au) in a sample of depressed American adults, using ITT analyses with multiple imputation. To be culturally relevant, the original content was modified for an American cohort (Appendix). Additionally, iCBT was compared to a monitored attention control (MAC) group, as a more stringent comparison than the waitlists more commonly employed in prior studies. We hypothesized that relative to MAC participants, MDD participants assigned to iCBT would show greater depression symptom reduction and higher rates of clinical response and remission, using interviewer-rated and self-report measures. Based on evidence that initial illness severity may predict differential outcome in iCBT (Andersson et al., 2005; Button, Wiles, Lewis, Peters, & Kessler, 2012; de Graaf, Hollon, & Huibers, 2010; Meyer et al., 2009; Warmerdam, van Straten, Twisk, Riper, & Cuijpers, 2008; Warmerdam, van Straten, Twisk, & Cuijpers, 2013), we also tested for moderating effects of baseline depression severity, number of prior depressive episodes, and anxiety disorder comorbidity.

2 | MATERIALS AND METHODS

2.1 | Recruitment and selection procedures

This study was approved by the Institutional Review Boards of McLean Hospital and Partners Healthcare, and registered at ClinicalTrials.gov (Identifier: NCT01598922). This was a single-center, parallel group RCT, with equal randomization to iCBT and MAC. Participants were recruited through Internet advertisements and community fliers. All participants provided written informed consent. This research was performed in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

The inclusion criteria were a primary diagnosis of current MDD according to the Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision (DSM-IV-TR; (American Psychiatric Association, 2000)), Patient Health Questionnaire-9 (PHQ-9; (Kroenke, Spitzer, & Williams, 2001)) scores between 10 and 23 (inclusive), age 18 to 45, ability to read English, regular access to a phone and computer with Internet access, absence of psychotropic medications for at least 2 weeks (6 weeks for fluoxetine, 6 months for neuroleptics), and right-handedness (due to the magnetic resonance imaging (MRI) component of the study). Exclusion criteria were severe depression (initial PHQ-9 total score >23), significant suicidal ideation (SI; initial PHQ-9 item 9 score >1), lifetime history of bipolar disorder or schizophrenia spectrum disorder, current or past alcohol or substance dependence, current alcohol abuse, current or past substance abuse, use of illicit drugs except cannabis within the past year, use of cannabis within the past month, current participation in any form of cognitive behavioral therapy, history of electroconvulsive therapy, less than ninth grade education, and contraindications for MRI.

Eligible participants completed three study visits (Fig. 1). Individuals who passed an initial telephone screen were scheduled for a Day 1 visit to provide written informed consent and to determine eligibility based on an initial PHQ-9 administration, the Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID; (First et al., 2002)), handedness screening, and MRI safety screening. If eligible, participants were scheduled for a Day 2 visit, as soon as possible following Day 1 (M = 14.68 days, range = 1–63 days). If more than 30 days elapsed between the first and second study visits, the SCID and PHQ-9 were re-administered to re-establish eligibility (n = 4 participants). Day 2 procedures consisted of questionnaires and interviews, including the Hamilton Rating Scale for Depression, 17-item version (HRSD; (Hamilton, 1960)), administered by doctoral-level clinicians blind to diagnostic group and eventual treatment group assignment. Participants then underwent an MRI scan. At the end of Day 2, MDD participants were randomly assigned to iCBT or MAC (Appendix). Participants were notified of their group assignment, and the MAC group was informed that they would receive access to the iCBT program after completion of the 10-week monitoring period and Day 3 visit. At the

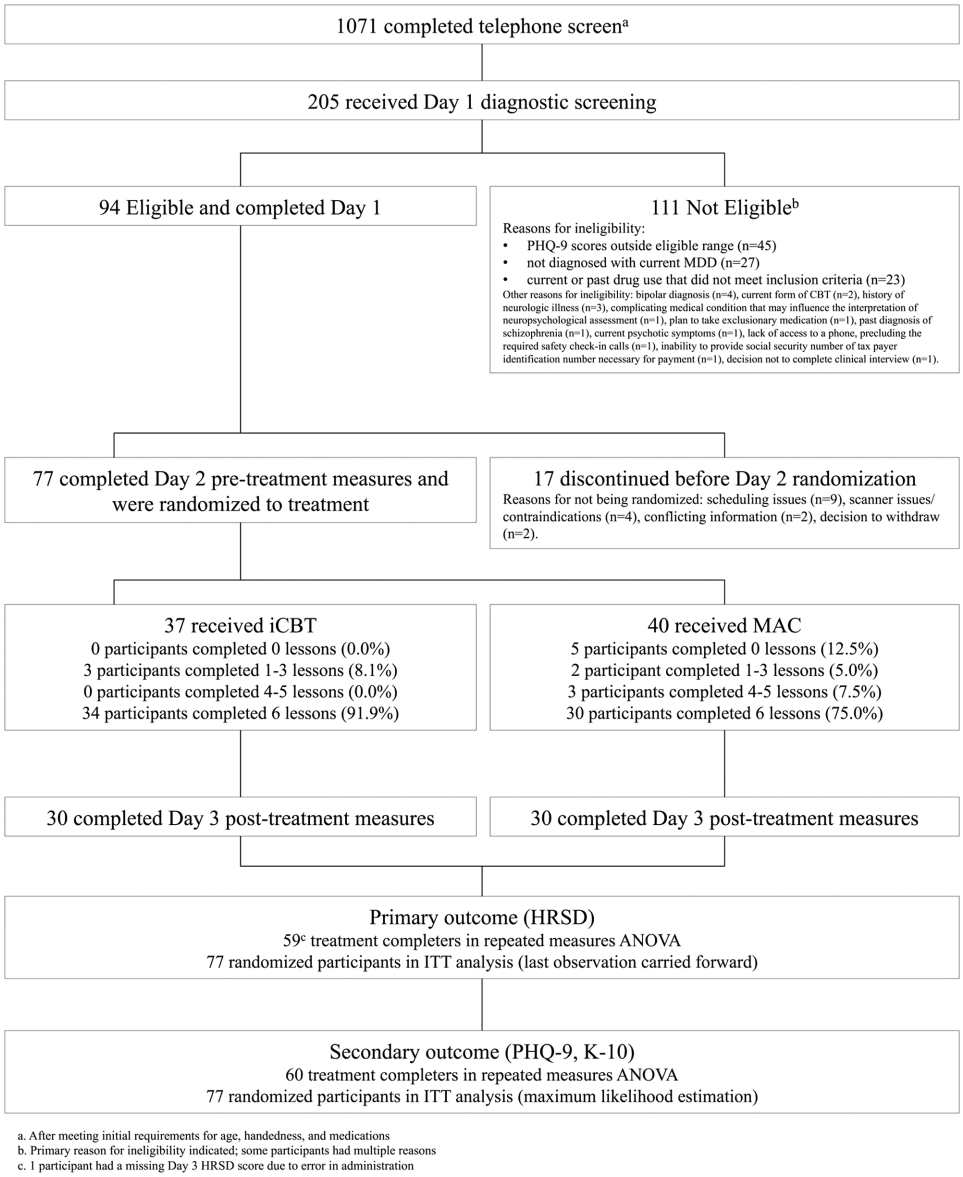

FIGURE 1 Flow of participants from initial telephone screen through randomization, outcome, and analysis

end of the 10-week treatment period, participants returned for the Day 3 visit. Day 3 repeated many of the Day 2 questionnaires and interviews, including the HRSD interview, administered by doctoral-level clinicians blind to treatment assignment.

2.2 | iCBT treatment program

This study used a modified version of the Sadness Program, a technician-assisted iCBT program developed at the UNSW (Perini et al., 2009; Titov et al., 2010). Modifications involved language adaptations and minor content alterations for relevance to American culture (e.g., providing MDD prevalence rates for the U.S. rather than Australia). We also modified the user interface and data collection procedures for HIPAA compliance, including implementation of confidential login procedures and de-identification of weekly self-report data.

Participants logged into the system six times during the 10-week intervention period. Immediately upon logging in, and prior to the start of each lesson, participants completed the PHQ-9 and Kessler Psychological Distress Scale (K-10; (Kessler et al., 2002)). Those in the iCBT group then had access to the weekly CBT lessons, completed sequentially from lesson 1 to 6 (Table 1). Lessons are presented as an illustrated cartoon strip about a character named "Jess" who experiences depression and anxiety symptoms, and who learns how to alleviate these symptoms using CBT. After completing each lesson, iCBT participants could download: a lesson summary, homework assignments, and optional supplemental resources. Participants in the MAC group also logged into the online system six times during the 10-week period, but their "lessons"consisted only of the PHQ-9 and K-10 questionnaires. Completion of the program was self-paced, with the exception that participants had to wait at least 5 days in between consecutive lessons and complete all lessons within 10 weeks. Remuneration was up to $500 based on the time invested in completing the study, including pre- and posttreatment MRI scanning, prorated for early termination; participants were not required to log into the system to be paid.

TABLE 1 iCBT weekly lesson plans

Lesson	Title	Content
1	The diagnosis: what is depression?	✓ Emotional, physical, cognitive and behavioral symptoms ✓ Cognitive behavioral model ✓ Identifying and tracking individual symptoms of depression
2	Monitoring thoughts and activities	✓ Breaking the cycle of depression by modifying behavior ✓ Monitoring and increasing activity ✓ Recognizing unhelpful thinking patterns ✓ Reducing depressive rumination ✓ Sleep hygiene
3	Learning to improve activities and thoughts	✓ Activity planning and tracking ✓ Challenging and changing unhelpful thoughts ✓ Recording positives
4	Facing fears	✓ Structured problem solving ✓ Facing fears gradually ✓ Developing a graded hierarchy to face fears
5	Being assertive	✓ Passive, aggressive, and assertive behaviors ✓ Effective communication and active listening
6	Preventing relapse	✓ Lapses and relapses ✓ Relapse prevention planning

2.3 | Contact with study staff

Participants in the iCBT and MAC groups received weekly check-in telephone calls from a trained bachelor-level research assistant (RA). These calls followed a script and were limited to 3–5 min, as previously published (Titov et al., 2010 Appendix).

During the 10-week trial, participants only spoke with a clinician if their weekly PHQ-9 or K-10 indicated elevated suicidality or distress. Specifically, an email alert was sent to the participants' assigned clinician in the case of: (1) significant SI on the PHQ-9 (i.e., a score of 2 or 3 on item 9); (2) any increase in SI on the PHQ-9; (3) greater than 5-point increase on the K-10; 4) total K-10 score greater than 35. Within one business day of the alert, participants were evaluated by telephone by doctoral-level clinicians, using the Columbia Suicide Severity Rating Scale (CSSRS; (Posner et al., 2011)). Fifteen (44%) of iCBT participants and 18 (47%) of MAC participants were evaluated ($X^2(1,77) = 0.156$, $P = .693$). No participants hitting alert triggers required additional in-person assessment or triage to standard care.

2.4 | Measures

Participants completed clinician-administered and self-report measures of depression. All interviewers were doctoral-level clinical psychologists or psychiatrists who were trained on the assessments, supervised by senior investigators, and blind to treatment condition. The SCID (First, Mentzel, & Miltner, 2002) was used to establish MDD diagnosis, and any other current and lifetime DSM-IV Axis I diagnoses.

2.4.1 | Primary outcome

The HRSD (Hamilton, 1960) was the primary measure of depression severity, administered by doctoral-level raters at pretreatment (Day 2) and posttreatment (Day 3). The 17-item version of the HRSD is the most often utilized and has become the standard as a criterion outcome measure in clinical trials of depression (Cusin, Yang, Yeung, & Fava, 2009). Each item is rated on a 3- or 5-point scale (0–2 or 0–4), and total HRSD scores can range from 0 to 52. The American Psychiatric Association's *Handbook of Psychiatric Measures* (Rush, First, & Blacker, 2008) defines "mild" depression as HRSD scores from 8 to 13, "moderate" depression from 14 to 18, "severe" depression from 19 to 22, and "very severe" depression as >22. The HRSD also has criteria to define treatment "response" as a 50% score reduction from baseline, and "remission" as a score ≤ 7 posttreatment.

2.4.2 | Secondary outcomes

The PHQ-9 (Kroenke et al., 2001) is a well-validated self-report measure of depression severity. The scale presents each of the nine DSM-IV MDD symptoms to be rated on a 4-point frequency scale ranging from "0" (not at all) to "3" (nearly every day). SI is assessed on item-9, which queries the frequency of "thoughts that you would be better off dead, or of hurting yourself in some way." On the Day 1 screening visit, only participants who endorsed scores of "0" (no SI) or "1" (passive SI) on item 9 were eligible for randomization. Total PHQ-9 scores range from 0 to 27 (Zuithoff et al., 2010). Participants completed the PHQ-9 at baseline, prior to each weekly lesson, and at posttreatment (eight time-points).

The K-10 (Kessler et al., 2002) is a self-report questionnaire of general psychological distress. Items assess frequency of ten depression and anxiety symptoms over the past thirty days (e.g., "About how often did you feel nervous?") on a 5-point scale from "1" ("none of the time") to "5" ("all of the time"). Total K-10 scores range from 10 to 50. The K-10 has good reliability and validity (Furukawa, Kessler, Slade, &

Andrews, 2003). Participants completed the K-10 at baseline, prior to each weekly lesson, and at posttreatment (eight time-points).

2.5 | Statistical analyses

Analyses were conducted with SPSS version 20 (IBM corporation), unless otherwise noted. We conducted ITT analyses that included all randomized participants (Fig. 1/Flowchart). Missing HRSD data were handled using multiple imputation (MI): we generated 20 datasets (Graham, Olchowski, & Gilreath, 2007), stratified by treatment group, using the fully conditional specification algorithm in SPSS. The imputation model employed pretreatment HRSD scores and all available PHQ-9 depression scores. We then used SPSS Generalized Linear Models (GENLIN) to conduct multilevel analyses of HRSD outcome in the imputed dataset, and to obtain pooled results. The analysis included treatment group as a predictor, and covaried for baseline HRSD, age, and sex. Treatment response and remission rates, based on HRSD criteria, were compared between groups using logistic regression analyses of pooled imputed data and controlling for age and sex.

Hierarchical Linear Modeling (HLM; (Raudenbush & Bryk, 2002)) was applied to the eight time-point PHQ-9 and K-10 data, which consisted of repeated symptom assessments nested within individuals. HLM analyses implemented mixed-effects repeated-measures models using SAS (version 9.2) PROC MIXED (SAS Institute Inc., Cary, NC). HLM models adjusted for baseline values of the dependent measures, and subject-specific slopes and intercepts were treated as randomly varying across individuals. *Group* × *Time* interactions were included to test for between-group differences in the slope of improvement of PHQ-9 and K-10 scores. Age and sex were entered as covariates. Cohen's *d* was used to represent magnitude of the effect sizes for within-group and between-group differences in pre- to posttreatment scores on outcome measures.

Moderation of primary outcome was examined using separate analyses for each of the three potential moderators. For the moderating effects of anxiety disorder comorbidity and number of depressive episodes, the analysis consisted of repeating the GENLIN analysis on imputed data, after adding as predictors the main effect of the moderator and its interaction with treatment group (McClelland & Judd, 1993). For the moderating effect of baseline depression severity, only the interaction term (i.e., *Baseline HRSD* × *Treatment Group*) needed to be added as a predictor since baseline HRSD score was already a covariate.

3 | RESULTS

3.1 | Participant enrollment, attrition, and sample characteristics

The randomization procedure yielded groups matched on age, sex distribution, and education (Fig. 1; Table 2). Forty-eight (63%) of the 77 randomized MDD participants met criteria for one or more current comorbid DSM-IV Axis I disorder(s), namely: social phobia (n = 27), generalized anxiety disorder (n = 16), posttraumatic stress disorder (n = 12), specific phobia (n = 3), anxiety disorder NOS (n = 2), panic disorder without agoraphobia (n = 1), agoraphobia (n = 1), obsessive compulsive disorder (n = 1), bulimia nervosa (n = 1). Within the randomized MDD sample, there was a significant positive correlation between age and baseline HRSD score ($r(75) = 0.25$, $P = .03$), and female MDD patients had significantly higher baseline HRSD scores than male MDD patients ($F(1,75) = 5.19$, $P = .03$). Thus, all subsequent analyses entered age and sex as covariates. Thirty-four (92%) of the randomized iCBT participants and 30 (75%) of the MAC participants completed all six lessons ($\chi^2(1,77) = 4.11$, $P = .04$). These attrition rates are comparable to those reported in the Australian trial (Titov et al., 2010).

3.2 | Primary outcome

Using GENLIN, there was a significant main effect of treatment group (β = 6.109, SE = 1.68, Wald χ^2 95% CI [2.814, 9.403], $P < .0001$), and the effects of baseline HRSD scores, age, and sex were not statistically significant. This reflected significantly lower posttreatment HRSD scores in the iCBT compared to the MAC group, corresponding to a large effect ($d = -0.80$) for iCBT (see Table 3). Calculation of pre- to posttreatment within-group effect sizes showed that HRSD scores declined significantly for iCBT participants ($d = -1.16$, 95% CI [−0.66, −1.65] but not MAC participants ($d = -0.34$, 95% CI [−0.10, −0.78]; Table 3).

Fifty-seven percent (n = 21) of the iCBT participants responded to the treatment compared with 18% (n = 7) of MAC participants. Logistic regression showed that this corresponded to a statistically significant group difference ($\beta = -2.041$, SE = .611, OR = 7.698, 95% CI OR [2.322, 25.519], $P < .001$). Similarly, 57% (n = 21) of iCBT and 14% (n = 5) of MAC participants remitted, a difference that was statistically significant ($\beta = -2.210$, SE = .649, OR = 9.118, 95% CI OR = [2.551, 32.586], $P < .001$). Clinically significant improvement also was significantly higher in iCBT than MAC using the Jacobsen and Truax (1991) definition (Appendix).

3.3 | Secondary outcomes

Using HLM, a significant *Group* × *Time* interaction emerged, indicating greater improvement in PHQ-9 scores in the iCBT group than in the MAC group ($F(1, 63.6) = 5.32$; $P = .024$; Fig. 2). A similar *Group* × *Time* interaction reflected significantly greater improvement in K-10 distress scores in the iCBT relative to the MAC group ($F(1, 62.1) = 9.82$; $P = .003$; Fig. 2).

3.4 | Exploratory analyses

Treatment group did not interact significantly with anxiety comorbidity, number of depressive episodes, or baseline depression severity in the prediction of posttreatment HRSD scores.

4 | DISCUSSION

In the current study, iCBT led to significant reduction of depression and distress symptoms as compared to MAC, which corresponded to large

TABLE 2 Baseline demographic and clinical characteristics of participants randomized to treatment, mean (SD) or *n* (%)

Baseline characteristic	iCBT (*n* = 37)	MAC (*n* = 40)
Female	23 (62.2)	30 (75.0)
Race		
White	23 (62.2)	25 (62.5)
Black	1 (2.7)	2 (5.0)
Asian	6 (16.2)	5 (12.5)
More than one race	1 (2.7)	3 (7.5)
Other	1 (2.7)	1 (2.5)
Unknown	5 (13.5)	4 (10)
Age	29.2 (7.69)	28.8 (6.74)
Marital status, married	5 (13.5)	8 (20.0)
Education (years)	15.4 (2.52)	15.3 (1.99)
College graduate	20 (54.0)	21 (52.5)
Employed outside the home	28 (75.7)	24 (60.0)
HRSD	15.7 (4.02)	15.7 (4.26)
Low (HRSD 8–19)	31 (83.8)	32 (80.0)
High (HRSD ≥ 20)	6 (16.2)	8 (20.0)
PHQ-9	13.9 (3.89)	15.6 (3.97)
K-10	28.9 (5.42)	30.8 (4.54)
Number prior MDEs	5.9 (10.85)[a]	3.2 (2.93)[b]
Recurrent MDD	29 (78.4)[c]	28 (71.8)
Age of onset first episode	16.5 (5.92)[d]	15.7 (4.41)[e]
Age of onset prior to 18	21 (63.6)[f]	28 (77.8)[g]
Any lifetime psychiatric hospitalizations	0 (0.0)	5 (12.5)*
Any current DSM-IV Axis I disorder	22 (59.5)	26 (65.0)
Any lifetime DSM-IV Axis I disorder	12 (32.4)	6 (15.0)
Any lifetime alcohol abuse	4 (10.8)	5 (12.5)

Notes $^{*}P < .05$.
[a]Number of episodes coded as too indistinct to count for four cases.
[b]Number of episodes coded as too indistinct to count for nine cases.
[c]Information not available for one case.
[d]Information not available for five cases.
[e]Information not available for eight cases.
[f]Information not available for four cases.
[g]Information not available for four cases.
HRSD, Hamilton Rating Scale for Depression; PHQ-9, Patient Health Questionnaire; K-10, Kessler Distress Scale; MDE, major depressive episode; MDD, major depressive disorder

TABLE 3 Pre- and posttreatment scores for outcome measures in iCBT participants (*n* = 37) and MAC participants (*n* = 40)

					Effect sizes *d* (95% CI)	
Outcome measure	Group	Pretreatment	Posttreatment	Pre-post mean difference (95% CI)	Within group	iCBT vs. MAC
HRSD						
	iCBT	15.73 ± 4.02	9.17 ± 6.92	6.56 (3.93, 9.19)	1.16 (0.66, 1.65)	−0.80 (−1.25, −0.32)
	MAC	15.70 ± 4.26	14.05 ± 5.34	1.65 (−0.50, 3.80)	0.34 (−0.10, 0.78)	
PHQ-9						
	iCBT	13.92 ± 3.89	7.70 ± 5.39	6.22 (4.04, 8.39)	1.32 (0.81, 1.81)	−0.79 (−1.24, −0.32)
	MAC	15.58 ± 3.97	11.88 ± 5.21	3.70 (1.64, 5.76)	0.79 (0.34, 1.25)	
K-10						
	iCBT	28.89 ± 5.42	20.14 ± 7.37	8.84 (5.84, 11.84)	1.35 (0.85, 1.86)	−0.95 (−1.42, −0.48)
	MAC	30.75 ± 4.55	26.73 ± 6.55	4.03 (1.51, 6.54)	0.71 (0.26, 1.16)	

Note: HRSD, Hamilton Rating Scale for Depression; PHQ-9, Patient Health Questionnaire; K-10, Kessler Distress Scale.

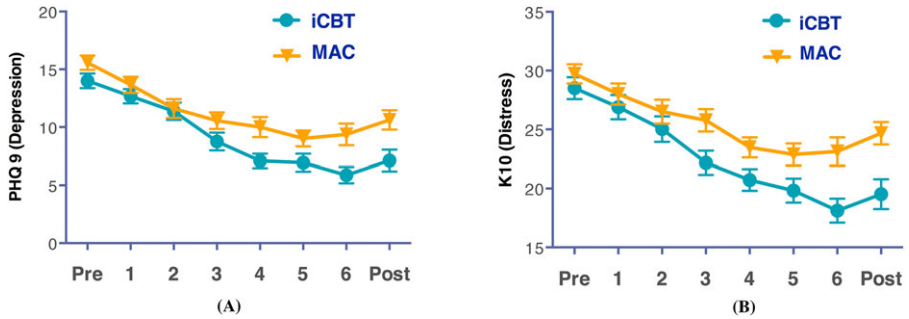

FIGURE 2 Weekly measures of self-reported depression (A) and distress (B) in iCBT ($n = 37$) and MAC participants ($n = 40$)

effect sizes. Moreover, greater than half of the iCBT group exhibited a clinically significant treatment response and remitted. Exploratory analyses indicated that the iCBT treatment effect was not moderated by baseline depression severity, number of depressive episodes, or anxiety disorder comorbidity.

Our results are consistent with prior reports on the efficacy of iCBT for depression generally (Andersson & Cuijpers, 2009; Andrews et al., 2010; Johansson & Andersson, 2012; Richards & Richardson, 2012). Two previous RCTs of the Sadness Program in Australian samples (Perini et al., 2009; Titov et al., 2010) demonstrated the efficacy of this iCBT protocol. Extending this work, our study used a standardized clinician-administered interview and an attention control group (e.g., (Titov et al., 2010)). As noted, treatment response and remission rates in the iCBT group were greater than in the MAC cohort. Whereas prior work relied on a waitlist-only comparison group, the control group of this study allows us to conclude that this iCBT intervention was effective above-and-beyond the non-specific factors present in the MAC condition, including symptom monitoring, staff contact, and support.

Symptom amelioration with iCBT was not affected by baseline illness severity. Specifically, none of the three indicators of severity at study entry (HRSD score, number of depressive episodes, anxiety disorder comorbidity) significantly moderated outcome. Thus, the iCBT program was effective in reducing depressive symptoms across the range of mild to moderate MDD represented in our sample, which is consistent with results of some studies finding no predictive or moderating role of baseline depression (Donker et al., 2013; Proudfoot et al., 2003). At the same time, there have been at least as many reports that pretreatment depression severity affects who will benefit from iCBT (e.g., (Andersson et al., 2005; Button et al., 2012; de Graaf et al., 2010; Spek et al., 2008)). Overall, additional research is required to determine whether pretreatment characteristics can be utilized to predict treatment response, as it would be valuable to identify patients most likely to benefit from this low-cost and easily disseminable treatment.

A challenge facing iCBT and similar technology-based psychotherapy applications is executing feasible dissemination from the experimental setting to the clinic without compromising treatment efficacy. There have been a handful of uncontrolled studies showing that iCBT for depression can produce medium-to-large within-subject effects in a general medical setting (Hadjistavropoulos et al., 2014; Hedman et al., 2014; Hoifodt et al., 2013; Newby et al., 2013; Ruwaard, Lange, Schrieken, Dolan, & Emmelkamp, 2012). In addition, one effectiveness study found a large, between-group effect when comparing iCBT and treatment as usual (Kivi et al., 2014). Of particular relevance to the current study, an effectiveness trial conducted with the Australian Sadness Program found that it led to reductions in depressive symptoms (including suicidality) when prescribed by primary care providers (Williams & Andrews, 2013), although the absence of a control group precluded the ability to separate superiority of treatment response from spontaneous remission or placebo response. Moving forward, there is a need for effectiveness studies with more rigorous comparator groups, and for additional non-inferiority trials (Donker et al., 2013) that would buttress the viability of iCBT in routine, American medical practice.

Our findings should be interpreted in the context of its limitations. First, assessment of treatment adherence for online protocols is challenging. We received confirmation emails when lessons were started and completed, and we assessed general progress via weekly phone calls, but we do not have detailed information on the type and duration of CBT activities that participants completed. Second, there are no follow-up data assessing maintenance of treatment effects. Data addressing the extended impact of iCBT are limited, although the initial findings are promising (Kessler et al., 2009; Kivi et al., 2014; Ruwaard et al., 2009). Third, it is unclear whether the current findings generalize to a real world clinical population. This sample was self-selected via response to advertisements, was willing to engage in the additional subject burden of neuroimaging, and was remunerated. Thus, adherence and completion rates remain to be determined for a real-world U.S. clinical sample. Third, although our MAC group controlled for some nonspecific factors that can contribute to symptom improvement, a placebo comparison group would have been more stringent. Finally, our analyses of potential moderators were underpowered. As we accrue data from a larger real world sample, we plan to address those important clinical questions, including identifying the subset of patients for whom iCBT may be an effective cost-efficient treatment, and those for whom it should be avoided due to nonadherence, likely inefficacy or other poor outcomes.

5 | Conclusions

In summary, this Americanized version of the Sadness Program was associated with statistically significant and clinically meaningful reductions in depression and distress-related symptoms. These therapeutic effects were above and beyond those associated with symptom

monitoring, staff contact, and support. The combination of medium-to-large treatment effect sizes and minimal required clinician time indicates that this Internet-delivered therapy can be an efficacious and cost-effective alternative to conventional CBT. With its potential to overcome treatment barriers and to increase access to evidence-based care, iCBT may become a valuable tool for reducing the public health burden of depression.

ACKNOWLEDGMENTS

We thank Garrett Fitzmaurice, DSc, for providing expert statistical consultation on multiple imputation analyses. We are grateful to the patients who participated this study, and to the MRI technologists of McLean Imaging Center. Research activities were funded by the US Army Military Operational Medicine Research Program (Award # W81XWH-12-1-0109; PI: Rauch).

CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

REFERENCES

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text revision (DSM-IV-TR), Vol. *1*). Washington, DC: American Psychiatric Association.

Andersson, G., Bergstrom, J., Hollondare, F., Carlbring, P., Kaldo, V., & Eskelius, L. (2005). Internet-based self-help for depression: Randomised controlled trial. *British Journal of Psychiatry, 187*, 456–461.

Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behavioral Therapy*, *38*, 196–205.

Andrews, G. (2001). Should depression be managed as a chronic disease? *British Medical Journal*, *322*, 419–421.

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS One*, *5*, e13196.

Button, K. S., Wiles, N. J., Lewis, G., Peters, T. J., & Kessler, D. (2012). Factors associated with differential response to online cognitive behavioural therapy. *Social Psychiatry and Psychiatric Epidemiology*, *47*, 827–833.

Cusin, C., Yang, H., Yeung, A., & Fava, M. (2009). Rating scales for depression. *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*. *121*, 7–35.

de Graaf, L., Hollon, S., Huibers, M. (2010). Predicting outcome in computerized cognitive behavioral therapy for depression in primary care: A randomized trial. *Journal of Consulting and Clinical Psychology*, *78*, 184–189.

Donker, T., Bennett, K., Bennett, A., MacKinnon, A., van Straten, A., Cuijpers, P., … & Griffiths, K. M. (2013). Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: Randomized controlled noninferiority trial. *Journal of Medical Internet Research*, *15*, e82.

Donker, T., Blankers, M., Hedman, E., Ljótsson, B., Petrie, K., & Christensen, H. (2015). Economic evaluations of Internet interventions for mental health: A systematic review. *Psychological Medicine*, *45*, 3357–3376.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders* (SCID-I/P) (Patient ed.). New York: New York: Biometrics Research, New York State Psychiatric Institute.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, *33*, 357–362.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213.

Hadjistavropoulos, H. D., Pugh, N. E., Nugent, M. M., Hesser, H., Andersson, G., Ivanov, G., …, & Austin, D. W. (2014). Therapist-assisted Internet-delivered cognitive behavior therapy for depression and anxiety: Translating evidence into clinical practice. *Journal of Anxiety Disorders*, *28*, 884–893.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62.

Hasin, D. S., Goodwin, R. D., Stinson, F. S., & Grant, B. (2005). Epidemiology of major depressive disorder: Results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Archives of General Psychiatry*, *62*, 1097–1106.

Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., Alaoui, S., Kraepelien, M., … & Lindefors, N. (2014). Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. *Journal of Affective Disorders*, *155*, 49–58.

Hedman, E., Ljótsson, B., & Lindefors, N. (2012). Cognitive behavior therapy via the Internet: A systematic review of applications, clinical efficacy and cost-effectiveness. *Expert Review of Pharmacoeconomics and Outcomes*, *12*, 745–764.

Hoifodt, R. S., Lillevoll, K. R., Griffiths, K. M., Wilsgaard, T., Eisemann, M., Waterloo, K., & Kolstrup, N. (2013). The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: Randomized controlled trial. *Journal of Medical Internet Research*, *15*, e153.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to denning meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Johansson, R., & Andersson, G. (2012). Internet-based psychological treatments for depression. *Expert Review of Neurotherapeutics*, *12*, 861–870.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., … & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976.

Kessler, R., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … & Walters, E. E. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, *289*, 186–187.

Kessler, D., Lewis, G., Kaur, S., Wiles, N., King, M., Weich, S., … & Peters, T. J. (2009). Therapist-delivered Internet psychotherapy for depression in primary care: A randomised controlled trial. *Lancet*, *374*, 628–634.

Kiluk, B. D., Sugarman, D. E., Nich, C., Gibbons, C. J., Martino, S., Rounsaville, B. J., & Carroll, K. M. (2011). A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: Toward improved standards for an emerging field. *American Journal of Psychiatry*, *168*, 790–799.

Kivi, M., Eriksson, M. C. M., Hange, D., Petersson, E. L., Vernmark, K., Johansson, B., & Björkelund, C. (2014). Internet-based therapy for mild to moderate depression in Swedish primary care: Short term results from the PRIM-NET randomized controlled trial. *Cognitive Behavioral Therapy*, *6073*, 1–10.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.

Meyer, B., Berger, T., Caspar, F., Beevers, C. G., Andersson, G., & Weiss, M. (2009). Effectiveness of a novel integrative online treatment for depression (Deprexis): Randomized controlled trial. *Journal of Medical Internet Research*, *11*, e15.

Newby, J. M., Mackenzie, A., Williams, A. D., McIntyre, K., Watts, S., Wong, N., & Andrews, G. (2013). Internet cognitive behavioural therapy for mixed anxiety and depression: A randomized controlled trial and evidence of effectiveness in primary care. *Psychological Medicine*, *43*, 2635–2648.

Perini, S., Titov, N., & Andrews, G. (2008). The climate sadness program of internet-based treatment for depression: A pilot study. *E-Journal of Applied Psychology*, *4*, 18–24.

Perini, S., Titov, N., & Andrews, G. (2009). Clinician-assisted Internet-based treatment is effective for depression: Randomized controlled trial. *Australian and New Zealand Journal of Psychiatry*, *43*, 571–578.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., … & Mann, J. J. (2011). The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*, *168*, 1266–1277.

Proudfoot, J., Goldberg, D., Mann, A., Everitt, B., Marks, I., & Gray, J. (2003). Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine*, *33*, 217–227.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. *2*). Thousand Oaks, CA: SAGE Publications.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*, 329–342.

Rush, A., First, M., & Blacker, D. (2008). Handbook of psychiatric measures *(2nd ed.)*. Washington, DC: American Psychiatric Publishing.

Ruwaard, J., Lange, A., Schrieken, B., Dolan, C. V., & Emmelkamp, P. (2012). The effectiveness of online cognitive behavioral treatment in routine clinical practice. *PLoS One*, *7*, e40089.

Ruwaard, J., Schrieken, B., Schrijver, M., Broeksteeg, J., Dekker, J., Vermeulen, H., & Lange, A. (2009). Standardized web-based cognitive behavioural therapy of mild to moderate depression: A randomized controlled trial with a long-term follow-up. *Cognitive Behavioral Therapy*, *38*, 206–221.

Spek, V., Nyklíček, I., Cuijpers, P., & Pop, V. (2008). Predictors of outcome of group and internet-based cognitive behavior therapy. *Journal of Affective Disorders*, *105*, 137–145.

Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., & Solley, K. (2010). Internet treatment for depression: A randomized controlled trial comparing clinician vs. technician assistance. *PLoS One*, *5*, e10939.

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. *Clinical Psychology Review*, *30*, 710–720.

Warmerdam, L., van Straten, A., Twisk, J., Riper, H., & Cuijpers, P. (2008). Internet-based treatment for adults with depressive symptoms: Randomized controlled trial. *Journal of Medical Internet Research*, *10*, e44.

Warmerdam, L., Van Straten, A., Twisk, J., & Cuijpers, P. (2013). Predicting outcome of Internet-based treatment for depressive symptoms. *Psychotherapy Research*, *23*, 559–567.

Williams, A., & Andrews, G. (2013). The effectiveness of Internet cognitive behavioural therapy (iCBT) for depression in primary care: A quality assurance study. *PLoS One*, *8*, e57447.

Zimmerman, M., Chelminski, I., & Posternak, M. (2004). A review of studies of the Hamilton depression rating scale in healthy controls: Implications for the definition of remission in treatment studies of depression. *Journal of Nervous and Mental Disease*, *192*, 595–601.

Zuithoff, N. P. A., Vergouwe, Y., King, M., Nazareth, I., van Wezep, M. J., Moons, K. G., & Geerlings, M. I. (2010). The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study. *BMC Family Practice*, *11*, 98. doi: 10.1186/1471-2296-11-98

How to cite this article: Rosso, I. M., Killgore, W. D. S., Olson, E. A., Webb, C. A., Fukunaga, R., Auerbach, R. P., Gogel, H., Buchholz, J. L., Rauch, S. L. (2016). Internet-based cognitive behavior therapy for major depressive disorder: a randomized controlled trial. *Depression and Anxiety*, XX, 1–10. DOI: 10.1002/da.22590

APPENDIX

METHODS

Randomization procedures

This was a single-center, parallel group RCT, with equal randomization to each group: iCBT and MAC. The website random.org was used to generate a list of 150 random integers (possible values 1 or 2) which were assigned sequentially and without any restriction to participants who participated in Day 2 visits. Participants in group "1" were assigned to the treatment group; whereas those with a value of "2" were assigned to the monitored attention control group. The random assignment sequence was in the sole possession of a RA in a password-protected document. After the Day 2 (baseline) HRSD assessment, the RA provided the randomly generated group assignment to the postdoctoral fellow responsible for registering patients to their assigned group. At the end of the Day 2 testing day, the postdoctoral fellow met with the participant and disclosed their group status. Thus, all Day 2 HRSD interviews (primary outcome) were conducted blind to treatment group status. Day 3 HRSD assessments were conducted by different clinicians who were also blind to treatment group.

Sample size and power

A priori estimation of targeted sample size was based on consideration of effect sizes from a previously published efficacy study of the same iCBT Sadness program in an Australian depressed sample (Titov et al., 2010). In that study, the between-group treatment effect sizes were large; the smallest effect was seen for differential change in K-10 scores between iCBT and waitlist, although it still corresponded to a large effect (Cohen's $d = .71$). Using the latter effect size, we determined that a sample of 30 trial completers in each group would provide at least 80% power for analyses of differences between two independent group differences (G*Power; (Faul et al., 2007)). Consequently, this was the sample size registered, in advance, at ClinicalTrials.gov (Identifier: NCT01598922). Participants were recruited from

17 October, 2012 through 28 August, 2015, and the trial ended when we reached our targeted sample size of 30 participants per group completing all study procedures. This corresponded to an ITT sample of 77, which was used to conduct an SPSS GENLIN analysis following multiple imputation of missing HRSD values. The GENLIN analysis ($n = 77$, alpha = .05, three covariates) had 93% power to detect a large main effect (Cohen's $f = 0.4$), and 80% power corresponded to detect a medium-to-large effect ($f = 0.32$) according to post hoc power analyses.

Contact with study staff

The weekly phone calls and support were delivered to both iCBT and MAC participants by Bachelor-level RAs with no formal clinical training. We used the telephone scripts generously provided by Gavin Andrews and *This Way Up* (www.thiswayup.org.au), in order to conduct the calls as done in the Australian RCT (Titov et al., 2010). The telephone calls were limited to 3–5 min each, and were designed to allow the participant to feel heard, supported, and engaged. The content could include statements of general encouragement (e.g., "You are making good progress through the program"), reflection ("It sounds like…"), normalizing (e.g., "Yes, learning these skills is not easy"), and prompts to check how the participant found the lesson and homework (e.g., "Did you complete the homework?"). During any period of the 3 years of recruitment, there were two to three RAs responsible for making these calls, with a maximum load of 16 patients per RA at peak periods (i.e., maximum 48–60 min spent on weekly phone calls). The RAs participated in weekly supervision with a doctoral-level psychologist.

Results

Clinical significance

In addition to examining HRSD response and remission criteria for assessing clinical significance, we examined the Jacobsen and Truax (1991) criteria: a decrease in HRSD scores was considered clinically significant if it placed the participant within two SDs of the mean of the normal population. We used data from a systematic review of studies of the HRSD in healthy adults (Zimmerman, Chelminski, & Posternak, 2004), where the mean and SD were both approximately equal to 3. Using this criterion, a posttreatment HRSD score of 9 points or less would be considered clinically significant. Accordingly, a clinically significant improvement was seen in 28(76%) of the iCBT group and 13(33%) of the MAC group. In a logistic regression analysis of pooled imputed data, this corresponded to a significant effect of treatment group ($\beta = -1.264$, SE = .557, OR = 3.541, 95% CI OR [1.185, 10.578], $P = .024$), controlling for age and sex.