A randomized noninferiority trial evaluating remotely-delivered stepped care for depression using internet cognitive behavioral therapy (CBT) and telephone CBT

David C. Mohr[a],[*], Emily G. Lattie[a], Kathryn Noth Tomasino[b], Mary J. Kwasny[a], Susan M. Kaiser[a], Elizabeth L. Gray[a], Nameyeh Alam[a], Neil Jordan[c],[d], Stephen M. Schueller[a]

[a] Department of Preventive Medicine, Center for Behavioral Intervention Technologies, Northwestern University, 750 N. Lakeshore Dr., 10th Floor, Chicago, IL, 60611, USA
[b] Department of Medicine, Northwestern University, NMH/Arkes Family Pavilion, Suite 1400, 676 N. Saint Clair St., Chicago, IL, 60611, USA
[c] Department of Psychiatry and Behavioral Sciences, Mental Health Services & Policy Program, Northwestern University Feinberg School of Medicine, 710 N Lake Shore Dr, 12th Flr, Chicago, IL, 60611, USA
[d] Center of Innovation for Complex Chronic Healthcare, Hines VA Hospital, 5000 S 5th Ave., Hines, IL, 60141, USA

ARTICLE INFO

ABSTRACT

This trial examined whether a stepped care program for depression, which initiated treatment with internet cognitive behavioral therapy, including telephone and messaging support, and stepped up non-responders to telephone-administered cognitive behavioral therapy (tCBT), was noninferior, less costly to deliver, and as acceptable to patients compared to tCBT alone. Adults with a diagnosis of major depressive episode (MDE) were randomized to receive up to 20 weeks of stepped care or tCBT. Stepped care (n = 134) was noninferior to tCBT (n = 136) with an end-of-treatment effect size of $d = 0.03$ and a 6-month post-treatment effect size of $d = -0.07$ [90% CI 0.29 to 0.14]. Therapist time in stepped care was 5.26 (SD = 3.08) hours versus 10.16 (SD 4.01) for tCBT (p < 0.0001), with a delivery cost difference of $-364.32 [95% CI $-423.68 to $-304.96]. There was no significant difference in pre-treatment preferences ($p = 0.10$) or treatment dropout (39 in stepped care; 27 in tCBT; $p = 0.14$). tCBT patients were significantly more satisfied than stepped care patients with the treatment they received (p < 0.0001). These findings indicate that stepped care was less costly to deliver, but no less effective than tCBT. There was no significant difference in treatment preference or completion, however satisfaction with treatment was higher in tCBT than stepped care.
Trial registration: clinicaltrials.gov Identifier: NCT01906476.

1. Introduction

Depression is prevalent and imposes a very high societal burden in cost, morbidity, quality of life, and mortality (Greenberg, Fournier, Sisitsky, Pike, & Kessler, 2015; Wells et al., 2002). While psychotherapy is both effective and acceptable to patients, a variety of barriers exist both to initiating and completing psychotherapy (Bedi et al., 2000; Mohr et al., 2010). Psychotherapy delivered remotely via telephone has been shown to increase treatment retention while being no less effective than face-to-face psychotherapy (Mohr et al., 2012; Mohr, Vella, Hart, Heckman, & Simon, 2008). Indeed, telephone psychotherapy is increasingly being used to overcome access barriers (Godleski, Darkins, & Peters, 2012; Turner, Brown, & Carpenter, 2018). However, increased treatment retention resulting from telephone psychotherapy may also increase treatment costs to healthcare organizations and payers.

Internet interventions, which combine digital tools with brief support from a coach or therapist, are another form of remote treatment. A large number of trials have consistently shown them to be effective at reducing depression (Andrews et al., 2018). As the evidence base for internet interventions accumulates, care systems are trying to understand how to integrate it into existing mental health services. The Dutch healthcare system, for example, has begun using blended care models, in which therapists can use internet treatments in combination with standard psychotherapy delivered in person or remotely. However, preliminary results suggest the use of internet interventions in blended care may actually increase therapist time and costs without improving outcomes (Kenter et al., 2015).

Stepped care has been proposed to be a solution for integrating technology-based treatments into existing care resources. However, the benefits of internet therapy as a low-intensity first step for treatment of

depression have not been tested (Cuijpers, Beekman, & Reynolds, 2012). Stepped care models typically have two features (Bower & Gilbody, 2005): they should be "least restrictive" in terms of cost and personal inconvenience, and they should be self-correcting. Thus, stepped care programs should initiate care with a low-intensity, less costly treatment, monitor outcomes systematically, and step patients up to more intensive and more costly treatments if criteria for improvement are not met. For example, stepped care protocols have been initiated treatment for depression with low-intensity psychoeducation, stepping up to a more intensive psychotherapy if a criterion such as dropping below a clinical threshold on a symptom severity measure is not met after a defined period (Araya et al., 2003; Bot, Pouwer, Ormel, Slaets, & de Jonge, 2010).

The aim of this study was to compare a remotely-delivered stepped care program to telephone administered cognitive behavioral therapy (tCBT) for the treatment of depression. The stepped care program initiated treatment using a guided internet CBT (iCBT), stepping up patients, who were non responsive, to tCBT. The use of tCBT as the comparison treatment, rather than face-to-face (Mohr et al., 2012; Mohr et al., 2008), also improves internal validity, as lower retention in face-to-face treatment could bias our findings in favor of our hypotheses. We evaluated three sets of hypotheses recommended for evaluating stepped care models (Bower & Gilbody, 2005): Stepped care would 1) be non-inferior to tCBT for depression outcomes; 2) have lower therapist costs, and; 3) be acceptable to patients.

2. Methods

2.1. Participants

Participants were recruited from February 23, 2015 through May 4, 2017 via online sources (e.g. ResearchMatch, Reddit) and healthcare systems in the United States (Group Health and Northwestern Medicine). Those recruited through healthcare systems were identified through electronic medical records. These recruitment methods and procedures, their effectiveness, and costs have been described in detail in two recent publications (Lattie et al., 2018; Palac et al., 2018).

Participants were included if they met criteria for major depressive episode (MDE) using the Mini International Neuropsychiatric Interview (Sheehan et al., 1998); had a score of 12 or greater on the Quick Inventory of Depressive Symptomatology – Clinician Rated (Rush et al., 2006); were aged 18 years or older, could speak and read English, and had a computer or tablet that would allow them to access a web intervention. Participants were excluded if they met diagnostic criteria for a severe psychiatric disorder, for which these treatments would be inappropriate, including psychotic disorders, bipolar disorder, or eating disorders; showed imminent risk of suicide in which both a plan and intent were reported using the Columbia Suicide-Severity Rating Scale (Posner et al., 2011); had visual or hearing impairments that would prevent participation; were receiving or planning to receive individual psychotherapy; or had initiated or modified antidepressant pharmacotherapy in the previous 14 days, to eliminate the effects of placebo response to medications. Participants receiving pharmacotherapy were permitted to continue their care and receive dose or medication adjustments. While alcohol or drug abuse or dependence were not exclusionary, potential participants who screened positive were further interviewed to assess their level of impairment. These audio-recordings were reviewed by the assessment supervisor; those whose impairment was judged to be severe enough that treatment for depression alone would be inappropriate were excluded. Excluded prospective participants were referred for appropriate treatment. All participants were consented and signed an online consent form in accordance with trial procedures that were approved by the Northwestern University institutional review board. The trial was monitored by an independent Data Safety Monitoring Board.

2.2. Treatments

Stepped care initiated treatment with guided iCBT and stepped up those participants who did not meet criteria for improvement to tCBT. The comparison condition was tCBT. Participants in both conditions were treated until they reached full remission, defined as a PHQ-9 < 5 over two consecutive weeks (Kroenke, Spitzer, & Williams, 2001), or for a maximum of 20 weeks, whichever came first. All patients completed a weekly PHQ-9 as part of their treatment, which was used to guide treatment decisions.

2.2.1. iCBT

iCBT used a highly interactive web-based intervention called ThinkFeelDo, which teaches basic CBT skills and has been piloted in several trials (Lattie et al., 2017; Schueller & Mohr, 2015; Tomasino, Lattie, Ho, et al., 2017). Didactic material was presented in "lessons" released 4 times per week. Two of the weekly lessons required approximately 10 min of time and include text and short animated video clips. The other 2 weekly lessons were brief summaries of previous material requiring 1–2 min to read. ThinkFeelDo included interactive tools supporting CBT skills (e.g., a calendaring tool for scheduling and monitoring pleasant events or a cognitive restructuring tool) and a coach interface that allowed the coach to see when and what the patient had accessed, the content of work with tools, and a secure messaging feature.

The iCBT coaching protocol was manualized (Tomasino, Lattie, Wilson, & Mohr, 2017) and based on the supportive accountability and efficiency models (Mohr, Cuijpers, & Lehman, 2011; Schueller, Tomasino, & Mohr, 2016). Coaches began with a 30–40 min engagement phone call, followed by weekly 10–15 min calls and 2–3 secure messages per week through the ThinkFeelDo site. After 3 weeks, patients could elect to move to messaging only or continuing to use telephone coaching on an as-needed basis.

Participants in iCBT were stepped up to tCBT if they met criteria that predicts non-response to iCBT and psychotherapy: PHQ-9 ≥ 17 from weeks 4–8, PHQ-9 ≥ 13 from weeks 9–13, or PHQ-9 ≥ 9 after week 13 (Schueller, Kwasny, Dear, Titov, & Mohr, 2015).

2.2.2. tCBT

tCBT used our manualized treatment based on a standard CBT approach, validated in numerous randomized trials and shown to be noninferior to face-to-face CBT (Mohr et al., 2012). Patients received a workbook, Beating Depression (Mohr, Ho., Duffecy, & Glazer-Baron, 2007), and spoke weekly with their assigned therapist for 45–50 min.

2.2.3. Therapists, training, and fidelity monitoring

Therapists included 5 PhD level clinical psychologists, 2 licensed clinical social workers, and a master's level therapist, all of whom were trained CBT therapists. Therapists provided both iCBT and tCBT and remained with the participant throughout treatment in both study arms.

tCBT training consisted of readings, 1 day of didactic training, weekly supervision on at least 3 cases until criteria for competence was met, which included a score of ≥ 40 on the Cognitive Therapy Scale (CTS) (Vallis, Shaw, & Dobson, 1986), and the supervisor's determination of competence. iCBT training consisted of listening to exemplary audiotaped coach calls and review of messaging logs, an additional 3–5 h of didactic training, review of the patient ThinkFeelDo site, and weekly individual supervision until the supervisor determined proficiency.

Supervision for tCBT and iCBT occurred separately. All phone therapy sessions in both iCBT and tCBT were audio-recorded. tCBT supervisors listened to at least one recording for each supervision session and rated it on the CTS. Supervision for therapists began weekly and was gradually reduced to not less than once monthly. The mean of the 112 therapist CTS rating scores was 50.8 (range = 40–61). iCBT

supervisors listened to call recordings and reviewed message logs. iCBT supervision occurred in a weekly group format, where at least one therapist's interactions with a participant were reviewed by reading messages and listening to the phone call if it occurred. As therapists began iCBT, they also received individual weekly supervision until the supervisor determined that they were competently following the protocol. Fidelity to iCBT was not monitored quantitatively. When iCBT participants stepped up to tCBT, supervision for the therapist also transferred to the tCBT supervisor.

2.3. Assessment

The primary clinical outcome was depressive symptom severity measured using the Quick Inventory of Depressive Symptomatology-Clinician version (QIDS) (Rush et al., 2006), an interviewer-based measure, administered by masked clinical evaluators. Comorbid diagnoses and MDE diagnosis as a secondary outcome were assessed using the MINI structured interview (Sheehan et al., 1998). All measures were administered at baseline, end of treatment (EOT), and at 3 and 6 months post-treatment. A mid-treatment assessment was also scheduled at week 10, however participants who remitted before week 10 would not have this mid-treatment assessment.

Bachelor-level research assistants, trained and supervised by a PhD-level psychologist, administered the MINI and QIDS by phone. Training included assigned readings from selected portions of the DSM-V (American Psychiatric Association, 2013), a 4-hour presentation on psychopathology, didactic training on MINI and QIDS administration, review of previously recorded assessments, and practice administrations. All interview assessments were audio-recorded. When the evaluator began assessing participants, the supervisor reviewed all recordings and provided individual feedback until competence was established (usually 5–10 assessments). Evaluators were supervised weekly as a group.

Time to remission (PHQ-9 > 5) was measured in weeks. Because the PHQ-9 was used to determine treatment decisions such as stepping and cessation of treatment, it was not used as an outcome measure in trial evaluation.

The primary delivery cost outcome was therapist cost. Therapists logged all time spent with a patient on phone sessions, messaging (iCBT), review of the dashboard and notes, outreach and call scheduling, and preparation. As these costs are intended to estimate the cost that might be expected in real-world implementation, we did not include time spent in training, supervision, and fidelity monitoring, as these are typically far more intensive in trials than in real-world care settings, where supervision is variable and often minimal. Therapist costs were calculated using an annualized psychologist wage of $80,510, based on the US Bureau of Labor Statistics 2017 estimates for Chicago. Consistent with previous work (Simon, Ludman, & Rutter, 2009), we added 28% fringe and 30% overhead, for a total annualized cost of $133,969, or an hourly cost of $74.43 based on 1800 work hours yearly.

Acceptability was explored at three times: pre-treatment preferences, adherence during treatment, and post-treatment satisfaction. Preference was evaluated using a previously employed method that provided a brief description of the treatments followed by a preference rating (Mohr et al., 2012). Non-adherence was defined as dropping out of treatment. Satisfaction was measured at the end of treatment using the Satisfaction Index for Mental Health Treatment (SIMH) measure, which has a possible range of 0–72, and a mean of 40 (SD = 13), based on a previous study of psychometric properties (Nabati, Shea, McBride, Gavin, & Bauer, 1998).

Baseline patient medication status was characterized by self-report. Dosing was categorized based on FDA recommendations (Education Medicaid Integrity Contractor, 2013).

2.4. Randomization and masking

Prior to the beginning of recruitment, a statistician produced a sequentially masked randomization scheme to the start of the trial, based on a 1:1 ratio, stratified by antidepressant medication status and therapist, with a block size of 4 within each stratum. To prevent allocation bias, all study personnel were masked to the randomization scheme. Clinical evaluators, who were masked to treatment assignment, enrolled and evaluated participants. Evaluators began each interview by asking participants not to say anything about the treatment they were receiving. After each assessment, evaluators reported if they had become unmasked, which occurred in 40 (1.7%) of the 2422 assessments. These participants were reassigned to another masked evaluator.

2.5. Statistical analyses

Noninferiority is established by showing that the true difference between two treatment arms is likely to be smaller than a prespecified noninferiority margin that separates clinically important from clinically negligible (acceptable) differences (D'Agostino, Massaro, & Sullivan, 2003; Nutt, Allgulander, Lecrubier, Peters, & Wittchen, 2008). Noninferiority trials of pharmaceuticals have used 30%–50% of the difference between treatment and control conditions to define noninferiority margins (Jones, Jarvis, Lewis, & Ebbutt, 1996; Nutt et al., 2008). A meta-analysis of CBT found an overall effect size of d = 0.82 (Cuijpers, Smit, Bohlmeijer, Hollon, & Andersson, 2010). We use face-to-face CBT rather than tCBT as the effect-size as this literature is far larger and therefore more reliable, and because there is no reason to believe that the effects of t-CBT are different from face-to-face CBT (Mohr et al., 2012). Using the midpoint of 40% for the acceptable criterion, we set d = 0.33 as the noninferiority criterion. A 1-sided test at a type I error rate of 5%, testing whether the difference in treatment groups is less than the noninferiority margin, is equivalent to testing whether a 2-sided 90% CI for the difference does not contain the noninferiority margin (Walker & Nowacki, 2011).

A sample of 115 per group provided 80% power to detect noninferiority using a one-sided, confidence interval for the difference in two independent means with type I error rate of 5% and a standardized margin of equivalence of 0.33. Assuming a 25% dropout rate, the study was designed to enroll 155 per treatment group.

Current literature suggests not directly imputing missing data for noninferiority analyses, but instead using mixed-effects models with repeated measures (Wiens & Rosenkranz, 2013). This method can estimate the difference between treatments at each time while accounting for observations that may be missing and adjusting for covariates. Hence, we used these models to estimate differences in primary outcomes (QIDS) between treatment allocation, assuming a random intercept and unstructured covariance structure. Additionally, we parameterized time as a nominal variable to not impose any structure to the trajectory of outcome over time. We provide the difference and 90% CI for treatment differences on the QIDS scale for interpretability. MDE, a secondary outcome, is binary and therefore used the effect size, h. All clinical outcome models were fit in SAS v.9.04. To verify that each arm produced reductions in symptoms, we evaluated the mean QIDS over time using mixed models with random intercepts and categorical indexing of time. Analyses for the primary clinical outcome was conducted using modified intention-to-treat, where all participants with at least one QIDS assessment after baseline were used in analyses. Additionally, as ITT analyses may increase the likelihood of falsely concluding non-inferiority, we performed a secondary analysis using a hybrid ITT/per-protocol (PP) approach, where participants who did not complete treatment were excluded. Non-trivial missing data was addressed using MLE-based ITT analysis via mixed models, as in our primary analysis (Matilde Sanchez & Chen, 2006).

Therapist cost differences and satisfaction across treatments were analyzed using *t*-tests. We calculated confidence intervals using
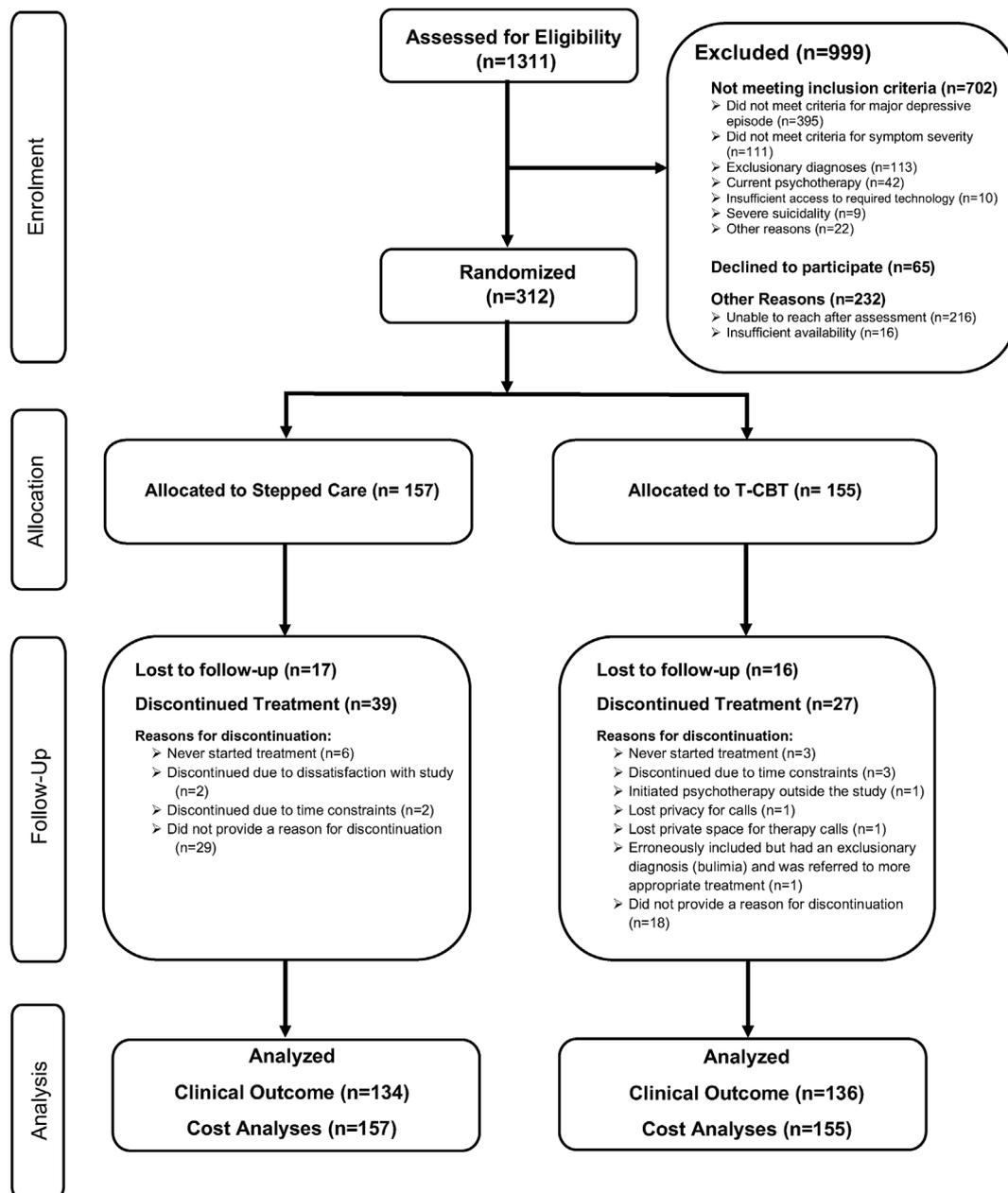
Fig. 1. Flow of participants through the study.

bootstrapping and Fieller's theorem (a probabilistic model) (Chaudhary & Stearns, 1996; Willan & O'Brien, 1996). Differences in acceptability were compared using logistic regression (to determine if dropout differed between groups), and linear regression was used to determine if satisfaction was different across groups after adjusting for covariates. Analyses of acceptability (treatment preferences, dropout, and satisfaction) and time to remission were analyzed using t-tests and chi-squared tests. These analyses were done in R v3.4.3.

3. Results

3.1. Participants

The flow of patients through the study is depicted in Fig. 1. A total of 270 (86.5%) of the 312 randomized had at least one QIDS assessment post baseline. There were no significant differences across treatment arms in lost-to-follow-up rates ($p = 0.99$). Table 1 summarizes the baseline demographics and characteristics of the randomized participants.

3.2. Depression outcomes

Group ratings by all interviewers on audiotaped QIDS were conducted revealing interrater reliabilities of 0.82. Fig. 2 displays the QIDS outcomes over time. Participants demonstrated clinically meaningful reductions in the QIDS at EOT in both stepped care ($\Delta = -6.8$ [95%CI = -7.6, -6.0]; $p < 0.0001$) and tCBT ($\Delta = -7.1$ [95%CI = -7.9, -6.2]). Participants remained improved at 6-month follow-up relative to baseline (ΔQIDS = -6.7 [95%CI = -7.6, -5.8] for stepped care; -6.5 [95%CI = -7.4, -5.7] for tCBT).

MDE also showed substantial reductions, with MDE at EOT of 15% in stepped care and 11% in tCBT, and 16% and 15%, respectively, at 6-month follow-up.

The primary outcome of QIDS showed little difference between groups (Fig. 3, Panel A). The effect size comparing QIDS between groups at EOT found an effect size of $d = 0.03$ (90%CI = -0.18, 0.24)

Table 1
Baseline characteristics of the intention-to-treat population.

	Stepped Care (157)	tCBT (155)
Age (Years) (median [IQR])	33.0 [27.0, 49.0]	32.0 [26.0, 45.5]
Sex (%)		
Female	118 (75.2%)	111 (71.6%)
Male	37 (23.6%)	44 (28.4%)
Missing	2 (1.3%)	0 (0.0%)
Race		
White	142 (90.4%)	133 (85.8%)
Black	8 (5.1%)	13 (8.4%)
Asian	6 (3.8%)	8 (5.2%)
Other	3 (1.9%)	5 (3.2%)
Ethnicity		
Non-Latino	141 (89.8%)	134 (86.5%)
Latino	12 (7.6%)	20 (12.9%)
Missing	4 (2.5%)	1 (0.6%)
Marital Status (%)		
Married/Partnered	70 (44.6%)	68 (43.9%)
Single	58 (36.9%)	67 (43.2%)
Separated/Divorced/Widowed	29 (18.5%)	20 (12.9%)
College		
High School or less	8 (5.1%)	7 (4.5%)
Some college	36 (22.9%)	42 (27.1%)
College degree	113 (72.0%)	106 (68.4%)
Household Income (median [IQR])	$53,500 [$30,500, $80,000]	$50,000 [$30,000, $82,000]
Antidepressant Use	68 (43.3%)	67 (43.2%)
Below minimal dosage	13 (19.4%)[a]	6 (9.0%)[a]
Minimal dosage	6 (9.0%)[a]	13 (19.4%)[a]
> Minimal to < 2x minimal dosage	9 (12.0%)[a]	7 (10.5%)[a]
≥2x minimal dosage	40 (59.7%)[a]	41 (61.2%)[a]
Baseline QIDS		
Mean (SD)	14.81 (2.42)	14.95 (2.57)
Median (IQR)	14.0 (13.0, 16.0)	14.0 (13.0, 17.0)
Baseline PHQ-9		
Mean (SD)	16.54 (3.66)	16.61 (3.97)
Median (IQR)	16.0 (14.0, 19.0)	17.0 (14.0, 19.0)
Comorbid Diagnoses (%)		
Generalized Anxiety Disorder	75 (47.8)	58 (37.4)
Social Anxiety Disorder	32 (20.4)	32 (20.6)
Panic Disorder	19 (12.3)	14 (8.9)
Obsessive-compulsive Disorder	12 (7.7)	5 (3.2)
PTSD	19 (12.1)	12 (7.7)
Agoraphobia	36 (22.9)	31 (20.0)
Alcohol Abuse/Dependence	9 (5.7)	5 (3.2)
Drug Abuse/Dependence	4 (2.5)	2 (1.3)
Number of comorbid diagnoses (%)		
0	46 (29.3)	54 (34.8)
1	60 (38.2)	53 (34.2)
2	31 (19.7)	30 (19.4)
3	11 (7.0)	10 (6.5)
≥4	9 (5.7)	8 (5.1)
Treatment Preferences		
Stepped Care	48 (30.6%)	36 (23.2%)
Telephone Cognitive Behavioral Therapy	53 (33.8%)	62 (40.0%)
No preference	56 (35.7%)	57 (36.8%)

[a] . Percentages for dosage are among those receiving antidepressant medications.

with an absolute QIDS difference of 0.13 (90%CI = −0.75, 1.01). The effect size at 6-month post-treatment was $d = -0.07$ (90%CI = −0.29, 0.14) with an absolute QIDS difference of −0.35 (90% CI = −1.35, 0.65). In the hybrid ITT/PP analysis, we excluded 66 (24%) of the 270 participants who discontinued treatment (Fig. 4, Panel A). The EOT effect size was $d = 0.01$ (90%CI = −0.21, 0.22) with an absolute QIDS difference of 0.03 (90%CI = −0.89, 0.94). The effect size at 6-month post-treatment was $d = -0.12$ (90%CI = −0.34, 0.10) with an absolute QIDS difference of −0.58 (90% CI = −1.63, 0.48). Model coefficients are provided in an online supplemental table.

The EOT effect size for MDE (Fig. 3, Panel B) was $h = 0.12$ (90%CI = −0.03, 0.27) with an absolute difference in rates of 3.9%

(90%CI = −3.1, 10.9). At 6 months post-treatment, the effect size for difference in MDE was $h = 0.03$ (90%CI = −0.12, 0.18) with an absolute difference in rates of 1.0% (90%CI = −6.7, 8.6). In the hybrid ITT/PP analysis (Fig. 4, Panel B), the EOT effect size was $h = 0.03$ (90%CI = −0.13, 0.18) with an absolute difference in rates of 0.8% (90%CI = −6.0, 7.6). At 6 months post-treatment, the effect size for difference in MDE was $h = -0.11$ (90%CI = −0.27, 0.05) with an absolute difference in rates of −3.7% (90%CI = −11.4, 4.0). Model coefficients are provided in an online supplemental table.

3.3. Stepping

Within the stepped care arm, a total of 36 (23%) stepped up to tCBT; 19 (53% of steppers) in weeks 4–8, 15 (42%) in weeks 9–13, and 2 (5%) after week 13.

3.4. Time to remission

There was no difference in the number of participants completing treatment prior to 20 weeks due to reaching the criterion for full remission [56 (36%) for stepped care; 56 (36%) for tCBT; 95% CI = −0.10, 0.11; $p = 1.0$] and no significant difference in the average number of weeks to completion due to full remission [9.5 (SD = 3.5) weeks for stepped care; 10.5 (SD = 3.5) for tCBT; 95% CI = −0.04, 0.18; $p = 0.10$].

3.5. Therapist cost

The average total therapist time per stepped care patient was 5.26 (SD = 3.08) hours compared to 10.16 (SD = 4.01) hours for tCBT ($p < 0.001$). Using the average hourly rate of $74.43, this translated into an average cost of $391.81 (SD = $229.06) per stepped care patient and $756.13 (SD = $298.64) per tCBT patient. Using probabilistic methods, the average difference between the cost to deliver stepped care and tCBT was $-364.32 [95% CI $-423.68, $-304.96]. Bootstrapping gave similar results, with an average difference of $-364.47 [95% CI $-422.89, $-305.07].

3.6. Acceptability

There was no significant difference in pre-randomization treatment preferences (Table 1; $p = 0.10$). There was no evidence that receiving or not receiving one's preferred treatment was significantly associated with change from baseline outcomes for any outcomes ($ps < 0.61$).

In stepped care, 39 (25%) participants dropped out from treatment, compared to 27 (17%; $p = 0.14$) in tCBT. We note that among the 36 participants (23%) who stepped up, only 1 participant dropped out at that point, failing to make the transition from iCBT to tCBT.

Participants receiving tCBT had significantly higher satisfaction scores (M = 62.3, SD = 7.58; N = 124) compared to stepped care (M = 57.4, SD = 10.61, N = 130; $p < 0.0001$).

3.7. Safety

One tCBT participant was briefly hospitalized for panic symptoms, which was determined by the data and safety monitoring board to be unrelated to treatment. There were no other adverse events.

4. Discussion

This is the first study to examine the integration of an internet intervention into a stepped care model for the treatment of depression. Stepped care processes using standard, non-technology-based treatments have generally been shown to be effective for depression (van Straten, Hill, Richards, & Cuijpers, 2015). This study extends this body of work, demonstrating that a stepped care program for depression, in
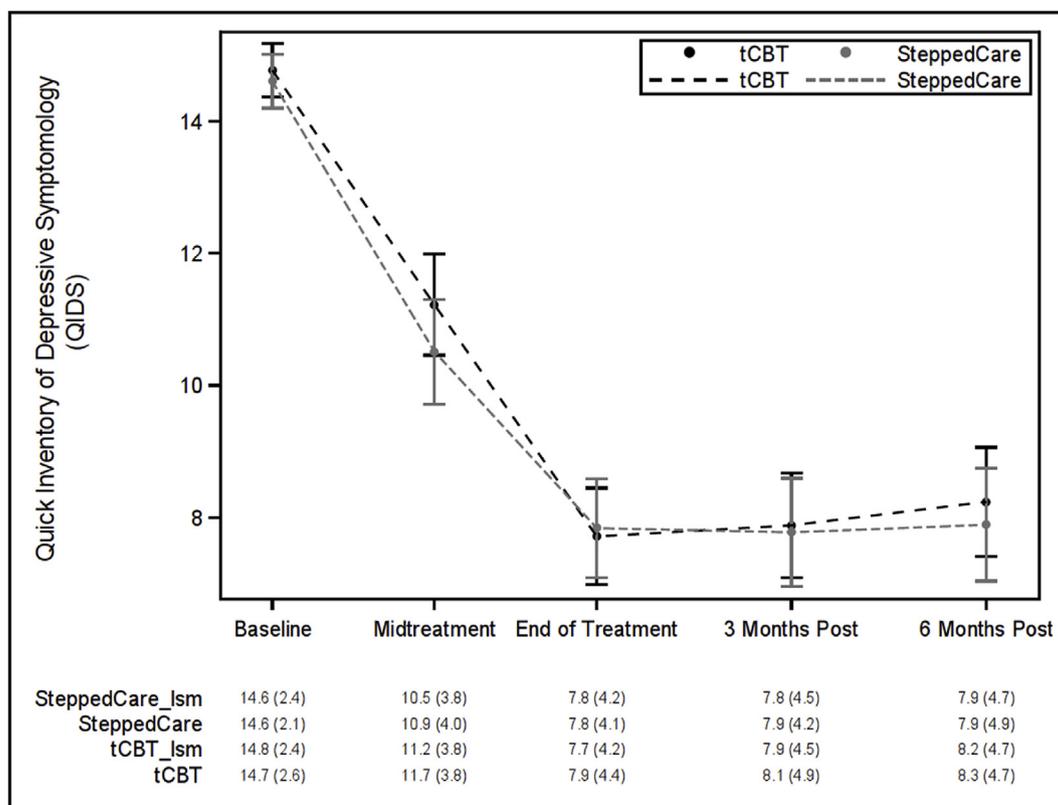
Fig. 2. QIDS over time by randomized group with least squares estimated and actual means (SDs)

Legend:

QIDS = Quick Inventory of Depressive Symptoms

tCBT = Telephone administered cognitive behavioral therapy

lsm = least squares model.

which patients begin with therapist-supported iCBT and move to tCBT if not meeting criteria for improvement, can be as effective as initiating treatment with tCBT. Both stepped care and tCBT achieved a reduction of approximately 7 points on the QIDS, which is equivalent to 7–8 points on the Hamilton Rating Scale for Depression and 10 points on the Beck Depression Inventory, indicating a substantial and meaningful reduction in symptom severity (Rush et al., 2003). Stepped care achieved these outcomes with approximately half the therapist time and costs as tCBT.

Stepped care was initially acceptable to patients, with no difference across treatment arms in pre-treatment preferences. Treatment completion rates were high and generally equivalent. This is notable, given that adherence rates for iCBT, even when accompanied by therapist support, can be lower than is commonly seen in tCBT (van Ballegooijen et al., 2014). It is possible that the comparatively high rates of adherence to iCBT in this trial were due to the website design (which utilized briefer more frequent interactions), the use of the telephone as a therapist contact medium (which was similar across treatment arms), or reassurance that participants would receive more standard care if iCBT did not work for them. However, while patients' satisfaction ratings with stepped care were good, they were still significantly lower than for patients who received tCBT. Care organizations may want to consider this finding in any stepped care implementation.

There have been two other trials that have used an internet treatment as part of a stepped care protocol. One trial examining stepped care for depression in primary care used a complex procedure that included multiple steps, with multiple options within each step. Treatment was initiated with active monitoring. The second step allowed patients to choose between bibliotherapy, an uncoached iCBT program, or telephone-administered psychotherapy, while the third step included outpatient psychotherapy or pharmacotherapy, and a fourth step combined psycho- and pharmacotherapy in out- or inpatient settings (Harter et al., 2018). While there was a significant effect for stepped care relative to treatment as usual in primary care, no data are presented on the proportion of patients who received any specific treatment (including iCBT or telephone-therapy), making it difficult to compare to the present study.

The other stepped care trial, which compared iCBT and face-to-face CBT to face-to-face CBT alone for panic or social anxiety disorders, found no significant difference between the two treatment arms (Nordgreen et al., 2016). There were some notable differences in methodology. It was neither designed nor analyzed as a non-inferiority trial. The study used an additive stepped care design (2 weeks for psychoeducation, stepping up non-responders to 9 weeks of iCBT followed by 10 weeks of CBT) for a potential total of 21 weeks, while the CBT condition was 12 weeks, confounding treatment arm with potential length of treatment and treatment. Nonetheless, similar to our study, the differences in outcomes were small, the stepped care program required substantially less therapist time, and treatment dropout was not high.

While our findings indicate that a stepped care approach to integrating iCBT may reduce therapist costs in a telemental health service, we offer some considerations in extending these findings to stepping from iCBT to face-to-face CBT. The stepping process is a sensitive period in which dropouts can occur (van Straten et al., 2015). We saw very little dropout at stepping. We expect that stepping from remotely delivered iCBT to face-to-face treatment would result in a higher dropout rate, consistent with the Nordgreen study (Nordgreen et al., 2016). Thus, we urge some caution in generalizing these findings to face-to-face treatment settings.
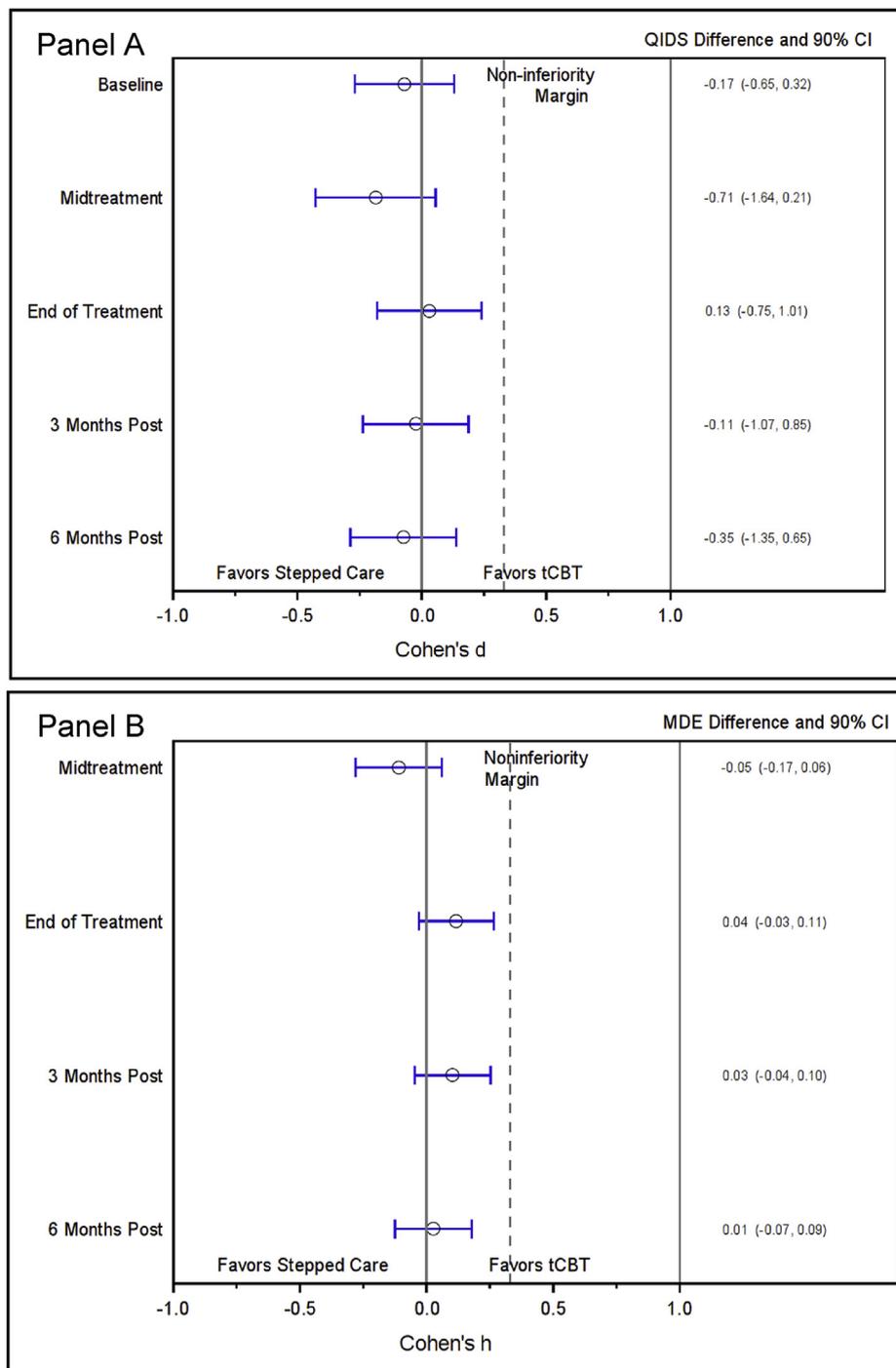
Fig. 3. Modified intention to treat effect size (Cohen's d & h) differences and actual differences between groups. Panel A: QIDS; Panel B: MDE.
Legend
MDE = Major Depressive Episode
QIDS = Quick Inventory of Depressive Symptoms
tCBT = Telephone administered cognitive behavioral therapy.

While this study was not integrated into a healthcare setting, our findings may have relevance for implementation. Stepped care utilized approximately half the amount of time required by tCBT due to the use of iCBT, and thus may extend the capacity of mental health professionals. However, by using less time, iCBT increases the number of patients a therapist must manage through briefer interactions. In our experience this resulted a few unique challenges. During brief calls, therapists had less time to reorient to a patient's problems and issues. Therapists reported that voice quality was often a helpful cue in

orienting the therapist to the patient. Messaging, which does not provide non-verbal cues, generally required some review of logged messages and notes to remind the therapist of patient's issues. These issues notwithstanding, therapists did not find the caseload or patient management burdensome in this trial.

iCBT has been successfully implemented as standalone eHealth clinics in Australia, Sweden, Denmark, Norway, and Canada, funded by national or regional healthcare systems (Titov et al., 2018). There have been numerous attempts at integrating digital mental health into value-
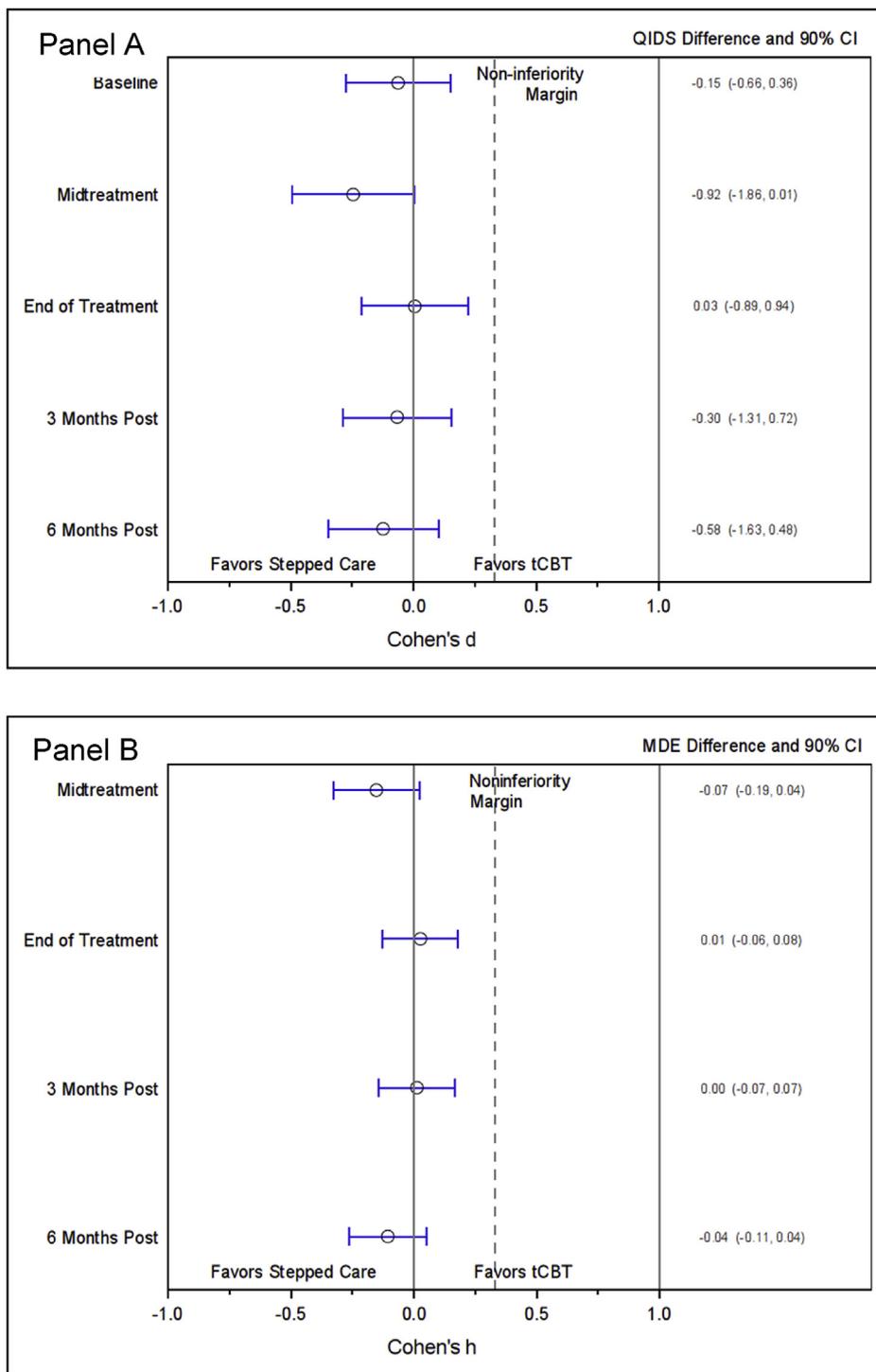
Fig. 4. Per protocol effect size (Cohen's d & h) differences and actual differences between groups. Panel A: QIDS; Panel B: MDE.
Legend
MDE = Major Depressive Episode
QIDS = Quick Inventory of Depressive Symptoms
tCBT = Telephone administered cognitive behavioral therapy.

based care systems in the United States, however these have not re-sulted, to the best of our knowledge, in successful, sustainable im-plementations (Bertagnoli, 2018; Bertagnoli et al., 2015; Mohr, Weingardt, Reddy, & Schueller, 2017). Problems have included diffi-culty integrating digital services into existing services, low compliance by healthcare workers, and low patient uptake. A growing number of studies have shown that more positive expectations about the value of

digital mental health by patients and providers, suggesting that edu-cation and marketing may improve uptake by patients and providers (Mira et al., 2019; Vis et al., 2018). However, greater efforts on the design of the services, technologies, and implementation plans to in-tegrate digital mental health into care systems and patient's lives will also likely be required (Mohr, Riper, & Schueller, 2018).

This trial has several strengths. It kept the treatment time consistent

across treatment arms, eliminating the biases that can occur when stepped care programs add treatments sequentially, which can extend time in treatment and the trial for stepped care relative to the comparator. We were able to retain a large portion of treatment dropouts in follow-up assessments, reducing the likelihood of bias associated with dropout.

There are also several limitations and caveats that should be considered in interpreting these data. First, patients in this sample may not be representative of real-world patients as research processes such as recruitment methods, screening, and consenting may select for a more motivated sample. This selection bias may have produced stronger adherence and engagement than would be seen in real-world settings. Second, as this was a noninferiority trial, we cannot rule out the possibility that the improvements seen in this sample were due to factors other than the treatments, such as measurement drift, regression to the mean, the natural course of the disorder, patient expectations and placebo effects, or any pharmacotherapy or adjustments to their pharmacotherapy they may have received. Third, we only looked at cost of therapist time, and not at a broad range of costs, such as medical costs or patient costs. Potential software costs were not included in the cost analysis, as there is no consistent pricing model. Fourth, the iCBT program was supported by telephone and messaging contact with a therapist; these findings should not be generalized to uncoached iCBT, which likely has much lower adherence and effectiveness (Karyotaki et al., 2017). Fifth, while we monitored fidelity through supervision, we did not conduct quantitative monitoring for iCBT. We note that since this study was conducted, an iCBT fidelity monitoring tool has been developed and validated (Hadjistavropoulos, Schneider, Klassen, Dear, & Titov, 2018). Finally, we did not observe any evidence that iCBT prior to stepping was any less effective than tCBT, raising the question as to whether iCBT alone would have performed as well. Given the relatively high rates of treatment dropout commonly seen in iCBT alone, and the 23% of patients who stepped up, we believe this is unlikely. However, we cannot rule out this possibility in the present study.

As internet treatments continue to demonstrate efficacy in a growing number of trials (Andrews et al., 2018), healthcare organizations are trying to understand how to integrate them into processes of care. This study indicates that use of a guided iCBT program, coupled with measurement-based stepping procedures to trigger transition to a standard manual-based tCBT program, can be a less costly, but no less effective method of providing treatment for depression than tCBT alone.

Funding

Declaration of competing interest

Dr. Mohr has accepted speaking fees from Apple Inc. and the American Psychological Association and has an ownership interest in Actualize Therapy. Dr. Schueller serves as a scientific advisor to Joyable, Inc. and receives stock options in Joyable. None of the other authors have any conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2019.103485.

References

American Psychiatric Association, D.-T. F.. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). Arlington, VA, US: American Psychiatric Publishing, Inc.

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., et al. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders, 55*, 70–78.

Araya, R., Rojas, G., Fritsch, R., Gaete, J., Rojas, M., Simon, G., et al. (2003). Treating depression in primary care in low-income women in santiago, Chile: A randomised controlled trial. *Lancet, 361*, 995–1000.

van Ballegooijen, W., Cuijpers, P., van Straten, A., Karyotaki, E., Andersson, G., Smit, J. H., et al. (2014). Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS One, 9*, e100674.

Bedi, N., Chilvers, C., Churchill, R., Dewey, M., Duggan, C., Fielding, K., et al. (2000). Assessing effectiveness of treatment of depression in primary care. Partially randomised preference trial. *British Journal of Psychiatry, 177*, 312–318.

Bertagnoli, A. (2018). Digital mental health: Challenges in implementation. *American psychiatric association* (New York, NY).

Bertagnoli, A., Trangle, M., Marx, L., Lacoutture, E., Rukanonchai, D., & Schuster, J. (2015). Behavioral health. *Annual meeting of the alliance of community health plans* (Washington, DC).

Bot, M., Pouwer, F., Ormel, J., Slaets, J. P., & de Jonge, P. (2010). Predictors of incident major depression in diabetic outpatients with subthreshold depression. *Diabetic Medicine, 27*, 1295–1301.

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry, 186*, 11–17.

Chaudhary, M. A., & Stearns, S. C. (1996). Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Statistics in Medicine, 15*, 1447–1458.

Cuijpers, P., Beekman, A. T., & Reynolds, C. F., 3rd (2012). Preventing depression: A global priority. *Journal of the American Medical Association, 307*, 1033–1034.

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *British Journal of Psychiatry, 196*, 173–178.

D'Agostino, R. B., Sr., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine, 22*, 169–186.

Education Medicaid Integrity Contractor (2013). Antidepressant medications: U.S. Food and drug administration-approved indications and dosages for use in adults. *CMS Medicaid integrity program*.

Godleski, L., Darkins, A., & Peters, J. (2012). Outcomes of 98,609 U.S. Department of Veterans Affairs patients enrolled in telemental health services, 2006-2010. *Psychiatric Services, 63*, 383–385.

Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *Journal of Clinical Psychiatry, 76*, 155–162.

Hadjistavropoulos, H. D., Schneider, L. H., Klassen, K., Dear, B. F., & Titov, N. (2018). Development and evaluation of a scale assessing therapist fidelity to guidelines for delivering therapist-assisted Internet-delivered cognitive behaviour therapy. *Cognitive Behaviour Therapy, 47*, 447–461.

Harter, M., Watzke, B., Daubmann, A., Wegscheider, K., Konig, H. H., Brettschneider, C., et al. (2018). Guideline-based stepped and collaborative care for patients with depression in a cluster-randomised trial. *Scientific Reports, 8*, 9389.

Jones, B., Jarvis, P., Lewis, J. A., & Ebbutt, A. F. (1996). Trials to assess equivalence: The importance of rigorous methods. *BMJ, 313*, 36–39.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., et al. (2017). Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: A meta-analysis of individual participant data. *JAMA Psychiatry, 74*, 351–359.

Kenter, R. F., van de Ven, P. M., Cuijpers, P., Koole, G., Niamat, S., Gerrits, R. S., et al. (2015). Costs and effects of Internet cognitive behavioral treatment blended with face-to-face treatment: Results from a naturalistic study. *Internet Interventions, 2*, 77–83.

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

Lattie, E. G., Ho, J., Sargent, E., Tomasino, K. N., Smith, J. D., Brown, C. H., et al. (2017). Teens engaged in collaborative health: The feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. *Internet Interventions, 8*, 15–26.

Lattie, E. G., Kaiser, S. M., Alam, N., Tomasino, K. N., Sargent, E., Rubanovich, C. K., et al. (2018). A practical do-it-yourself recruitment framework for concurrent eHealth clinical trials: Identification of efficient and cost-effective methods for decision making (Part 2). *Journal of Medical Internet Research, 20*, e11050.

Matilde Sanchez, M., & Chen, X. (2006). Choosing the analysis population in non-inferiority studies: Per protocol or intent-to-treat. *Statistics in Medicine, 25*, 1169–1181.

Mira, A., Soler, C., Alda, M., Banos, R., Castilla, D., Castro, A., et al. (2019). Exploring the relationship between the acceptability of an internet-based intervention for depression in primary care and clinical outcomes: Secondary analysis of a randomized controlled trial. *Frontiers in Psychiatry, 10*, 325.

Mohr, D. C., Cuijpers, P., & Lehman, K. (2011). Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. *Journal of*

Medical Internet Research, 13, e30.

Mohr, D. C., Ho, J., Duffecy, J., & Glazer-Baron, K. (2007). Beating depression: Cognitive behavioral therapy patient workbook. *Northwestern university*. Northwestern University.

Mohr, D. C., Ho, J., Duffecy, J., Baron, K. G., Lehman, K. A., Jin, L., et al. (2010). Perceived barriers to psychological treatments and their relationship to depression. *Journal of Clinical Psychology, 66*, 394–409.

Mohr, D. C., Ho, J., Duffecy, J., Reifler, D., Sokol, L., Burns, M. N., et al. (2012). Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: A randomized trial. *Journal of the American Medical Association, 307*, 2278–2285.

Mohr, D. C., Riper, H., & Schueller, S. M. (2018). A solution-focused research approach to achieve an implementable revolution in digital mental health. *JAMA Psychiatry, 75*, 113–114.

Mohr, D. C., Vella, L., Hart, S., Heckman, T., & Simon, G. (2008). The effect of telephone-administered psychotherapy on symptoms of depression and attrition: A meta-analysis. *Clinical Psychology: Science and Practice, 15*, 243–253.

Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three problems with current digital mental health research . . . and three things we can do about them. *Psychiatric Services, 68*, 427–429.

Nabati, L., Shea, N., McBride, L., Gavin, C., & Bauer, M. S. (1998). Adaptation of a simple patient satisfaction instrument to mental health: Psychometric properties. *Psychiatry Research, 77*, 51–56.

Nordgreen, T., Haug, T., Ost, L. G., Andersson, G., Carlbring, P., Kvale, G., et al. (2016). Stepped care versus direct face-to-face cognitive behavior therapy for social anxiety disorder and panic disorder: A randomized effectiveness trial. *Behavior Therapy, 47*, 166–183.

Nutt, D., Allgulander, C., Lecrubier, Y., Peters, T., & Wittchen, U. (2008). Establishing non-inferiority in treatment trials in psychiatry: Guidelines from an expert consensus meeting. *Journal of Psychopharmacology, 22*, 409–416.

Palac, H. L., Alam, N., Kaiser, S. M., Ciolino, J. D., Lattie, E. G., & Mohr, D. C. (2018). A practical do-it-yourself recruitment framework for concurrent eHealth clinical trials: Simple architecture (Part 1). *Journal of Medical Internet Research, 20*, e11049.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., et al. (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.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., et al. (2006). An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: A sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry, 59*, 493–501.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., et al. (2003). The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*, 573–583.

Schueller, S. M., Kwasny, M. J., Dear, B. F., Titov, N., & Mohr, D. C. (2015). Cut points on the Patient Health Questionnaire (PHQ-9) that predict response to cognitive-behavioral treatments for depression. *General Hospital Psychiatry, 37*, 470–475.

Schueller, S. M., & Mohr, D. C. (2015). Initial field trial of a coach-supported web-based depression treatment. *The proceedings of pervasive health '15*Turkey: IEEE. Istanbul.

Schueller, S. M., Tomasino, K. N., & Mohr, D. C. (2016). Integrating human support into behavioral intervention technologies: The efficiency model of support. *Clinical Psychology: Science and Practice, 27*–45.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33 quiz 34-57.

Simon, G. E., Ludman, E. J., & Rutter, C. M. (2009). Incremental benefit and cost of telephone care management and telephone psychotherapy for depression in primary care. *Archives of General Psychiatry, 66*, 1081–1089.

van Straten, A., Hill, J., Richards, D. A., & Cuijpers, P. (2015). Stepped care treatment delivery for depression: A systematic review and meta-analysis. *Psychological Medicine, 45*, 231–246.

Titov, N., Dear, B., Nielssen, O., Staples, L., Hadjistavropoulos, H., Nugent, M., et al. (2018). ICBT in routine care: A descriptive analysis of successful clinics in five countries. *Internet Interventions, 13*, 108–115.

Tomasino, K. N., Lattie, E. G., Ho, J., Palac, H. L., Kaiser, S. M., & Mohr, D. C. (2017a). Harnessing peer support in an online intervention for older adults with depression. *American Journal of Geriatric Psychiatry, 25*, 1109–1119.

Tomasino, K. N., Lattie, E. G., Wilson, R. E., & Mohr, D. C. (2017b). *Coaching manual for the ThinkFeelDo internet intervention program.* Chicago, IL: Northwestern University.

Turner, J., Brown, J. C., & Carpenter, D. T. (2018). Telephone-based CBT and the therapeutic relationship: The views and experiences of IAPT practitioners in a low-intensity service. *Journal of Psychiatric and Mental Health Nursing, 25*, 285–296.

Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54*, 381–385.

Vis, C., Mol, M., Kleiboer, A., Buhrmann, L., Finch, T., Smit, J., et al. (2018). Improving implementation of eMental health for mood disorders in routine practice: Systematic review of barriers and facilitating factors. *JMIR Mental Health, 5*, e20.

Walker, E., & Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine, 26*, 192–196.

Wells, K., Miranda, J., Bauer, M., Bruce, M., Durham, M., Escobar, J., et al. (2002). Overcoming barriers to reducing the burden of affective disorders. *Biological Psychiatry, 52*, 655.

Wiens, B. L., & Rosenkranz, G. K. (2013). Missing data in noninferiority trials. *Statistics in Biopharmaceutical Research, 5*, 383–393.

Willan, A. R., & O'Brien, B. J. (1996). Confidence intervals for cost-effectiveness ratios: An application of fieller's theorem. *Health Economics, 5*, 297–305.