Published in final edited form as:

Cogn Behav Ther. 2009 December ; 38(4): 222–234. doi:10.1080/16506070802675353.

Randomized Effectiveness Trial of an Internet, Pure Self-Help, Cognitive Behavioral Intervention for Depressive Symptoms in Young Adults

Greg Clarke, Chris Kelleher, Matt Hornbrook, Lynn DeBar, John Dickerson, and Christina Gullion

Kaiser Permanente Center for Health Research (CHR), Portland, Oregon

Abstract

This study evaluated an Internet-delivered, cognitive behavioral skills training program versus a treatment-as-usual (TAU) control condition targeting depression symptoms in youth ages 18 to 24. Potential participants were mailed a recruitment brochure; if interested they accessed the study website to complete an online consent and baseline assessment. Intervention participants could access the website at their own pace and at any time. Reminder postcards were mailed periodically to encourage return use of the intervention. The pure self-help intervention was delivered without contact with a live therapist. The primary depression outcome measure was the Patient Health Questionnaire (PHQ-8), administered at 0, 5, 10, 16, and 32 weeks after enrollment. A small but significant between-group effect was found from week 0 to week 32 for the entire sample ($n=160$; $d=.20$, 95% CI=0.00-0.50), with a moderate effect among females ($n=128$; $d=.42$, 95% CI=0.09-0.77). Greater depression reduction was associated with two measures of lower website usage, total minutes and total number of page hits. While intervention effects were modest, they were observed against a background of substantial TAU depression pharmacotherapy and psychosocial services. Highly disseminable, low-cost, and self-help interventions such as this have the potential to deliver a significant public health benefit.

Keywords

cognitive behavioral treatment; pure self-help; treatment-as-usual; randomized trial; effectiveness; college-age

Introduction

Interventions for mental disorders such as depression have been created for delivery through the Internet (Andersson et al., 2005; Christensen et al., 2004; Christensen et al., 2006; Clarke et al., 2005; Seligman et al., 2005), typically offering variants of evidence-based treatments (EBTs), such as cognitive-behavioral therapy (CBT), originally created for use as traditional psychotherapies. Reviews and meta-analyses suggest that small to moderate effects are achievable with these Internet interventions (Andersson, 2006; Griffiths & Christensen, 2006; Pull, 2006; Spek et al., 2007a; Wantland et al., 2004; Ybarra & Eaton, 2005). Larger effects have generally been obtained when the Internet component is paired with brief, in-person, telephone or email contacts with a counselor or behavioral coach (Andersson, 2006; Palmqvist et al., 2007; van Straten, Cuijpers, & Smits, 2008). This approach is often referred

to as “guided self-help” (Fairburn & Carter, 1997; Perkins et al., 2006) or “minimal contact therapy” (Newman et al., 2003).

Alternatively, a “pure” self-help approach, also known as “self-administered therapy” (Newman et al., 2003), delivers only the Internet psycho-educational materials with no accompanying contact with therapists. While these pure self-help interventions have generally yielded smaller effects than those obtained with guided self-help Internet programs, some pure self-help interventions have produced moderate intervention effects (Spek et al., 2007b). Pure self-help programs have the potential advantage of much lower marginal costs per patient—although studies have yet to examine the cost-effectiveness of either guided or pure self-help Internet programs (Palmqvist et al., 2007).

We previously conducted two trials of a pure self-help Internet CBT program for depression in adults (Clarke et al., 2002; Clarke et al., 2005). Both trials compared the intervention to a strong treatment-as-usual (TAU) control condition. Nearly all participants in both experimental and control conditions received non-experimental TAU depression services, primarily antidepressant medication and traditional psychotherapy. The experimental intervention provided tutorials and exercises in evidence-based cognitive therapy (Beck et al., 1979). Consistent with the pure self-help format, participants did not have any contact with a research therapist. The first trial failed to find any significant program effect, possibly due to very low use of the intervention website (Clarke et al., 2002). The second trial (Clarke et al., 2005) found a modest effect when the intervention was paired with either brief telephone or postal reminders to return and use the website. Despite the small but positive effects of the second trial, we concluded that this older intervention -- developed in 1999 -- failed to provide the depth of content that might yield more potent effects.

Therefore, we developed a new, pure self-help depression intervention with significantly more interactivity, a behavioral therapy tutorial (Lewinsohn et al., 1984), and updated cognitive therapy elements from the earlier intervention. This report provides the results of our first trial of the new intervention with depressed older adolescents and young adults, ages 18 to 24. We focused on this sample because depression incidence rises rapidly during this age period (Kessler & Walters, 1998). We hypothesized that the improved intervention, paired with postal reminders, would yield a significantly better self-reported depression response rate compared to a TAU control condition.

We also examined participant utilization of non-experimental health care. We did not necessarily expect any “utilization offset” effects of the intervention, but we were interested in whether different patterns of use or service mix might emerge. For example, use of the website might result in lower use of specialized mental health services compared to use of primary care. However, given that very large samples are typically required to detect health care utilization/cost effects (Lynch & Clarke, 2006), we did not expect to be adequately powered for these exploratory analyses. Nonetheless, even non-significant trends in service mix patterns might help formulate testable hypotheses for future studies.

Methods

Setting, Sample, and Recruitment

We conducted the study in a health maintenance organization (HMO) with over 450,000 members in the northwest USA. The Human Subjects Committee for the HMO approved study procedures.

In 2005 we employed the HMO’s electronic medical record to identify two recruitment groups: a “depressed” group ages 18 to 24 ($n=4,396$), who received depression medication or

psychotherapy in the previous 6 months and/or had a chart diagnosis of depression; and an roughly equal-sized “nondepressed” group of the same ages ($n=4,578$), who did not receive such services and did not have an chart diagnosis of depression, but who did show elevated health care utilization. We included the latter group because persons with elevated health care utilization are more likely to have unrecognized depression (Smith et al., 2005).

We mailed a recruitment brochure to each potential participant, explaining the study and providing the study website address. Interested participants confirmed their identity at the website, then completed the online informed consent and the baseline self-report depression measure: the 8-item Patient Health Questionnaire (PHQ-8) (Lowe et al., 2004; Spitzer et al., 1999).

Consenting participants were randomly assigned by a website program to one of two conditions. Randomization was 1:1 with no stratification or blocking. Participants assigned to the TAU control group ($n=77$) were not granted access to the intervention but were instead linked to an HMO website that provided static information about depression but no interactive skills training. Participants in the intervention condition ($n=83$) were given immediate access to the intervention. The Consort Figure (Figure 1) provides a summary of the study design and participant progression through key events. Participants in both arms of the study were permitted to continue or initiate TAU health care services for depression.

Assessment

All participants in both study conditions followed the same assessment schedule, at enrollment (week 0) and 5, 10, 16, and 32 weeks after enrollment. All participants were prompted by email reminders to return to the password-protected assessment portion of the study website, followed by a brief telephone reminder if they failed to respond to email reminders. Reminder calls were made by non-clinician research staff who limited interactions to non-therapeutic topics such as completing assessments and resetting passwords. Participants received an electronic \$10 Amazon gift certificate for completing each assessment. The study assessment website provided participants in both conditions with links to the non-research, HMO psychiatric emergency services.

Our primary depression outcome measure was the Patient Health Questionnaire-8 (PHQ-8), a self-report version of the clinician-responder PRIME-MD (Spitzer et al., 1994). The PHQ-8 is a version of the PHQ-9 (Ackermann et al., 2005; Kroenke & Spitzer, 2002) with the suicide/self-harm item removed. The psychometric properties of the PHQ-8 are nearly identical to those of the complete PHQ-9 (Corson et al., 2004); it possesses excellent sensitivity (.97), specificity (.97), and positive predictive value (.75) for detecting major depression (Kroenke & Spitzer, 2002).

We collected each participant’s HMO health services utilization data using the HMO administrative data systems for up to one year after each participant’s completion of the baseline assessment questionnaire.

Intervention

The experimental intervention is an Internet-based program for young adults who are coping with depression. This pure self-help site is “unattended” in that it is not staffed by live personnel. Instead, it provides self-guided, interactive behavioral (Lewinsohn et al., 1975) and cognitive therapy (Beck et al., 1979) tutorials to help users overcome depression. The website programming offers significant interactivity and tailoring that distinguishes it from static, traditional bibliotherapy sites that deliver identical content to every user. The stand-alone aspect of the program distinguishes it from guided self help programs where the Internet content

is augmented by mental health professionals conducting person-to-person psychotherapy or counseling through e-mail, telephone, or in-person exchanges with patients.

The website consists of four main sections. One section, called “Measure Your Mood,” permits users to complete a brief, auto-scored depression scale and review their graphically displayed depression scores over time. During our study, this rating was completed as desired by the experimental condition participants and thus was not considered an outcome measure. A second section, called “Facts about Depression,” offers general information pages about depression (its symptoms, causes, treatments) and associated mental and emotional problems. A third, “Journal,” section permits users to record their thoughts and concerns in a private online journal or diary. Users can electively “publish” journal entries for viewing by other website users, subject to review by research staff for inappropriate content.

The fourth and final section, called “Improve Your Mood,” provides users with brief, interactive tutorials in cognitive restructuring and behavioral therapy methods (users can choose one or both tutorials). We consider this section to be the most “curative” element of the website. The behavioral therapy tutorial provides users with personalized feedback that simulates how behavioral therapy for depression is delivered in conventional, face-to-face sessions. For example, a website tutorial guides users through identifying 8 to 12 pleasant activities that matter most to their mood (Figure 2). The website then assists with creating a personalized self-contract to make small but consistent increases in the frequency of these activities. Users are prompted to return to the website every few days to record total daily activities and a daily mood rating (on a 7-point Likert scale) for each intervening day. Website algorithms process these tracking data to generate personalized feedback regarding the association (or lack of) between each user’s daily mood and his or her pleasant activity level. The website identifies those activities that users report occurring infrequently or not at all, or those that are not associated with increased mood, and suggests that these might be replaced with other, more mood-lifting activities.

The “Thought Helper” function is a representative element of the cognitive restructuring module. After a tutorial on the basics of cognitive restructuring, a user can type a personal negative or irrational thought into a text search box. The program searches a predefined list of 300+ common negative thoughts and displays the best matches with the negative thought submitted by the user. Each user selects the negative thought closest to their original, and the program then returns a list of several possible realistic counter-thoughts relevant to that belief. Users are encouraged to create a personalized counter-thought using relevant portions of the provided examples and enter it into the website for storage. Later, a user can retrieve these personal counter-thoughts, unrealistic beliefs, and activating situations.

Intervention Exposure

Participants randomized to the intervention condition were able to use the program at any time during the study duration. Each received a U.S. mail postcard reminder at 2, 8, and 13 weeks after enrollment. Postcards highlighted a different feature of the website at each mailing, designed to entice the participant to make a return visit. Most website use was concentrated in the first four weeks after enrollment (25% of sessions within 1 week, 50% of sessions within 4 weeks, and 75% of visits within 15 weeks). The site collected detailed information about frequency of user access to the site, the duration and “flow” of these visits, and the collection of answers to online questions. We created several measures of each subject’s use and exposure to the program (“dose”) to examine whether individuals with greater exposure to the program improved to a greater degree.

Analysis Plan

The planned primary analysis was a between-groups comparison of change over repeated measures in the PHQ-8 score. We conducted intent-to-treat analyses; that is, all participants randomized were analyzed in the arm to which they were randomized and without regard for actual participation in the intervention. As noted above, intervention participants were encouraged to continue returning to the CBT website throughout the study.

The study assessment schedule involved five repeated measures of depression levels for each participant from enrollment (week 0) through the final assessment at week 32. Participants could respond at a varying number of measurement occasions, possibly resulting in intermittent missing data as well as some permanent dropout. Also the mathematical assumptions of Gaussian least squares models, such as homogenous variances and uncorrelated measures over time, are often violated in repeated measures data (Littell et al., 1998). In view of these considerations, we planned to use generalized mixed modeling, with residual maximum likelihood estimation (REML) to deal with incomplete data and a structured covariance matrix to adjust for correlations between timepoints. REML is more efficient and less biased than widely used alternatives, such as the method of last observations carried forward (LOCF) (Gibbons et al., 1993). Other authors have published excellent articles on the application of mixed models to repeated measures (Brown et al, 2008; Fidler et al 2008; Singer, 1998; Littell et al 1998).

The specific model we used was a hierarchical mixed model, in which the repeated measures data from baseline to week 32 were nested within participant, the intercept (participant effect) was treated as a random effect and treatment group as a fixed effect. This is also known as a random coefficients model. The test of treatment group difference in change is a test of the difference in slope parameters over time (i.e., the interaction between group and time). Covariation within subjects over time was modeled as a spatial power structure, to adjust for the unequal spacing between assessments. Data analyses were conducted using the SASTM System for Windows, Release 9.1.3, with mixed modeling in PROC MIXED.

We evaluated the treatment effect in a series of models, beginning with the primary analysis (all participants) and continuing with examination of the effect in specified subgroups selected on the basis of prior studies reporting more pronounced effects in these groups (Clarke et al., 2002; 2005). These subgroups were: (a) females, (b) participants with high PHQ-8 score at baseline (≥ 10), and (c) participants enrolled from the “depressed” recruitment sample. In all of these, we included age and race as planned covariates. Gender was included as a covariate in the latter two subgroup models.

We also evaluated whether the form of the trend over time was linear, quadratic or cubic. A linear form assumes that the rate of change is constant over time points, whereas quadratic and cubic forms provide estimates of variation in the rate of change over time. We report results from the best model for each subgroup.

We report the “net” treatment effect size, i.e., the difference between arms in the model-based (i.e., adjusted) mean change from baseline to week 32, divided by the adjusted standard deviation (i.e., Cohen’s *d*). The mean change and standard deviation of change were calculated from the model-based, predicted participant scores at baseline and week 32. Because this effect size is based on predictions from a model of change over time, it implicitly takes into account the pattern of change over all interim assessments as well.

We evaluated the dose effect on PHQ-8 scores within the intervention group only using a hierarchical mixed modeling structure, with intercept as a random effect and the selected time-varying dose measure as a fixed effect. For each of our dose measures, we evaluated the form

of the trend over time (linear, quadratic, or cubic). We present results from the best-fitting models. Our measures of dose were total number of website sign-ins, total number of “page hits” (i.e., mouse clicks), and total duration of website use (minutes). We constructed the variables as time-varying predictors by measuring cumulative exposure up to and including the day each PHQ-8 score was obtained. Because we could not track Internet site use outside the intervention website, we could not reliably determine the end of sessions when users failed to formally log out and instead navigated to another, nonstudy website. After reviewing patterns of web use across all available data, we chose to treat any web session as terminated when a single page had been viewed for more than 30 minutes. Elapsed time on the final page was censored to 30 minutes in calculating total website duration.

For health care utilization data, we used the chi-square test to evaluate hypotheses about differences between groups in the proportions of participants in each condition who had at least one instance of each type of health care service. We also conducted logistic regression analyses to test the hypothesis that use of each type of health care service was associated with study condition or baseline PHQ depression score.

Results

Subject Flow

The CONSORT Flow Chart (Figure 1) summarizes participant progression through key study events. Brochures were sent to a total 8,974 HMO members, 163 registered on the website, and 160 completed the consent and baseline assessment and were randomized: 83 to the intervention condition and 77 to the TAU control condition. Table 1 provides baseline demographics by experimental condition.; This was a predominantly female, white sample, with no significant differences across study arms on baseline measures, suggesting that randomization produced comparable groups.

The “depressed” recruitment group yielded 109 enrolled participants ($n=109/4396$ or 2.5% of the mailed invitations to this group), with a mean baseline PHQ-8 score of 10.7 ($SD=5.6$). The “high utilization” recruitment group yielded 51 enrolled participants ($51/4578$ or 1.1% of the total mailed invitations to this group), with a mean baseline PHQ-8 score of 6.9 ($SD=3.9$) ($t = -4.4$, $p < .01$, $r^2 = 0.11$). Table 1b presents baseline demographics for the “depressed” and “high health care utilization” recruitment groups from which participants were enrolled. As expected, baseline depression rate (PHQ-8 score 10 or higher) also differed between the two recruitment groups ($X^2 = 16.6$, $p < .01$), with participants from the “depressed” recruitment population having a higher rate (56.0%; $61/109$) than those from the “high utilization” population (21.6%; $11/51$) ($OR=4.6$, 95% $CI: 2.1-10.0$). There were no significant differences in race, age, or gender across participants recruited from these different populations.

Follow-up assessment completion rates are shown in Figure 1. None of the follow-up completion measures was related to experimental condition, gender, baseline depression scores, or recruitment population. Older participants were somewhat more likely than younger participants to complete follow-up assessments (Odds Ratio = 1.15 for an age increase of one year, 95% $CI = 1.01-1.31$), but this difference was not related to experimental condition.

Intervention Exposure

Table 2 (overall exposure) and Table 3 (exposure by intervention content area) summarize the intervention exposure in the experimental arm of the study. One participant ($1/83$, 1.2%) did not return after the randomization visit but was kept in analyses per the ITT analysis plan. Users spent the greatest time on the “Improve Your Mood” section (a median of 25 minutes per user)

where they viewed an average of almost 60% of the pages offered. This suggests that the putatively curative elements had the greatest attraction for enrolled users.

Main Effects

Table 4 presents the model-based predicted PHQ-8 depression scores by time, with p value for the treatment effect on rate of improvement over time (group-by-quadratic time interaction) for the total sample and each subsample examined. Both the linear and quadratic components of trend over time were significant, and the addition of a cubic component did not improve variance accounted for in any model.

A significant treatment effect ($t=-1.97$, error $df=442$, $p=.05$) favoring the CBT condition was obtained for the total sample (all enrolled participants), with a small effect size of Cohen's $d=.20$ (95% CI = 0.00-0.50) for the model-based treatment effect on change from week 0 to week 32. Figures 3a and 3b shows the actual and model-predicted PHQ-8 means over time, respectively, for participants in the two study arms, overlaid with the slopes that were the subject of the primary efficacy test.

Dose Effects

In the experimental arm only, greater depression symptom reduction on the PHQ-8 was associated with fewer minutes of website usage ($F[1,221]=5.84$, $p=.02$) and with fewer page hits ($F[1,221]=8.35$, $p=.004$); in both cases the linear model was the best fitting. We did not find an association between PHQ-8 outcomes and the total number of sign-ons. The third dose variable, number of log-ons, was not significantly associated with outcomes. This variable had much lower variability and a more restricted range than either page hits or duration of website use. This restricted range may partly account for the log-on measure's lack of association with outcome. We conducted additional post-hoc analyses to better understand the counterintuitive dose results. We first divided the intervention participants into two groups using the median total time of website usage. Participants using the website more intensively had higher average baseline PHQ-8 scores (11.2; $SD=6.1$) compared to participants who used the website less intensively (7.6; $SD=4.1$) ($t=3.18$, $p<.01$, $r^2=0.11$); that is, participants who were most depressed at baseline were more likely to continue using the website throughout the study period.

Figure 4 presents minutes of website usage for participants who had improved ($PHQ-8 < 9$) and who still had significant depression ($PHQ-8 \geq 10$) among the subset of participants randomized to the intervention who had a baseline PHQ-8 score of 10 or greater (in order to detect improvement). Although at three of the four follow-up assessments (weeks 5, 16, 32) participants who had *not* improved ($PHQ-8 \geq 10$) appear to have used the website much more often, at no follow-up point were the minutes of website use significantly different between the two groups, probably due to the relatively small sample.

Subgroup Effects

We also report results for the subsamples of female participants and persons with higher depression scores at baseline ($PHQ-8$ total score ≥ 10), as previous research suggested better outcomes among these subgroups (Clarke et al., 2002; Clarke et al., 2005). We found an improved significant group $\times$ time effect for the female-only population ($DF=363$, $t=-1.96$, $p=0.05$, with a between-group effect size of $d=0.42$ (95% CI=0.09-0.77) across the entire study period (weeks 0 to 32). We did not find a significant group $\times$ time treatment effect for the male-only population ($p=0.74$), although this may have been due to small numbers of enrolled males (TAU had 61 females and 16 males, intervention had 67 females and 16 males). We also did not find a significant group $\times$ time treatment effects within any of the other subgroups (Table 4).

Health Care Utilization

Table 5 summarizes participant use of non-experimental HMO health care services, by condition, over the 12 months post-enrollment. There were no significant differences between study conditions. Slightly more than half of the sample obtained some form of mental health care (medication and/or specialty mental health visits) during this time.

Discussion

We found a small but significant between-group, or net effect size ($d = .20$) for model-based change from baseline to the final follow-up assessment at week 32. Among female participants we found a moderate net effect size for the same time frame ($d = .42$). This is similar to effects obtained in a previous trial testing an earlier version of this intervention (Clarke et al., 2005), except that in the present study the intervention and TAU control conditions appear to converge by the final follow-up point, primarily as a consequence of improving depression scores in the TAU condition.

Between-group or net effects of this size are typically classified as small (Cohen, 1988). This may be an accurate indicator of the potency of this intervention relative to other self-help CBT programs such as that of Spek and colleagues (2007b) which yielded a larger pre-post, between-group effect size ($d = .55$) at the end of acute treatment. However, the between-group effect size is not just a function of the potency of the experimental intervention, but is also a function of the magnitude of change observed in the control condition. Detecting the effects of an experimental intervention relative to a strong TAU or alternative treatment control condition typically yields smaller between-group effects than comparing against a waitlist or no-treatment control condition (Kazdin & Bass, 1989). In the present study our results were obtained when compared against a background of treatment-as-usual (TAU) received by both experimental and control group participants. Over 50% of all participants received traditional depression pharmacotherapy and/or psychotherapy (Table 5), and 71.6% [78/109] of those from the “depressed” recruitment group received this type of service (there was no difference between study arms). Therefore, it is difficult to directly compare the between-group effect size obtained in this study with the larger effects obtained in trials of self-help internet CBT program conducted by Andersson et al. (2005) and Spek et al (2007b) which employed waitlist control conditions. Neither approach is right or wrong; both types of study designs are ultimately necessary in this progression of research. However, they represent different ends of the efficacy (controlled) and effectiveness (real-world) spectrum.

Further context for these findings comes from meta-analyses comparing traditional, face- to-face CBT with other active treatments—as was the case in this study. These studies have found an average between-group, pre-post treatment effect of $d = .27$ favoring CBT (Gaffan et al., 1995; Gloaguen et al., 1998), only slightly larger than the comparable effect obtained in this trial ($d = .20$).

The counter-intuitive association between lower dose and greater improvement has been observed before in other mental health trials, especially where treatment duration (dose) is not fixed or randomized but is determined by the patients themselves (Barkham et al., 2006; Feaster, Newman, & Rice, 2003). Our post-hoc analyses are consistent with, but not proof of, the explanation that participants who improved more rapidly found the website less necessary and thus discontinued treatment early. In contrast, those persons with more persistent depression may have continued using the website for longer periods in the hopes of obtaining relief (Barkham et al., 2006).

This study had several limitations. First, this sample of 160 was enrolled as a pilot study not formally powered to detect main effects. A post-hoc power analysis suggests that, had we

assumed an intervention effect size of $d = .25$ based on an earlier study (Clarke et al., 2005), we would have needed a sample of 216 to achieve power of .80 as recommended by Cohen (1988), assuming a 2-tailed $\alpha=.05$, compound symmetry covariance structure, and $\pi=.07$. Clearly these results must be replicated in a larger study. The small sample was even more of a limitation for the health care outcomes, as the detection of experimental impacts on health care services generally requires thousands of cases (Lynch & Clarke, 2006).

A further limitation was our reliance on the PHQ-8 as our sole outcome measure. While it is a reasonable instrument with demonstrated ability to discern treatment outcome effects, future trials should include outcomes beyond depression, and a modest number of predictors, moderators, and mediators.

By the final (32-week) assessment, PHQ-8 scores were essentially indistinguishable for the two conditions, in part as a result of improving scores among the TAU participants for the last two assessment points. Most experimental participants were not using the website with any frequency after week 15; the majority of participants' exposure to the intervention was concentrated in the early weeks immediately after enrollment. While we employed some low-intensity methods to encourage return visits to the website (postcards mailed to users' homes), in future studies we should consider more intensive methods (e.g., live telephone contacts from therapists) to encourage ongoing use of the website and CBT skills, especially in the latter months.

Several other Internet-enabled CBT interventions for depression have obtained positive effects (reviewed by Griffiths & Christensen, 2006; Ybarra & Eaton, 2005). Of these, the most well-researched are the MoodGYM and BluePages programs (Christensen et al., 2004; Christensen et al., 2006), the Internet version of the Coping with Depression course (Spek et al., 2007b), and Andersson's (2005) program. Different trial designs, control conditions, samples, and outcome measures complicate direct effect comparisons of these interventions to ours. However, large, simple, practical trials (March et al., 2005; Tunis et al., 2003) directly comparing these different programs could be conducted entirely over the Internet, with few limits on participant qualification, with a simplified design and outcome measures. These trials would enable the enrollment of very large samples quickly and at a relatively low cost, thus powering the ability to detect small differences in outcome.

Widely disseminable, low-cost, self-help interventions for depression have the potential to deliver a significant public health benefit. Although the average depression effect is likely to be relatively modest for any one patient, these interventions may be offered to thousands of persons for a marginal additional cost of pennies for each additional recipient and thus hold the potential to improve the depression status of whole populations.

Acknowledgments

We wish to thank Kevin Rogers, Lynnette Rogers, and Stephanie Hertert for their assistance. This trial is registered at ClinicalTrials.gov with Identifier NCT00145054.

Reference List

- Ackermann RT, Rosenman MB, Downs SM, Holmes AM, Katz BP, Li J, et al. Telephonic case-finding of major depression in a Medicaid chronic disease management program for diabetes and heart failure. *General Hospital Psychiatry* 2005;27:338–343. [PubMed: 16168794]
- Andersson G. Internet-based cognitive-behavioral self help for depression. *Expert Review of Neurotherapeutics* 2006;6:1637–1642. [PubMed: 17144778]

- Andersson G, Bergstrom J, Hollandare F, Carlbring P, Kaldø V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. *British Journal of Psychiatry* 2005;187:456–461. [PubMed: 16260822]
- Barkham M, Connell J, Stiles WB, Miles JN, Margison F, Evans C, et al. Dose-effect relations and responsive regulation of treatment duration: the good enough level. *Journal of Consulting & Clinical Psychology* 2006;74:160–167. [PubMed: 16551153]
- Beck, AT.; Rush, AJ.; Shaw, BF.; Emory, G. *Cognitive therapy of depression*. Guilford Press; New York: 1979.
- Brown, CH.; Costigan, TE.; Kendziora, KT. Data analytic frameworks: Analysis of variance, latent growth, and hierarchical models. In: Nezu, AM.; Nezu, CM., editors. *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. Oxford University Press; New York: 2008. p. 285-313.
- Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. *British Medical Journal* 2004;328:265. [PubMed: 14742346]
- Christensen H, Griffiths KM, Mackinnon AJ, Brittcliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. *Psychological Medicine* 2006;36:1737–1746. [PubMed: 16938144]
- Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. *Journal of Medical Internet Research* 2002;4:E14. [PubMed: 12554545]
- Clarke GN, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. *Journal of Medical Internet Research* 2005;7:e16. [PubMed: 15998607]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates; Hillsdale, New Jersey: 1988.
- Corson K, Gerrity MS, Dobscha SK. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *American Journal of Managed Care* 2004;10:839–845. [PubMed: 15609737]
- Fairburn, CG.; Carter, JC. Self-help and guided self-help for binge-eating problems. In: Garner, DM.; Garfinkel, PE., editors. *Handbook of Treatment for Eating Disorders*. 2nd ed.. Guilford Press; New York: 1997. p. 494-500.
- Feaster DJ, Newman FL, Rice C. Longitudinal Analysis when the Experimenter does not Determine when Treatment Ends: What is Dose-Response? *Clinical Psychology and Psychotherapy* 2003;10:352–360. [PubMed: 16724159]
- Fidler, F.; Faulkner, C.; Cumming, G. Analyzing and presenting outcomes. In: Nezu, AM.; Nezu, CM., editors. *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. Oxford University Press; New York: 2008. p. 315-334.
- Gaffan EA, Tsaousis I, Kemp-Wheeler SM. Researcher allegiance and meta-analysis: the case of cognitive therapy for depression. *Journal of Consulting & Clinical Psychology* 1995;63:966–980. [PubMed: 8543719]
- Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. Application to the NIMH treatment of Depression Collaborative Research Program dataset. *Archives of General Psychiatry* 1993;50:739–750. [PubMed: 8357299]
- Gloaguen V, Cottraux J, Cucherat M, Blackburn IM. A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders* 1998;49:59–72. [PubMed: 9574861]
- Griffiths KM, Christensen H. Review of randomised controlled trials of Internet interventions for mental disorders and related conditions. *Clinical Psychologist* 2006;10:16–29.
- Kazdin AE, Bass D. Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting & Clinical Psychology* 1989;57:138–147. [PubMed: 2925964]
- Kessler RC, Walters EE. Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. *Depression & Anxiety* 1998;7:3–14. [PubMed: 9592628]

- Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals* 2002;32:1–7.
- Lewinsohn, PM.; Biglan, A.; Zeiss, AM. Behavioral treatment of depression. In: Davidson, PO., editor. *The behavioral management of anxiety, depression and pain*. Brunner/Mazel; New York: 1975. p. 91-146.91-146
- Lewinsohn, PM.; Steinmetz, JL.; Antonuccio, DO.; Teri, L. A behavioral group therapy approach to the treatment of depression. In: Upper, D.; Ross, SM., editors. *Handbook of behavioral group therapy*. Plenum; New York: 1984. p. 303-329.
- Littell RC, Henry PR, Ammerman CB. Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science* 1998;76:1216–1231. [PubMed: 9581947]
- Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders* 2004;81:61–66. [PubMed: 15183601]
- Lynch FL, Clarke GN. Estimating the economic burden of depression in children and adolescents. *American Journal of Preventive Medicine* 2006;31:S143–151. [PubMed: 17175409]
- March JS, Silva SG, Compton S, Shapiro M, Califf R, Krishnan R. The case for practical clinical trials in psychiatry. *American Journal of Psychiatry* 2005;162:836–846. [PubMed: 15863782]
- Newman MG, Erickson T, Przeworski A, Dzus E. Self-help and minimal-contact therapies for anxiety disorders: Is human contact necessary for therapeutic efficacy? *J Clinical Psychologist* 2003a; 59:251–274.
- Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? *Expert Review of Pharmacoeconomics & Outcomes Research* 2007;7:291–297.
- Perkins SJ, Murphy R, Schmidt U, Williams C. Self-help and guided self-help for eating disorders. *Cochrane Database Systematic Review* 2006;3 CD004191.
- Pull CB. Self-Help Internet Interventions for Mental Disorders. *Current Opinion Psychiatry* 2006;19:50–53.
- Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. *American Psychologist* 2005;60:410–421. [PubMed: 16045394]
- Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* 1998;24:323–355.
- Smith RC, Gardiner JC, Lyles JS, Sirbu C, Dwamena FC, Hodges A, et al. Exploration of DSM-IV criteria in primary care patients with medically unexplained symptoms. *Psychosomatic Medicine* 2005;67:123–129. [PubMed: 15673634]
- Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychological Medicine* 2007a;37:319–328. [PubMed: 17112400]
- Spek V, Nyklíček I, Smits N, Cuijpers P, Riper H, Keyzer J, Pop V. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. *Psychological Medicine* 2007b;37:1797–1806. [PubMed: 17466110]
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. Journal of the American Medical Association* 1999;282:1737–1744. [PubMed: 10568646]
- Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV III, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *Journal of the American Medical Association* 1994;272:1749–1756. [PubMed: 7966923]
- Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association* 2003;290:1624–1632. [PubMed: 14506122]
- van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. *Journal of Medical Internet Research* 2008;10:e7. [PubMed: 18364344]

- Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The Effectiveness of Web-Based vs. Non-Web-Based Interventions: A Meta-Analysis of Behavioral Change Outcomes. *Journal of Medical Internet Research* 2004;6:e40. [PubMed: 15631964]
- Ybarra ML, Eaton WW. Internet-based mental health interventions. *Mental Health Services Research* 2005;7:75–87. [PubMed: 15974154]

Figure 1.
CONSORT figure of study events

home | measure | improve | personal | BB | journals | facts | settings | CRISIS

1 2 3 4 5

Choose Fun Stuff – Create Your Personal Activities List

■ ■ ■ Use the arrows to add and remove activities to and from your Personal Activities List. You'll use this list later to help you do more fun activities. Please pick between 8 and 12 in any order.

300+ Fun Activities

- Downloading stuff from the internet
- Making snacks for yourself
- Getting new software
- Finding or listening to music on the internet
- Shopping online
- Singing to myself
- Eating a really good snack
- Restoring antiques, refinishing furniture
- Going to the movies by yourself
- Renting movies to watch alone
- Cooking meals for yourself
- Redecorating my personal space
- Being in a sporty or expensive car
- Getting a manicure or pedicure
- Staying up late

Personal Activities List

- Going shopping by yourself
- Buying stuff at internet auction sites
- Taking a relaxing bath
- Getting new computer hardware
- Going to a restaurant by yourself

Previous Next

CRISIS | Site Feedback | Site Help | Privacy | Security | Staff Entry

Figure 2.
Screenshot of behavioral therapy page

Figure 3.
Average PHQ-8 scores by week and randomization condition

Website Use by Recovery Status (PHQ $\leq$ 9 vs. $\geq$ 10)

Figure 4. Minutes of website usage among recovered (PHQ-8 $\leq$ 9) versus unrecovered (PHQ-8 $\geq$ 10) participants.

Table 1

Comparison of experimental condition on baseline demographics

	Intervention (n=83)		Control (n=77)		Significance*
	Mean or %	(SD)	Mean or %	(SD)	
Age	22.6	(2.3)	22.7	(2.7)	0.88
PHQ-8	9.4	(5.5)	9.7	(5.4)	0.67
Female	81%		79%		0.81
White	82%		81%		0.82
"Depressed" at baseline**	42%		48%		0.45
"Likely Depressed" case-finding***	68%		69%		0.85

* Participant age and PHQ-8 score were compared with an ANOVA. All other comparisons were made with chi-square analyses.

** PHQ8 score of 10 or higher.

*** Identified at case-finding as likely depressed through pharmacy and encounter data

Table 1b

Baseline demographics of participants enrolled from different recruitment pools

	Enrolled from depressed recruitment pool (n=109)		Enrolled from high utilization recruitment pool (n=51)	
	Mean or %	(SD)	Mean or %	(SD)
Age	22.6	(2.6)	22.6	(2.4)
PHQ-8	10.7	(5.6)	6.9	(3.9)
PHQ-8 $\geq$ 10 at baseline	56%		22%	
Female	79%		82%	
White	79%		86%	

* Statistical tests were not run comparing these groups as we designed the samples so that they would be similar on all measures except depressive symptoms.

Table 2

Frequency of website usage for intervention condition only (n=83), excluding randomization visit at week 0

	Median	Mean (SD)	Range
Number of Visits per Person	6	8.5 (14.2)	0-111
Total Time (minutes)	52	115.1 (176.1)	0-1088
Total Hits	48	84.8 (112.9)	0-742

Table 3

Cumulative website usage by major component, for intervention condition only

	Median	Mean (SD)	Range
<i>Total Time per Visit by Content Area (minutes)</i>			
Improve Your Mood*	25	35.8 (58.7)	0 - 463
Facts about Depression	2	10.2 (26.9)	0 - 200
Measure Mood	6	10.0 (13.5)	0 - 68
Journal	5	43.7 (100.6)	0 - 692
<i>% of Web Pages Visited per Visit by Content Area</i>			
Improve Your Mood (of 24 pages)	58.3%	38.6% (29.4%)	0 - 92%
Facts about Depression (of 56 pages)	5.4%	9.4% (15.3%)	0 - 79%
Measure Mood (of 3 pages)	100.0%	64.3% (43.8%)	0 - 100%
Journal (of 7 pages)	28.6%	31.2% (28.3%)	0 - 100%

* 18 of 83 intervention participants (21.7%) did not use the "Improve Your Mood" content area

Table 4
Model-predicted self-reported depression outcomes (PHQ8) for the total sample and selected subsamples

Group	Study Condition	N	Baseline		5-Week		10-Week		16-Week		32-Week		P value*	Overall Effect Size*
			Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)		
Total Sample	Intervention	83	10.0	(0.8)	9.1	(0.7)	8.5	(0.8)	8.0	(0.8)	8.6	(0.8)	.05	0.20
	Control	77	10.3	(0.8)	10.1	(0.7)	10.0	(0.7)	9.7	(0.8)	9.1	(0.8)		
Baseline PHQ ≥ 10	Intervention	35	14.3	(0.9)	12.5	(0.8)	11.2	(0.9)	10.3	(0.9)	12.0	(1.1)	.11	0.25
	Control	37	14.2	(1.0)	13.4	(0.9)	12.8	(0.9)	12.4	(1.0)	13.2	(1.2)		
Depressed Recruitment Case**	Intervention	56	10.5	(0.8)	9.4	(0.8)	8.6	(0.8)	8.1	(0.9)	9.2	(0.9)	.08	0.17
	Control	53	10.7	(0.8)	10.4	(0.8)	10.2	(0.8)	9.9	(0.9)	9.5	(1.0)		
Female	Intervention	67	8.8	(0.7)	8.0	(0.7)	7.4	(0.7)	7.0	(0.7)	7.5	(0.8)	.05	0.42
	Control	61	10.0	(0.7)	10.0	(0.7)	9.9	(0.7)	9.8	(0.8)	9.1	(0.8)		

* P value and overall effect size are presented for the mixed model regression comparing the slope of the intervention condition with control condition over entire study period from baseline to week 32. Reported means adjusted for differences in age, gender, and race.

** Enrolled in study from recruitment cases identified as likely depressed through pharmacy and encounter data

Table 5

Non-experimental, Treatment-As-Usual (TAU) health care services in the 12 months post-randomization, by study condition

	Intervention (n = 83)			Control (n = 77)		
	Mean	SD	N (%) > 0	Mean	SD	N (%) > 0
Mental Health Services						
Mental Health Visits	1.3	(3.1)	23 (28%)	0.9	(2.7)	20 (26%)
SSRI ^a Dispenses	2.2	(2.9)	38 (46%)	1.9	(2.8)	35 (45%)
TCA ^b Dispenses	0.3	(1.6)	3 (4%)	0.1	(0.5)	1 (1%)
Any MH Service			44 (53%)			44 (57%)
Other Health Services						
Outpatient Visits	6.6	(9.4)	71 (86%)	5.7	(4.9)	69 (90%)
Other Rx Dispenses	9.5	(22.1)	67 (81%)	7.7	(8.2)	63 (82%)
						.86

* The "p" column displays the alpha for chi-square analyses comparing the proportion with any of each type of services, for the two conditions

^a selective serotonergic reuptake inhibitor (SSRI) medications

^b tricyclic antidepressant (TCA) medications