Internet Interventions 2 (2015) 48–59

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.invent-journal.com/

Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial

Björn Meyer [a,b,*], Julia Bierbrodt [c], Johanna Schröder [c], Thomas Berger [d], Christopher G. Beevers [e], Mario Weiss [a], Gitta Jacob [a,f], Christina Späth [g], Gerhard Andersson [h,i], Wolfgang Lutz [j], Martin Hautzinger [k], Bernd Löwe [l,m], Matthias Rose [n,o], Fritz Hohagen [g], Franz Caspar [d], Wolfgang Greiner [p], Steffen Moritz [c], Jan Philipp Klein [g]

[a] Research Department, Gaia, Hamburg, Germany
[b] Department of Psychology, City University London, United Kingdom
[c] Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[d] Department of Clinical Psychology and Psychotherapy, University of Bern, Switzerland
[e] Institute for Mental Health Research, University of Texas at Austin, Austin, USA
[f] Department of Clinical Psychology and Psychotherapy, University of Freiburg, Germany
[g] Department of Psychiatry and Psychotherapy, University of Luebeck, Germany
[h] Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden
[i] Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[j] Department of Clinical Psychology and Psychotherapy, University of Trier, Germany
[k] Department of Psychology, Clinical Psychology and Psychotherapy, Eberhard Karls University Tuebingen, Tuebingen, Germany
[l] Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Germany
[m] Schön Klinik Hamburg Eilbek, Germany
[n] Department of Psychosomatic Medicine, Charité University Medical Center, Berlin, Germany
[o] Department of Quantitative Health Sciences, University of Massachusetts Medical School, USA
[p] Department of Health Economics and Health Care Management, Bielefeld University, Bielefeld, Germany

ARTICLE INFO

Article history:
Received 22 July 2014
Received in revised form 13 December 2014
Accepted 15 December 2014
Available online 24 December 2014

Keywords:
Depression
Internet-based treatment
Randomized controlled trial

ABSTRACT

Background: Studies have shown that certain Internet interventions can help alleviate depression. However, many such interventions contain personal support elements, making it difficult to ascertain whether the program or the support drives the effects. Studies are needed to investigate whether Internet interventions contribute to symptom reduction even when they are delivered without personal support, and even among severely depressed individuals who often receive other forms of treatment.
Objective: This randomized controlled trial aimed to examine the effect of an Internet intervention that was deployed without personal support ("Deprexis") among adults with initially severe depression symptoms.
Methods: Adults recruited from a range of sources who had exceeded the threshold for severe depression (PHQ-9 ≥ 15) in a pre-screening assessment and met inclusion criteria were randomized ($N = 163$) to the intervention (3 months program access; $n = 78$) or care-as-usual/waitlist control ($n = 85$). A diagnostic screening interview was administered by telephone at baseline to all participants. Online assessments were administered at baseline, 3 months (post-treatment), and 6 months (follow-up). The main outcome was the Patient Health Questionnaire (PHQ-9) between baseline and post-treatment.
Results: Eighty-two percent of randomized participants were reached for the post-treatment assessment. Results for the intention-to-treat (ITT) sample showed significant intervention effects on depression reduction between baseline and post-treatment (linear mixed model [MM], $F_{1,155.6} = 9.00$, $p < .01$, for the time by condition interaction), with a medium between-group effect size, Cohen's $d = 0.57$ (95% CI: 0.22–0.92). Group differences in depression severity at follow-up were marginally significant in the ITT sample, $t\ (119) = 1.83$, $p = 0.07$, and smaller than at post-treatment (PHQ-9, $d = 0.33$, 95% CI: −0.03–0.69). The number needed to treat (NNT) at post-treatment was 5, with 38% of participants in the intervention group achieving response (at least 50% PHQ-9 symptom change, plus post-treatment score <10), compared to 17% in the control group, $p < 0.01$. Effects on secondary outcomes, including anxiety, health-related quality of life, and somatic symptoms, were not significant, with the exception of significant effects on anxiety reduction in PP analyses. Early ratings of program

* Corresponding author at: Research Department, Gaia, Gertigstr. 12-14, 22303 Hamburg, Germany. Tel.: +49 40 3510 5231; fax: +49 40 3510 5210.
E-mail address: bjoern.meyer@gaia-group.com (B. Meyer).

http://dx.doi.org/10.1016/j.invent.2014.12.003

helpfulness/alliance (after 3 weeks) predicted pre–post depression reduction, controlling for baseline severity and early symptom change.

Conclusions: These results replicate and extend previous findings by showing that Deprexis can facilitate symptomatic improvement over 3 months and, perhaps to a lesser degree, up until 6 months among adults with initially severe depression.

1. Introduction

Over the past decade, at least 25 randomized controlled trials (RCTs) and several systematic reviews and meta-analyses have demonstrated that Internet interventions can reduce depression symptoms among program users and thereby contribute to improving the quality of care for persons with unipolar depression (Andersson and Cuijpers, 2009; Andrews and Titov, 2010; Andrews et al., 2010; Barak et al., 2008; Cuijpers et al., 2011; Johansson and Andersson, 2012; Richards and Richardson, 2012). Consensus seems to be emerging that well-designed depression-focussed Internet interventions are efficacious, and that more relevant research challenges now concern the identification of moderators and mediators, health economic evaluations, the transfer of interventions to routine care settings, and exploring combinations of Internet interventions with pharmaceutical or face-to-face psychotherapeutic treatments (Andersson and Titov, 2014).

This positive appraisal of the research evidence has been expressed enthusiastically in recent years (Andrews and Titov, 2010; Titov, 2011; Kazdin and Blase, 2011), although it is not shared by all. For example, a recent review noted several methodological shortcomings plaguing this field, including the scarcity of non-inferiority comparisons with established treatments, the absence of valid diagnostic procedures and outcome assessments beyond self-report, and the predominant focus on short-term effects (Arnberg et al., 2014). Moreover, most currently studied interventions come from a small number of countries, particularly from research groups in Sweden, Australia, and the Netherlands (Arnberg et al., 2014), although others may be catching up.

Questions also remain with regard to the merit of purely stand-alone interventions, which are typically associated with smaller effect sizes, compared to programs that are accompanied by personal support (e.g., "blended", "guided", or "supported" online interventions) (Johansson and Andersson, 2012; Richards and Richardson, 2012). As Kazdin and Blase remarked (Kazdin and Blase, 2011), though, "the arbiter of the value of a treatment is not necessarily in its effect size … but where that intervention fits into a broader portfolio to help reduce the burden of mental illness" (p. 29). Conceivably, unsupported or minimally supported interventions could also play a useful role, especially when combined with available treatments, although little research has addressed this issue. Interestingly, recent research also suggests that the effectiveness of unsupported interventions can be enhanced by integrating automated forms of support, such as regular reminder e-mails (Titov et al., 2013), so there are several good reasons for continuing to investigate such programs. Our goal in this study, therefore, was to further examine the benefits of an unsupported version of an Internet-based depression intervention in a trial that has several methodological improvements over earlier studies.

In previous work, members of our research group have developed and examined a depression-focussed Internet intervention that has been studied in three published RCTs (Berger et al., 2011; Meyer et al., 2009; Moritz et al., 2012): the Deprexis program. We regard the results from these trials as encouraging, particularly when compared with other "self-guided" (i.e., unsupported) Internet interventions (Cuijpers et al., 2011). The Deprexis program is primarily cognitive–behavioural in content focus but also includes integrative elements, such as acceptance and mindfulness, interpersonal relatedness, positive psychology, and the facilitation of basic psychological needs (Meyer et al., 2009). It can be deployed both in an unsupported (Meyer et al., 2009; Moritz et al., 2012) or a supported version (Berger et al., 2011), which allows clinicians to track program use and symptom trajectories in order to intervene accordingly. There is a need to replicate the program's effects in methodologically stringent trials, as previous studies have suffered from shortcomings such as lack of diagnostic interviews (Meyer et al., 2009; Moritz et al., 2012), small sample sizes (Berger et al., 2011), or high attrition rates (Meyer et al., 2009).

There is also a need to examine in more detail whether Internet interventions such as Deprexis are particularly beneficial for certain patient groups, such as those with milder versus more severe symptoms or with versus without concurrent treatment. Although Internet-based depression interventions are often regarded as an appropriate first step primarily for persons presenting with mild to moderate symptoms (National Institute for Health and Clinical Excellence, 2009), evidence from a patient-level meta-analysis suggests that more severely depressed patients might derive greater benefit (Bower et al., 2013). For severely depressed patients, personally supported rather than stand-alone Internet interventions may seem like a more suitable option, though, not only for ethical and safety reasons but also because they have been found to produce larger effects, on average (Johansson and Andersson, 2012; Richards and Richardson, 2012; Berger et al., 2011). However, we believe that it is also important to examine the effects of online interventions that are offered without support among the severely depressed, for several reasons: Firstly, interventions that do not require personal support resources can be offered reliably and efficiently to large numbers of people in need, even when resources are scarce (Kazdin and Blase, 2011). Secondly, whereas it may be hard to tease apart which aspects of "guided treatments" are responsible for any observed clinical benefits (i.e., the quality or extent of support, the therapeutic relationship, or the software itself), studying unsupported treatments offers the possible methodological advantage of clear causal interpretability. Thirdly, many people with severe depression already receive some form of treatment involving personal contact – perhaps as many as 80% in Europe (McCracken et al., 2006) – so it seems relevant to ask whether offering an Internet intervention as an additional tool might have an incremental or synergistic effect when combined with treatments patients already receive.

The goal of the present trial, then, was to examine the effects of the Deprexis program among adults screened for severe depression symptoms in a moderately sized pragmatic RCT. A large and publically funded RCT (>1000 participants) focussing on the program's effects among people with mild to moderate depression symptoms is currently underway (the EVIDENT trial (Klein et al., 2013)). However, participants with severe symptoms were not admitted for participation in that trial, even though Internet interventions may be most beneficial among more severely depressed persons (Bower et al., 2013). Therefore, the present trial was initiated for those who had expressed interest in participating in the EVIDENT trial but did not meet inclusion criteria due to scoring above the severity threshold on a screening questionnaire, as described below.

As a pragmatic trial (Zwarenstein et al., 2008), the present RCT aimed to test whether offering Deprexis to this group, above and beyond any to standard care they might receive, would reduce depression severity at a greater rate than what would be observed without this program (i.e. in a control group). We also aimed to examine secondary

outcomes, including anxiety, somatic symptoms, and health-related quality of life, and to examine whether early perceptions of program helpfulness would predict pre–post symptom reduction.

2. Method

2.1. Participants

Participants who had expressed interest in participating in the EVIDENT trial (Klein et al., 2013) but were excluded from it because their depression symptoms were beyond the mild to moderate range (that is, >14 on PHQ-9) were automatically referred to a separate study website on which the goals of the study were explained, and they were asked to complete a baseline questionnaire. In the EVIDENT trial, participants were recruited from a broad range of settings, including outpatient and inpatient treatment clinics, Internet depression forums, health insurances, and newspaper advertisements (Klein et al., 2013). Participants were included in the present trial if they surpassed the PHQ-9 score of 14 in the EVIDENT screening, even if they no longer scored above this threshold at the baseline assessment.

Inclusion criteria were: age 18–65, ability to read German, willingness to participate in a telephone diagnostic interview, score of 15 to 27 (i.e., no upper limit) on the PHQ-9 in the initial screening described above, and written informed consent. Exclusion criteria were a lifetime diagnosis of schizophrenia or bipolar disorder (as determined by a telephone diagnostic interview) or current suicidality (as determined in the telephone interview; see below).

2.2. Procedure and study design

This was a parallel-group pragmatic randomized controlled trial. Participants were randomized to two groups: (1) Care-as-usual (CAU) or (2) CAU plus the Internet intervention (i.e. access to the Deprexis program for 3 months). Participants randomized to the treatment group were invited to use the online intervention without guidance or support. That is, there was no contact between participants and study-related personnel, with the exception of the pre-randomization telephone-administered diagnostic screening interview. This design was chosen because it allowed us to evaluate the intervention effect in relatively pure form, without the potential boost conferred by personal support or the therapeutic relationship, while retaining the advantage of collecting valid diagnostic information from a structured interview.

Diagnostic status (presence of current Major Depressive Episode (MDE) or Dysthymic Disorder (DD) according to DSM-IV criteria) was determined with the Mini International Neuropsychiatric Interview (MINI (Lecrubier et al., 1997)), which was administered via telephone by trained interviewers with a university degree in psychology who were in advanced professional training to become licensed psychotherapists. The MINI has been shown to be an efficient tool that can help improve the accuracy of diagnoses, compared to routine physician diagnoses, which are often unreliable (Pinninti et al., 2003; Mitchell et al., 2009). The MINI was also used to confirm the absence of clinical exclusion criteria (likely diagnosis of schizophrenia or bipolar disorder, or suicidality). To ensure patient safety, a personal emergency plan was also discussed with each participant during the telephone interview. That is, emergency contact options were discussed, including the contact details of local psychiatrists, physicians and clinics, as well as emergency telephone hotlines.

Randomization was conducted with an allocation schedule of random numbers that was created by a computerized random number generator. Participants who were deemed eligible after the telephone interview were consecutively placed on this list by one of the researchers (J.B.), who did not conduct telephone interviews and did not have contact with or knowledge of individual study participants. Participants were informed of the outcome of the randomization by a standard e-mail, in which those in the treatment condition received their access voucher that enabled them to log onto the intervention and use it freely (without time-limit and at no cost) for a subsequent 90 day period.

The study was approved by the ethics committee of the German Psychological Association (DGPs, reference number SM 042012_amd_122012), and the trial was entered in an international trials registry (NCT02178631 at ClinicalTrials.gov). Before providing consent, all participants received written information about the aims and procedures used in the study and were informed that they could withdraw at any point without negative consequences. The study was conducted in compliance with the Declaration of Helsinki (World Medical Association, 2001) and took place between 02/2013 and 04/2014.

2.3. Measures

2.3.1. Primary outcome measure

Patient Health Questionnaire - 9 items (PHQ-9; (Gräfe et al., 2004; Kroenke et al., 2001; Löwe et al., 2004; Martin et al., 2006)). The PHQ-9, one of the most thoroughly validated self-report scales for the assessment of depressive symptom severity, serves as the primary outcome measure in the EVIDENT trial (Klein et al., 2013) and, therefore, was also selected as the primary outcome for the present trial. The psychometric properties of the PHQ-9 are excellent, with high internal consistency, test–retest reliability, criterion validity, comparatively favourable sensitivity and specificity, and good sensitivity to change (Kroenke et al., 2010). The PHQ-9, as well as all other measures, was administered via an online survey platform, as there is evidence that this produces valid results and offers the advantage of efficiency (van Gelder et al., 2010; Fann et al., 2009). Clinically significant improvement on the PHQ-9 can be defined as a five-point decline (Löwe et al., 2004; Kroenke et al., 2010) or 50% decline plus post-treatment score below 10 (McMillan et al., 2010).

2.3.2. Secondary outcome measures

Generalized Anxiety Disorder - 7 (*GAD-7*) (Spitzer et al., 2006). The GAD-7 targets symptoms of generalized anxiety disorder, which is highly comorbid with depression and may share a common genetic vulnerability (Gorwood, 2004). The GAD-7 has been found to correlate highly with other anxiety scales, to have a high level internal consistency ($\alpha > .85$), and to have good sensitivity and specificity for the detection of not only GAD but also panic, social anxiety disorder and post-traumatic stress disorders (Kroenke et al., 2010). It is frequently used in Internet intervention trials as a generic measure of anxiety severity (Johansson et al., 2012; Titov et al., 2011; Dear et al., 2011).

Patient Health Questionnaire - 15 items (*PHQ-15*) (Kroenke et al., 2002). The PHQ-15 is an efficient measure of somatic complaints that are commonly encountered in primary care and in the general population, which are also frequently comorbid with both depression and anxiety. As summarized in a recent review, the PHQ-15's psychometric properties are adequate, as evidenced by acceptable internal consistency (α around .80); good convergent validity with measures of functioning, symptom burden and healthcare utilization, and good sensitivity to change (Kroenke et al., 2010). The PHQ-15 can be regarded as a measure of somatic symptom severity as well as potential somatization (Kroenke et al., 2010; Kroenke et al., 2002).

Short Form Health Survey - 12 (*SF-12*) (Ware et al., 1996). The SF-12 is a commonly used measure of general health functioning or health-related quality of life and yields two summary scores: a physical and a mental health index. The SF-12 has been validated internationally and is commonly used in epidemiological surveys, although it has also been used among patients with psychiatric illness, including depression (Sugar et al., 1998; Gandek et al., 1998; Salyers et al., 2000). Psychometric properties, including test–retest reliability, internal consistency and convergent validity are good to excellent, and comparable to its longer version, the SF-36 (Gandek et al., 1998; Jenkinson et al., 1997).

2.3.3. Treatment satisfaction and alliance/helpfulness ratings

An adapted version of a patient satisfaction questionnaire that is widely used in Germany, the ZUF-8 (Schmidt et al., 1989), was used in this study. This brief and reliable instrument was originally developed as a translation of the Client Satisfaction Questionnaire (CSQ-8 (Attkisson and Greenfield, 1999)) and was originally intended to assess satisfaction with inpatient treatment. The items were reworded slightly to focus on satisfaction with the Internet intervention examined in this trial.

We also administered an adapted version of the 11-item Helping Alliance Questionnaire (HAQ-11 (Alexander and Luborsky, 1986; Bassler et al., 1995)) three weeks after randomization to participants allocated to the treatment group. The HAQ-11 yields either a total score (used here) or two subscale scores, indicating the extent to which patients feel that the treatment is helpful, seems to view problems in the same way and seems to share their goals. A six-point response scale was used, with values of −1 to −3 indicating negative alliance/helpfulness perceptions, whereas values of +1 to +3 indicate positive alliance/helpfulness perceptions. The mean of 11 items was computed to indicate patients' impressions of alliance/helpfulness early. The HAQ is one of the most commonly used alliance questionnaires and has been found to have good psychometric properties and predict various aspects of psychotherapeutic process and outcome, although it is not always clear whether the alliance, as measured by scales such as the HAQ, is a reflection of third variables, such as early symptom improvement, rather than an independent predictor of process and outcome (Elvins and Green, 2008).

2.4. Treatment

The Deprexis program is an integrative Internet intervention for adults with symptoms of unipolar depression (Meyer et al., 2009), developed and operated by a research-focussed public health company in Hamburg, Germany (Gaia). The program uses a software technology (broca®) that was designed to allow for the tailoring of content to dynamically changing user requirements. To accomplish this, the core of the program revolves around a series of "simulated dialogues" in which patients are continuously asked to select one of several response options, after which the program then presents content to match the chosen response. Content covered in these "dialogues" ranges from traditional cognitive–behavioural topics, such as cognitive restructuring, behavioural activation, problem-solving, and acceptance/mindfulness to optional topics such as positive psychology and emotion-focussed techniques (for a more detailed description, see Meyer et al. (2009); Meyer et al. (2009)).

The Deprexis program in its current version also includes daily text (SMS) messages that are deployed optionally and convey content that has been covered in the program. If users consent, they receive a different daily message for 3 months, which is intended to briefly remind them of key ideas conveyed by the program and encourage them to apply these techniques in their current situational context (Examples: (a) "Which thoughts are going through your mind right now? Are they helpful? If so, let them be! If not, let them go calmly"; (b) "Helpful response to mistakes: 'I've made a mistake, which is normal and human!' Support yourself compassionately…"). Additionally, the program integrates features such as symptom-tracking (e.g., questionnaires such as the PHQ-9 are offered in two-week intervals and graphical as well as text feedback is provided), worksheets and summaries in printable format, audio recordings, and photos as well as illustrations. The program has been CE-certified as a medical product in Europe and is currently available in German, English and Swedish language versions. Previous trials have demonstrated the program's effectiveness (Berger et al., 2011; Meyer et al., 2009; Moritz et al., 2012), and above-average effects have been found for the program in meta-analyses (Cuijpers et al., 2011; Richards and Richardson, 2012). Participants in the Deprexis group were also permitted to use whichever components of care-as-usual (CAU) were available to them, including pharmacotherapy and/or psychotherapy (just as in the control condition).

2.5. Control condition

Participants in the control condition were not influenced or advised to change their existing treatment patterns (that is, they received CAU). They were informed that they could receive access vouchers to the Deprexis program after the study period of 6 months, if they wished. Therefore, with respect to gaining access to Deprexis, this is a waitlist control condition, albeit with the caveat that participants were permitted to use any other treatments available to them. This CAU plus waitlist comparison condition was chosen in line with the logic governing pragmatic randomized control trials: to maximize external validity and test whether the intervention improves outcome compared to the heterogeneous care realities characterizing most healthcare systems (Zwarenstein et al., 2008). The extent to which control (and intervention) participants used concurrent treatments, such as antidepressant medication or psychotherapy, was assessed and examined.

2.6. Statistical analysis

Following current standards (Meyer et al., 2009; Mackinnon et al., 2008; Andersson et al., 2012a; Ivarsson et al., 2014), we used a linear mixed effects models (MM) approach with full information maximum likelihood estimation. This approach has been recommended because it uses all available data and can handle missing data appropriately (Gueorguieva and Krystal, 2004). The approach is based on the assumption that data are missing at random, and it is not assumed that missing data remain stable, as in the no-longer-recommended last-observation-carried-forward (LOCF) approach (Blankers et al., 2010). We used a first-order autoregressive covariance structure to model both post-treatment and follow-up effects.

Consistent with CONSORT recommendations (Eysenbach and CONSORT-EHEALTH Group, 2011), we report MM analyses across multiple groups: (1) The intention-to-treat (ITT) group, which includes data from all randomized participants, regardless of whether they used the intervention or activated their access vouchers to enter the program, (2) the Per-Protocol (PP) group of participants in the treatment condition who had activated their vouchers, thus indicating an interest in it (referred to as the PP-AV sample; AV: "activated voucher"), (3) the PP sample of treatment participants who were able to engage with the intervention for a reasonable amount, defined here as having started at least four sessions and spent a total of at least 60 min actively engaged within the program (referred to as the 4S group; 4S: "at least four sessions"). The definition of four sessions for reasonable intervention use (adequate dosage) is consistent with other research groups' definitions (Ivarsson et al., 2014). Also, completion of four sessions exposes users to the content we deem minimally sufficient (e.g. psychoeducation, behavioural activation, cognitive restructuring). Participants in the PP 4S group are, therefore, considered to be adherent to the program.

We do not report analyses for participants who completed "all modules" because this number varies per user, given the tailored nature of content delivery in Deprexis (Meyer et al., 2009). For example, whereas core content covering behavioural activation and cognitive restructuring is always presented early on, content on emotion-focussed interventions is offered only for participants who indicate an express wish for such topics. We would not expect that every user must view all content in order to benefit.

Following evidence-based recommendations, we defined response on our primary outcome measure, the PHQ-9, as a post-treatment (3 month) score of below 10 combined with improvement of at least 50%, among participants who had scored above 10 at baseline (McMillan et al., 2010). This definition has been shown to be closely aligned with alternative approaches, such as using the reliable change criterion developed by

Jacobson and Truax (Jacobson and Truax, 1991). At baseline, 159 participants (98%) scored above 10 on the primary outcome measure (PHQ-9) and could, therefore, be included in the responder analyses.

Effect sizes are presented as Cohen's *d* throughout; that is, differences between means at post- and follow-up tests, divided by the pooled standard deviation of the respective subsample. All effect sizes were computed from the observed means of the respective groups (e.g., ITT, PP-AV, PP-4S), consistent with current practices used by other groups (Andersson et al., 2012a).

An a-priori power analysis revealed that the study was adequately powered (i.e. >0.80) to detect a post-treatment group differences corresponding to $d = 0.45$, which seemed realistic, given that previous research has found similar effects (Johansson and Andersson, 2012; Berger et al., 2011), particularly in studies involving pre-treatment contact, such as the initial diagnostic interview employed here.

3. Results

3.1. Enrolment and baseline characteristics

As shown in Table 1, about 3 out of 4 participants were women, and the average age was just above 40. The typical (modal) participant had completed secondary education, was not working full-time, and had found the study via news reports, an Internet search, or posts in depression Internet forums. Participants were also recruited via clinical settings or informed by their health insurance companies. About half of the participants were currently receiving antidepressant medication, and about half reported being in current psychiatric or psychotherapeutic treatment. When asked about receiving any current treatment, 120 (74%) of the 163 randomized participants reported either being in some form of psychiatric or psychological treatment, taking antidepressants, or receiving psychotherapy, whereas 43 (26%) reported that they currently did not receive any such treatment.

The vast majority of enrolled subjects (84%) reported having received a clinical diagnosis of a depressive disorder from a healthcare provider, and most (88.3%) stated that their depression might be (51.5%) or was definitely (36.8%) a chronic problem. On average, participants had suffered from depression symptoms for over 15 years, given that mean symptom onset was reported at age 25, with average current age above 40. Nearly half (48%) reported early onset of depressive symptoms (before age 21). Almost all (87%) reported using the Internet almost daily, which is in line with population usage patterns: 79% of Germans above age 10 used the Internet daily or almost daily in the first quarter of 2013 (German Federal Statistical Office (Statistisches Bundesamt), 2014).

Table 1
Demographic and clinical variables.

		Treatment (n = 78)		Control group (n = 85)		Total (n = 163)	
Variable	Sub-variable	n	%	n	%	n	%
Gender	Male	20	25.6	21	24.7	41	25.2
	Female	58	74.4	64	75.3	122	74.8
Age	Mean age (SD)	44 (11.02)		40 (11.48)		42 (11.39)	
	Range	21–62		18–62		18–62	
Marital status	Married	28	35.9	17	20	45	27.6
	Separated	5	6.4	4	4.7	9	5.5
	Single	19	24.4	29	34.1	48	29.4
	In relationship	14	17.9	26	30.6	40	24.5
	Divorced	10	12.8	9	10.6	19	11.7
	Widowed	2	2.6	0	0	2	1.2
Education*	Lower secondary	4	5.1	3	3.5	7	4.3
	Middle secondary	23	29.5	14	16.5	37	22.7
	Higher secondary	14	17.9	12	14.1	26	16
	Highest secondary	37	47.4	53	62.4	90	55.2
	Other	0	0	3	3	3.5	1.8
Employment status	Full time	38	48.7	31	36.5	69	42.3
	Regular part-time	9	11.5	16	18.8	25	15.3
	Not working	31	39.7	38	44.7	69	42.3
Recruitment source	Via clinic/doctor	11	14.1	13	15.3	24	14.7
	Internet forum	18	23.1	23	27.1	41	25.2
	Health insurance	13	16.7	7	8.2	20	12.3
	Other (e.g., Internet search, newspaper)	36	46.2	42	49.4	78	47.9
Currently on antidepressants	Yes	39	50	45	52.9	84	51.5
	No	39	50	40	47.1	79	48.5
Currently in psychiatric/psychotherapeutic treatment	Yes	43	55.1	45	52.9	88	54
	No	35	44.9	40	47.1	75	46
Waiting for psychotherapy	Yes	22	28.2	24	28.2	46	28.2
	No	56	71.8	61	71.8	117	71.8
Age of depression onset	Mean Age (SD)	27 (13.88)		24 (12.47)		25 (13.22)	
Chronicity of depression	Definitely not chronic	11	14.1	8	9.4	19	11.7
	Unsure	40	51.3	44	51.8	84	51.5
	Definitely chronic	27	34.6	33	38.8	60	36.8
Clinical diagnosis of a depressive disorder	Yes	66	84.6	71	83.5	137	84
	No	12	15.4	14	16.5	26	16
Diagnosis (MINI)	Only MDE	29	37.2	32	37.6	61	37.2
	Only DYS	10	12.8	7	8.2	17	10.4
	Both MDE and DYS	29	37.2	35	41.2	64	39.3
	Neither	10	12.8	11	12.9	21	12.9
Frequency of Internet use	≤3 times per week	12	15.4	9	10.6	21	12.9
	(Almost) every day	66	84.6	76	89.4	142	87.1

* Secondary education according to the German classification: "Hauptschule" ("lower", 9 years, until age 15/16), "Realschule" ("middle", 10 years, until age 16/17), "Fachhochschulreife" (12 years, until age 17/18), "Abitur" (12 or 13 years, until age 17–19).

Mean depression severity at baseline, as measured by the PHQ-9, was in the severe range, at 16.92 ($SD = 3.66$); 75% of participants scored above the cut-off of 15 for severe depression; 23% scored in the moderate range (10–14), and only 2% scored below 10.

There were no statistically significant differences in any baseline demographic or clinical characteristics between participants in the treatment versus control groups, with the exception that those in the treatment group were slightly older, $t(161) = 2.18$, $p = .03$. Therefore, age was entered as a covariate in all inferential outcome analyses below (although results were not substantially affected by this inclusion; that is, the significance of all findings reported below remained the same regardless of whether age was included as a covariate or not).

Information collected with the MINI interview revealed that 87% of randomized participants fulfilled diagnostic criteria for a current depressive syndrome or disorder. Specifically, 77% met criteria for current Major Depressive Episode (MDE), either alone (37%) or in the form of double-depression (that is, with additional diagnosis of dysthymic disorder, 39%). An additional 10% met diagnostic criteria for dysthymic disorder but not current MDE (see Table 1). The diagnoses were distributed evenly across the treatment versus control groups, as confirmed by χ^2 tests, $ps > .40$.

3.2. Patient flow and treatment adherence

The numbers and percentages of retained participants at the 3- and 6-month time-points are shown in the CONSORT flow chart (Fig. 1). The attrition rate at post-treatment was 18% and can thus be considered low, particularly for an unsupported Internet intervention. In terms of treatment adherence, 80% ($n = 56/70$) of the participants in the treatment condition who had activated their voucher completed at least four sessions and used the program for a minimum of one hour and, thus, can be regarded as treatment-adherent.

The average (mean) time these adherent users engaged online with the program was estimated to be 457 min ($SD = 240$) or just over seven and a half hours. The mean time that all users (including the non-compliant group) who had activated their vouchers spent online within the program was estimated to be 390 min ($SD = 254$) or six and a half hours. In the computation of these usage times, periods of inactivity of 5 min or longer were subtracted. Usage time did not correlate with the amount of pre to post or pre to follow-up depression change on the PHQ-9, $p > .20$.

3.3. Treatment effects on continuous outcomes at post-assessment

3.3.1. Intention-to-treat analyses

Table 2 presents the observed means of both the ITT and PP groups; effect sizes and 95% confidence intervals are presented in Table 3. Between-group effects in the ITT sample on depression severity at post-treatment were in the medium range, conventionally defined as $d = 0.50$ (Cohen, 1992). Specifically, a between-groups effect of $d = 0.57$ was observed at post-treatment and a small effect of $d = .33$ at follow-up.

We observed large pre–post reductions in depression on the primary outcome variable (PHQ-9) in the treatment group, with a within-group pre–post effect size of $d = 1.32$; for the baseline to follow-up period: $d = 1.13$. Depression reduction was also substantial in the control group, though, with a moderate to large pre–post effect of $d = 0.71$ on the primary outcome measure (PHQ-9), and a similar moderate to large baseline to follow-up within-group effect of $d = 0.73$. MM analyses revealed a significant time by treatment condition interaction for the PHQ-9 at post-treatment in the ITT sample ($F_{1,155.6} = 9.00$, $p < .01$).

Effect sizes for secondary outcomes were smaller than those for depression measures, with the exception that the effect on anxiety (GAD-7) approached that observed for depression. However, with GAD-7 as the dependent variable, the interaction between time and treatment

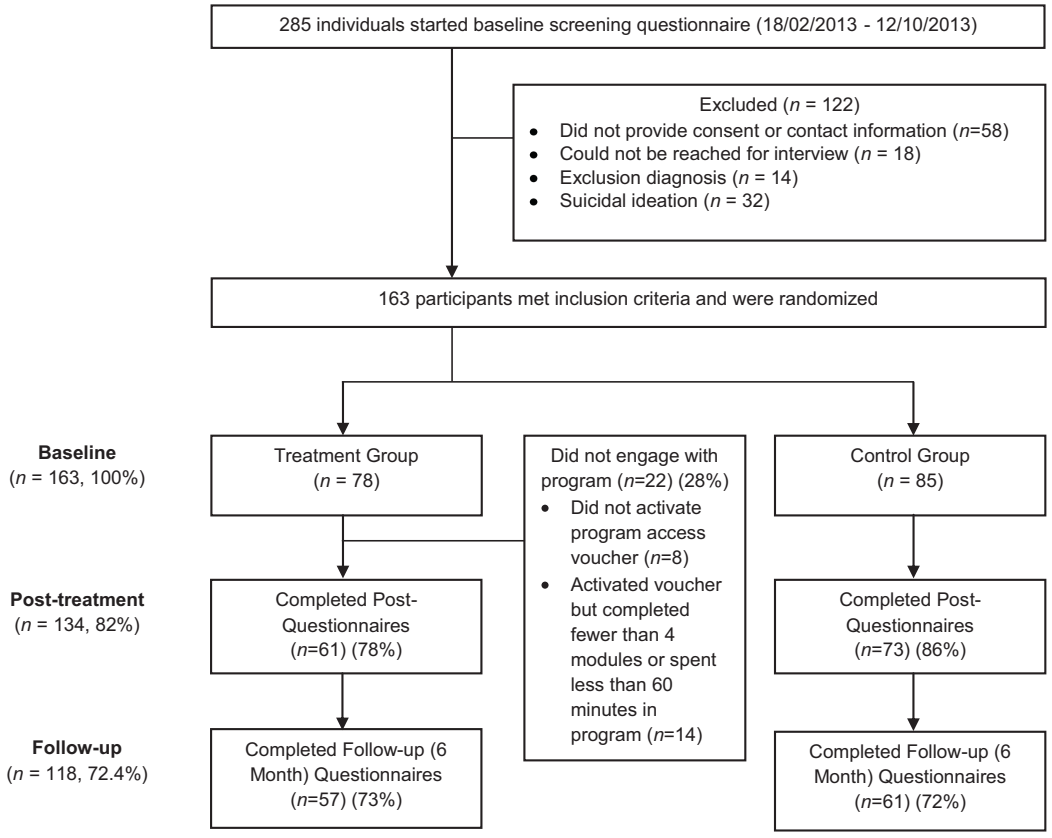

Fig. 1. CONSORT-R participant flow chart.

Table 2
Results of outcome measures: Means and standard deviations.

		ITT							PP					
		Pre n (treat) = 78 n (control) = 85		Post n (treat) = 60–61 n (control) = 72–73		Follow-up n (treat) = 54–57 n (control) = 62–64			Pre n (AV) = 70 n (4S) = 56		Post n (AV) = 58–59 n (4S) = 51		Follow-up n (AV) = 52–55 n (4S) = 45–48	
Measure	ITT subgroup	Mean	SD	Mean	SD	Mean	SD	PP treatment subgroup	Mean	SD	Mean	SD	Mean	SD
Primary outcome														
PHQ-9	Treatment	16.62	3.44	10.08	6.37	11.28	6.04	AV	16.56	3.40	9.78	6.15	11.05	6.02
	Control	17.20	3.86	13.64	6.14	13.39	6.59	4S	16.54	3.33	9.53	6.35	11.15	6.03
Secondary outcomes														
GAD-7	Treatment	13.03	3.77	8.69	5.06	9.19	4.78	AV	13.00	3.74	8.56	5.10	9.07	4.81
	Control	13.44	3.90	10.73	5.19	10.61	5.76	4S	13.04	3.76	8.29	5.14	9.06	4.61
PHQ-15	Treatment	13.06	4.65	10.89	5.32	11.67	5.21	AV	12.90	4.70	10.66	5.20	11.45	5.16
	Control	12.99	4.51	12.00	5.77	11.98	5.88	4S	12.73	3.92	10.39	5.06	11.46	4.78
SF-12 — physical	Treatment	45.39	10.72	45.82	9.45	41.60	9.14	AV	45.63	10.74	45.83	9.60	41.93	9.11
	Control	43.79	10.45	43.74	11.58	41.88	11.36	4S	46.33	9.80	46.44	9.53	41.80	9.17
SF-12 — mental	Treatment	25.93	6.95	33.70	11.24	34.58	11.74	AV	25.88	6.62	34.06	11.24	34.90	11.81
	Control	24.65	7.31	30.62	11.61	32.89	11.82	4S	25.98	6.22	34.12	11.38	34.88	11.43

ITT: Intention-to-treat sample (data from all subjects); PP: Per-protocol sample (data in treatment group only from subjects who logged on and started at least 4 modules); Pre: Pre-treatment, Post: Post-treatment (3 month); Follow-up: 6 month; PHQ-9: Patient Health Questionnaire 9-Item; GAD-7: Generalized Anxiety Disorder 7-Item; SF-12: Short Form Health Survey – 12 Items; AV = activated voucher to access program; 4S: completed at least 4 modules and spent ≥60 min in program.

condition was only marginally significant in the ITT sample, $F_{1,150.9} = 3.46$, $p = .07$. No other significant interaction effects were observed in the ITT sample.

3.3.2. Per-protocol analyses

Observed means for the PP groups (PP-AV and PP-4S) are presented in Table 2; effect sizes and 95% confidence intervals are shown in Table 3. In general, the effects in the AV and 4S groups were similar in magnitude, with slightly larger effects among those in the PP-4S group. Large pre–post depression reductions were observed in the PP samples, with effects of $d = 1.40$ for the PHQ-9. Between-group effects were in the medium-large range for the PP groups.

MM analyses revealed significant treatment effects for depression in both PP groups. That is, in the prediction of PHQ-9 scores at post-treatment, the interaction between time and treatment group was significant for the PP-AV sample, $F_{1,148.8} = 11.07$, $p < .01$, as well as for the PP-4S sample, $F_{1,136.6} = 11.14$, $p < .01$.

Tables 2 and 3 show that effect sizes for the secondary outcome measures were somewhat smaller, compared to the primary outcomes. However, the interaction between time and treatment group in the prediction of GAD-7 reached significance, both for the PP-AV group, $F_{1,145.6} = 4.18$, $p < .05$, and for the PP-4S group, $F_{1,134.6} = 4.86$, $p < .03$. No other significant interaction effects between time and treatment condition were observed in the PP samples, although in the prediction of PHQ-15, this interaction was marginally significant ($p = .08$, PP-4S group).

3.4. Treatment response at post-treatment and follow-up

Of 163 randomized participants, 159 (98%) scored above 10 on the primary outcome measure (PHQ-9) and were thus included in the responder analyses. In the ITT sample, 38.2% ($n = 29/76$) of participants in the treatment group were classified as responders at post-treatment, compared to 16.9% ($n = 14/83$) of those in the control group. This difference was statistically significant, $\chi^2(1) = 9.11$, $p < .01$ and corresponded to a number needed to treat (NNT) of 5 (CI-95% = 2.9–13.0). In the PP-AV sample, at post-treatment, the proportion of participants classified as responders was 42.6% for the PP-AV group ($n = 29/68$; NNT = 4, CI-95% = 2.5–8.7) and 49.1% in the PP-4S group ($n = 27/55$; NNT = 4, CI-95% = 2.1–6.0). These proportions also differed significantly from the proportion of 16.9% ($n = 14/83$) observed in the control group (for the PP-AV group: $\chi^2(1) = 12.20$, $p < 0.001$; for the PP-4S group: $\chi^2(1) = 16.45$, $p < .001$).

When the more lenient criterion for clinically significant change recommended by Kroenke et al. (Kroenke et al., 2010) was used (5-point pre–post PHQ-9 decline), response rates were higher. That is, 37% of control participants responded at post-treatment, compared with 53% (ITT), 59% (PP-AV), or 63% (PP-4S) in the treatment groups, respectively. The differences between the treatment groups versus control group were also statistically significant when this definition of post-treatment response was used, $ps < .05$.

At follow-up, there were few differences in response rates between the treatment and control groups, as would be expected, given the trend toward symptom severity convergence (Table 2). Specifically, using the response definition of at least 50% pre to follow-up PHQ-9 score reduction plus a score of below 10 at follow-up, response rates did not differ significantly between groups. Collapsing across groups, 23% could be classified as responders according to this definition in the ITT groups, and there were no significant group differences when the per-protocol groups were examined. However, when the more lenient response definition of at least 5-point PHQ-9 reduction was used, one significant difference was observed. That is, among the treatment-adherent participants (PP-4S group), 50% were classified as responders at follow-up (28/56), whereas this was true for only 32% of control participants (27/85). This difference was statistically significant, $\chi^2(1) = 4.72$, $p < 0.05$. This corresponds to an NNT of 6 (CI-95% = 2.9–54.9).

3.5. Stability of treatment effects up to 6 months

Tables 2 and 3 show that between-group differences at 6 months tended to be slightly smaller than those observed at 3 months, and this was due to a tendency toward convergence between the groups. For example, whereas participants in the control group reported a slight reduction on the PHQ-9 between month 3 and 6, on average, PHQ-9 scores in the treatment group rose slightly. To examine the course of symptom changes from post-treatment to follow-up, we conducted MM analyses in which only the time-span between post-treatment to follow-up was considered. These analyses were again conducted in the ITT group and both PP groups.

With the PHQ-9 as dependent variable, we observed significant main effects for treatment condition, $ps < .03$, but no significant effects for time in these analyses. This showed that the significantly reduced

Table 3
Effect sizes (Cohen's *d*, 95% CI).

Measure	ITT				PP				
		WG (pre–post)		BG (treatment vs. control)			WG (pre–post)		BG (treatment vs. control[1])	
	ITT subgroup	0–3 months	0–6 months	3 months	6 months	PP treatment subgroup	0–3 months	0–6 months	3 months	6 months
Primary outcome										
PHQ-9	Treatment	1.32 (.95–1.69)	1.13 (.77–1.50)	.57 (.22–.92)	.33 (−.03–.69)	AV	1.40 (1.01–1.78)	1.16 (.78–1.55)	.63 (.28–.98)	.37 (.01–.73)
	Control	.71 (.38–1.03)	.73 (.40–1.07)			4S	1.40 (.98–1.83)	1.13 (.71–1.55)	.66 (.29–1.03)	.35 (−.02–.73)
Secondary outcome										
GAD-7	Treatment	.99 (.64–1.35)	.91 (.55–1.27)	.40 (.05–.74)	.27 (−.09–.63)	AV	1.01 (.64–1.37)	.93 (.55–1.30)	.42 (.07–.77)	.29 (−.07–.65)
	Control	.60 (.28–.92)	.59 (.26–.92)			4S	1.06 (.66–1.47)	.95 (.55–1.36)	.47 (.11–.83)	.29 (−.08–.67)
PHQ-15	Treatment	.44 (.10–.78)	.28 (−.06–.63)	.20 (−.14–.54)	.06 (−.30–.41)	AV	.45 (.10–.80)	.30 (−.06–.65)	.24 (−.10–.59)	.10 (−.27–.46)
	Control	.19 (−.12–.51)	.20 (−.13–.52)			4S	.52 (.13–.91)	.29 (−.09–.68)	.29 (−.07–.65)	.10 (−.28–.47)
SF-12 — physical	Treatment	.04 (−.30–.38)	.38 (.03–.72)	.20 (−.15–.54)	.03 (−.34–.39)	AV	.02 (−.33–.37)	.37 (.01–.73)	.19 (−.15–.54)	.00 (−.36–.37)
	Control	.00 (−.31–.32)	.18 (−.15–.50)			4S	.01 (−.37–.39)	.48 (.08–.87)	.25 (−.11–.61)	.01 (−.38–.39)
SF-12 — mental	Treatment	.86 (.51–1.21)	.94 (.57–1.30)	.27 (−.08–.61)	.14 (−.22–.51)	AV	.91 (.54–1.27)	.98 (.60–1.36)	.30 (−.05–.65)	.17 (−.20–.54)
	Control	.63 (.31–.95)	.87 (.53–1.21)			4S	.90 (.50–1.30)	1.00 (.58–1.41)	.30 (−.06–.66)	.17 (−.21–.56)

ITT: Intention-to-treat sample (data from all subjects); PP: Per-protocol sample (data in treatment group only from subjects who logged on and started at least 4 modules); Pre: Pre-treatment, Post: Post-treatment (3 month); Follow-up: 6 month; PHQ-9: Patient Health Questionnaire 9-Item; GAD-7: Generalized Anxiety Disorder 7-Item; SF-12: Short Form Health Survey – 12 Items; AV = activated voucher to access program; 4S: completed at least 4 modules and spent ≥60 min in program.
[1] Values for the control group do not differ between ITT and PP analyses as control participants, per definition, did not use the program.

depression scores in the treatment group remained lower than those in the control group when the period from 3 to 6 months was considered, despite the relative convergence of means and reduced between-groups effect size at follow-up (see Tables 2 and 3). One significant interaction between time and treatment group was observed. That is, in the PP-4S sample, participants in the treatment group reported a slight increase in PHQ-9 scores between post-treatment and follow-up, whereas there was a minimal reduction over this period in the control group, $F_{1,114.47} = 4.95$, $p < .03$. Even in this subgroup, though, PHQ-9 scores at six months remained lower in the treatment than the control group (see Table 2).

Mean comparisons at post-treatment and follow-up were also conducted and showed that PHQ-9 scores differed significantly between the intervention and control groups at post-treatment, both in the ITT and PP samples, $ps < .01$. At follow-up, however, the difference in PHQ-9 scores was only marginally significant in the ITT sample, $t(119) = 1.83$, $p = 0.07$, and in the PP-4S sample, $t(110) = 1.85$, $p = 0.07$. In the PP-AV sample, PHQ-9 scores at follow-up were significantly lower in the intervention group than the control group, $t(117) = 2.01$, $p < 0.05$.

3.6. Treatment moderators

In line with previous trials, we explored whether diagnostic status[1], initial depression severity, onset age (before vs. after age 21), current age, education, gender, concurrent psychotherapy, or taking antidepressant medication during the treatment period moderated pre–post intervention effects. For this purpose, MM analyses were performed, focussing on the three-way interaction between time (pre to post-treatment), condition (intervention vs. control), and the respective moderator variable. Separate analyses were conducted for the ITT and both PP groups.

Diagnostic status, initial depression severity, onset age, current age, gender, education, and receiving psychotherapy during the pre–post period did not appear to moderate the effect of treatment condition on symptom change over 3 months, $ps > .05$. However, reported antidepressant use during the pre–post interval appeared to moderate the treatment effect. That is, there was a significant interaction among time, condition, and antidepressant use, regardless of whether the ITT, PP-AV, or PP-4S group was considered; for the ITT group: $F_{1,134} = 11.51$, $p < .01$; for the PP-AV group: $F_{1,132} = 10.88$, $p < .01$; for the PP-4S group: $F_{1,124} = 8.96$, $p < .01$.

To follow up on the three-way interaction, we examined two-way interactions among those with versus without antidepressant use in the ITT group. Among participants who did not use antidepressants in the pre–post interval, time and condition did not interact significantly, $p > .5$. Among those who reported concurrent antidepressant use, though, the interaction between time and condition was significant, $F_{1,69} = 18.10$, $p < .001$. Among treatment group participants on antidepressants, PHQ-9 scores reduced from 16.89 ($SE = 0.93$) to 8.79 ($SE = 0.93$), on average (estimated marginal means, ITT group), which is a reduction by 48.0% from pre-treatment scores. Among control participants on antidepressants, by contrast, PHQ-9 scores reduced from 16.88 ($SE = 0.77$) to 14.54 ($SE = 0.77$), on average, corresponding to a 13.9% reduction from pre to post-treatment. Among participants on antidepressants, PHQ-9 scores did not differ at pre-treatment between the treatment vs. control groups, $F_{1,69} = 0.001$, $p > .9$, but the groups differed at post-treatment,

[1] For these analyses, we used a maximally conservative definition of diagnostic status. That is, the diagnosis of major depressive disorder (MDD) was only awarded if (1) MINI diagnostic criteria for major depressive episode were met, (2) the diagnosis was confirmed by the diagnostic algorithm of the PHQ-9 at baseline, (3) the threshold of sum-score at least 15 on the PHQ-9 was met, and (4) the participant reported having been diagnosed with a depressive disorder by a healthcare provider, such as a physician. Using this approach, 52% (n = 85/163) were diagnosed with MDD. These diagnostic cases were distributed evenly in the treatment versus control group (52% in the control group with MDD, vs. 53% in the treatment group, $\chi^2(1) = 0.01$, $p > .90$).

$F_{1,69} = 16.32$, $p < .001$, and the between-group difference was maintained at follow-up, $F_{1,61} = 4.57$, $p < .04$.

3.7. Treatment satisfaction and alliance/helpfulness ratings

Sixty participants in the treatment group completed the adapted ZUF-8 at post-treatment (missing data for one participant on this scale). The sum score of the ZUF-8 was 24.17 ($SD = 4.82$), and the mean score per item was 3.02 ($SD = 0.60$), reflecting a generally positive evaluation, on average (4 = most positive response; 1 = most negative). The ZUF-8 sum score did not differ significantly from that reported for a normative sample of 664 inpatients who had been treated in psychosomatic hospitals in Germany ($M = 25.12$; $SD = 4.63$) (Kriz et al., 2008). Subjective program evaluation correlated with treatment response, as measured by percentage symptom reduction (PHQ-9) between baseline and 3 months, $r = 0.37$, $p < .01$. Patients who responded to the treatment (>50% pre–post reduction plus post-treatment score <10 on PHQ-9) were more satisfied ($M = 26.14$, $SD = 3.93$) than non-responders ($M = 22.20$, $SD = 4.94$), $t(57) = 3.81$, $p < .01$. On average, nearly 80% endorsed a favourable response (77.5%), indicating general satisfaction. For example, 83.3% stated they would probably or definitely recommend Deprexis to a friend, and 88.3% rated the quality of the program as good or excellent.

Sixty-two participants in the treatment group also completed the HAQ-11 an average of three weeks ($M = 21.7$ days, $SD = 5.9$) after activating their program access voucher. Overall, 71% had a positive impression of the program's helpfulness after three weeks, as indicated by scores above the response-scale midpoint of zero. A correlation of $r = .42$ ($p < .01$) with pre to post-treatment PHQ-9 change and $r = .46$ ($p < .01$) with pre to post-treatment PHQ-9 *percentage change* showed that early perceived helpfulness/alliance predicted pre–post symptom reduction. These correlations remained significant when controlling for baseline symptom severity (PHQ-9 pre-treatment scores), partial $r = .49$ with sum-score change, partial $r = .47$ with percentage change, $ps < .01$. In fact, the correlation between the HAQ-11 and pre–post symptom reduction also remained significant when controlling for early symptom change. That is, $n = 49$ participants in the treatment group completed the PHQ-9 twice within the first 35 days *within the Deprexis program* (on average, first PHQ-9 completed on day of initial log-in; second PHQ-9 after 24.8 days, $SD = 7.4$), as doing so is a standard feature of the program. Controlling for early PHQ-9 change, as measured by the difference between these first two within-program PHQ-9 scores, the HAQ-11 was still significantly correlated with pre–post PHQ-9 sum score change, partial $r = .34$, $p < .02$, and with pre–post PHQ-9 percentage change, partial $r = .40$, $p < .01$.

To explore whether these early alliance ratings could predict clinical response, a median split was used to classify respondents into those with clearly positive (HAQ-11 mean score > 1, $n = 30$, 48%) versus less positive (HAQ-11 < 1, $n = 32$, 52%) alliance/helpfulness ratings. Among those with positive early alliance ratings, 69% ($n = 20/29$) achieved clinically significant pre–post response (>50% PHQ-9 pre–post reduction plus post-treatment score below 10), whereas only 25% ($n = 8/32$) of those with less positive early alliance ratings achieved such response, $\chi^2(1) = 11.84$, $p < .001$.

A correlation of $r = .60$, $p < .001$ also showed that alliance/helpfulness ratings after 3 weeks were strongly associated with overall treatment satisfaction (ZUF-8) after treatment termination.

3.8. Subsidiary analyses: PHQ-9 changes among diagnosed participants with severe depression

The analyses reported above focussed on all participants included in this trial, which included those who exceeded the severe depression threshold (PHQ-9 > 14) at a pre-screening but not necessarily at the baseline assessment. All of the included participants had exceeded the severity criterion initially but 25% of them scored below 15 at baseline, and 13% did not meet criteria for MDD or DD on the MINI. Therefore, the question arises whether findings hold up among participants who exceeded the severity criterion at baseline as well and who were eligible for a diagnosis of MDE or DD. Thus, subsidiary analyses were conducted with this more stringently defined subsample of participants with diagnosed depression and severe symptoms at baseline. These analyses focussed on the trajectory of PHQ-9 symptoms between baseline, post-treatment, and follow-up. We conducted ITT analyses among this subgroup with the PHQ-9 serving as dependent variable.

Using the inclusion criteria of confirmed MDE or DD plus baseline severity of PHQ-9 > 14, $n = 109$ (67%) were included. Within this subsample, 81% were reached at post-treatment and 72% at follow-up. Focussing on the baseline to post-treatment period, MM analyses yielded a significant time by treatment condition interaction, $F_{1,106.3} = 10.23$, $p < .01$. Mean PHQ-9 scores changed by 7.68 points over this period within the treatment group, from 18.24 ($SD = 2.72$) to 10.56 ($SD = 6.26$), whereas in the control group, a mean PHQ-9 reduction of 3.78 points was observed, from 18.76 ($SD = 3.11$) to 14.98 ($SD = 6.40$). The within-groups effect size was very large for the treatment group ($d = 1.66$, 95% CI: 1.19–2.12), and about half as large (although still large in absolute terms) for the control group ($d = 0.78$, 95% CI: 0.36–1.20). The between-groups effect size at post-treatment was moderate to large ($d = 0.70$, 95% CI: 0.27–1.13).

Mean PHQ-9 scores increased slightly in the treatment group between post-treatment and follow-up ($M = 12.69$, SD = 6.34), whereas they remained slightly higher but stable in the control group ($M = 15.00$, $SD = 6.53$). The pre-to-follow-up within-group effect for the treatment group was large ($d = 1.21$, 95% CI: 0.75–1.68) and somewhat smaller but still substantial for the control group ($d = 0.78$, 95% CI: 0.36–1.19). The between-groups effect size at follow-up was small to moderate ($d = 0.36$, 95% CI: −0.09–0.81). MM analyses focussing on the period between post-treatment to follow-up showed a significant time by treatment condition interaction, $F_{1,79.3} = 5.11$, $p < .03$, suggesting that symptom trajectories differed between groups over this period. However, a significant treatment group effect suggested that participants in the treatment group had significantly lower depression scores when scores were collapsed across the post-treatment and follow-up timepoints, $F_{1,\ 90.4} = 6.45$, $p < .02$.

4. Discussion

The present study aimed to test the effectiveness of the Deprexis program among adults presenting with depression symptoms in the severe range, most of whom were receiving heterogeneous forms of concurrent treatment. Of note, we examined the effects of an Internet intervention without personal support in a population that is comparatively difficult to treat, given the elevated symptom severity levels and high rates of depression chronicity as well as early onset. The main findings were: (1) that the effectiveness of Deprexis over a 3-month period, delivered without personal program support, could be replicated, compared to a CAU control condition; (2) that effects on secondary outcomes such as somatic symptoms and quality of life were weaker, with the exception of significant effects on anxiety in per-protocol analyses (marginally significant in ITT analyses); (3) that the treatment effects on depression severity could be largely maintained up to six months, although group differences were smaller at that point, (4) that concurrent antidepressant use might moderate the effects of the intervention, such that substantial benefit may occur among patients on antidepressants who use Deprexis as an adjunctive treatment tool. Additionally, we replicated previous findings by showing a high level (around 80%) of patient satisfaction with the program. We also showed for the first time that perceptions of early alliance/helpfulness after an average of three weeks could predict treatment response over 3 months quite powerfully, even when controlling for early symptom reduction.

We interpret these findings as further evidence supporting the specific effects of the Deprexis program on its intended outcome: depression reduction. There was also evidence for sustained effects after treatment discontinuation (i.e. group differences when the post-treatment to 6 month follow-up period was examined), although the differences between the treatment and control group were smaller at that point, and responder analyses revealed few significant group differences at follow-up. Several explanations may be invoked for these smaller differences at follow-up: Firstly, Deprexis usage was limited to a 90 day period. Our findings suggest that this may be too brief and should perhaps be changed to a longer period, at least for severely depressed individuals. Secondly, we note that even at 6 months, between-group effect sizes in the small to medium range were observed for the primary outcome (i.e. between 0.20 and 0.50). Thirdly, it is important to note that large improvements were observed among control participants, which may be explained, in part, by regression to the mean, although other processes (e.g. differential treatment-seeking) may also be involved. Indeed, it seems reasonable to expect that control participants would actively seek treatment when symptoms do not improve over time, particularly in a country such as Germany, where a great variety of inpatient and outpatient treatment services are available at no direct cost to patients (i.e. covered by mandatory health insurance) (Gaebel et al., 2009). Thus, the equivalence of the treatment and control groups over longer periods cannot be assumed, and future investigations should attempt to disentangle how differential treatment utilization might influence symptom trajectories over time.

In summary, though, these results replicate previous findings by demonstrating that considerable rates of depression reduction appear to occur as a consequence of using the program, with effect sizes closely resembling those observed in previous trials. The findings also suggest that the benefit of the program is particularly strong if combined with antidepressant medication (or possibly with psychotherapy). These findings suggest, then, that Deprexis is an effective Internet-based psychological intervention for adults reporting depression symptoms in the severe range. Given that several previous trials have shown similar effects, the effects associated with this program can be considered robust, in our opinion. Among participants who engaged with the program with reasonable intensity (at least four sessions and at least one hour), clinically meaningful symptom reductions of at least 5 points on the PHQ-9 (Kroenke et al., 2010) were observed in nearly 2 out of 3 cases at post-treatment, even though such change also occurred in the control condition in about 1 out of 3 cases.

An intriguing preliminary finding was that effects were particularly strong among those who were on concurrent antidepressants and used the program as an adjunctive treatment tool. We can only speculate about explanations for this finding. One potential explanation is that participants on antidepressants who participated in this trial were disenchanted with the effects of their medication and, therefore, strongly desired additional help. Once they received it from the online intervention, they may have benefitted particularly strongly. Another possible explanation is that simultaneously receiving antidepressants and engaging with cognitive–behavioural techniques produces synergistic effects. For example, antidepressants can improve hippocampal function, which facilitates learning and thus may potentiate the effects produced by cognitive–behavioural interventions (Craighead, 2014). It remains to be seen, though, whether this interaction effect can be replicated and, if so, whether broader dissemination of the program as an adjunct to antidepressant medication or psychotherapy is a feasible dissemination avenue. Such combinations of Internet-based and conventional treatments are regarded as promising future directions by many (Andersson and Titov, 2014). Indeed, the question of when, how, and for whom psychotherapeutic and pharmacological interventions should be combined is a topic of considerable research interest, and evidence currently supports the idea that chronically depressed patients, in particular, may benefit more than others from combination treatment (Craighead, 2014). Clearly, the ways in which software-based as well as face-to-face psychological interventions should be combined with pharmacotherapy deserve further research attention.

Our findings also showed that program users seem to develop a good sense early on of how helpful the intervention will eventually be for them, even as early as three weeks after first logging on. Early alliance/helpfulness ratings predicted outcome rather powerfully, consistent with decades of psychotherapy research suggesting that the alliance between patients and therapists is a robust predictor of outcome (Martin et al., 2000). An emerging literature is examining the role of the alliance in Internet-based treatments, although most studies have focussed on interventions in which personal support is included, and findings so far seem mixed, with the alliance sometimes predicting outcome and sometimes only weakly or not at all (Knaevelsrud and Maercker, 2006; Andersson et al., 2012b). Our findings suggest that exploring early impressions of perceived fit, helpfulness or common ground between patients' perceptions and the program's approach might be fruitful. Previous research suggests that early alliance ratings might reflect, in part, early symptom change or baseline patient characteristics (Webb et al., 2011; Klein et al., 2003); however, it is noteworthy that alliance ratings predicted pre–post symptom reduction even when controlling for baseline symptom severity and early symptom change. We acknowledge that the "alliance" between a program user and a software-program is not equivalent with the human alliance or emotional bond emerging between a patient and his or her therapist. Instead, what we measured with the adapted HAQ-11 may reflect users' sense of perceived helpfulness, plausibility and personal fit – and these early impressions appear to be good predictors of ultimate clinical benefit, even if they are not synonymous with the concept of the therapeutic alliance between two persons.

Several strengths as well as limitations of the trial ought to be noted. Among the strengths, diagnoses were confirmed with a validated structured interview, study attrition rate was low (<20%)[2], intervention uptake was good (80% of those who logged on received what we consider to be a minimally sufficient "dose"), and the study included a sample that was larger than many other trials in this field. However, limitations must be kept in mind, including limited power to detect small to moderate effects, the limited ability to ensure equivalence between the control and treatment groups over a longer period of time, lack of outcome ratings beyond self-reports, relatively short follow-up period, and the attenuated baseline depression range due to the high initial inclusion criterion. Some of these limitations will be rectified in the much larger EVIDENT trial (Klein et al., 2013), although that trial targets mild to moderate depressive symptoms. Other Deprexis studies are underway (e.g., see ClinicalTrials.gov: ISRCTN20165665 and NCT01663649 or German Clinical Trials Register [www.drks.de]: DRKS00003564), which may shed further light on questions such as moderation and mediation of effects, optimal deployment settings, and robustness of effects when using rater-based measures and considering longer follow-up periods.

In conclusion, we believe that these findings contribute to the literature by replicating and extending effects of an Internet-based depression intervention and by examining moderators and response predictors, such as concurrent treatment and early perceived helpfulness. More research is needed to address remaining methodological limitations and to improve our understanding of the psychological and biological mechanisms that explain how the kinds of treatment effects we observed unfold.

[2] Attrition was much lower in the present trial than one might expect for unsupported Internet interventions. At least two explanations for this seem plausible: Firstly, pre-treatment diagnostic interviews are rare in trials with unguided interventions, but they were used here and may have bolstered commitment to participate (Johansson and Andersson, 2012; Berger et al., 2011). Secondly, the relationship between depression severity and attrition perhaps deserves research attention because the high symptom load among more severely depressed participants might motivate them to use the intervention more, and complete assessments more diligently, than might be true for their less depressed counterparts.

Acknowledgments and Declaration of Conflict of Interest

This study received financial support by Gaia AG, the developer and owner of Deprexis, the Internet intervention examined in this study. BM, GJ, and MW are employed at Gaia, and JB was employed by Gaia during the duration of the study. None of the other authors are employed by Gaia or have received remuneration for participating in this project. The authors thank Galina Dedova, Julian van Ulardt and Dina Al-Saydali, who conducted the MINI interviews, and thanks are also due to all members of the EVIDENT research group, including Sandra Nolte, Viola Gräfe, David Rosenbaum, and Kristina Fuhr.

References

Alexander, L.B., Luborsky, L., 1986. The Penn Helping Alliance Scales. In: Greenberg, L.S. (Ed.), The Psychotherapeutic Process: A Research Handbook. Guilford Press, New York, NY, US, pp. 325–366.

Andersson, G., Cuijpers, P., 2009. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn. Behav. Ther. 38, 196–205. http://dx.doi.org/10.1080/16506070903318960.

Andersson, G., Titov, N., 2014. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatr. 13, 4–11.

Andersson, G., Paxling, B., Roch-Norlund, P., Östman, G., Norgren, A., Almlöv, J., Georen, L., Breitholtz, E., Dahlin, M., Cuijpers, P., 2012a. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother. Psychosom. 81, 344–355.

Andersson, G., Paxling, B., Wiwe, M., Vernmark, K., Felix, C.B., Lundborg, L., Furmark, T., Cuijpers, P., Carlbring, P., 2012b. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. Behav. Res. Ther. 50, 544–550.

Andrews, G., Titov, N., 2010. Is internet treatment for depressive and anxiety disorders ready for prime time? Med. J. Aust. 192, S45.

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

Arnberg, F.K., Linton, S.J., Hultcrantz, M., Heintz, E., Jonsson, U., 2014. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS ONE 9, e98118.

Attkisson, C.C., Greenfield, T.K., 1999. The UCSF Client Satisfaction Scales: I. The Client Satisfaction Questionnaire-8. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 2nd ed. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 1333–1346.

Barak, A., Hen, L., Boniel-Nissim, M., Shapira, N., 2008. A comprehensive review and a meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions. J. Technol. Hum. Serv. 26, 109–160. http://dx.doi.org/10.1080/15228830802094429.

Bassler, M., Potratz, B., Krauthauser, H., 1995. Der "Helping Alliance Questionnaire"(HAQ) von Luborsky. Möglichkeiten zur Evaluation des therapeutischen Prozesses von stationärer Psychotherapie. Psychotherapeut 40, 23–32.

Berger, T., Hämmerli, K., Gubser, N., Andersson, G., Caspar, F., 2011. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn. Behav. Ther. 40, 251–266. http://dx.doi.org/10.1080/16506073.2011.616531.

Blankers, M., Koeter, J.M., Schippers, M.G., 2010. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 12, e54.

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., Meyer, B., Huibers, M., Smit, F., van Straten, A., Warmerdam, L., Barkham, M., Bilich, L., Lovell, K., Liu, E.T.H., 2013. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. Br. Med. J. 346, f540.

Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155.

Craighead, W.E., 2014. Combination psychotherapy and antidepressant medication treatment for depression: for whom, when, and how. Annu. Rev. Psychol. 65, 267–300.

Cuijpers, P., Donker, T., Johansson, R., Mohr, D.C., van Straten, A., Andersson, G., 2011. Self-guided psychological treatment for depressive symptoms: a meta-analysis. PLoS ONE 6, e21274. http://dx.doi.org/10.1371/journal.pone.0021274.

Dear, B.F., Titov, N., Sunderland, M., McMillan, D., Anderson, T., Lorian, C., Robinson, E., 2011. Psychometric comparison of the generalized anxiety disorder scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. Cogn. Behav. Ther. 40, 216–227.

Elvins, R., Green, J., 2008. The conceptualization and measurement of therapeutic alliance: an empirical review. Clin. Psychol. Rev. 28, 1167–1187.

Eysenbach, G., CONSORT-EHEALTH Group, 2011. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J. Med. Internet Res. 13.

Fann, J.R., Berry, D.L., Wolpin, S., Austin-Seymour, M., Bush, N., Halpenny, B., Lober, W.B., McCorkle, R., 2009. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. Psycho-Oncology 18, 14–22.

Gaebel, W., Janssen, B., Zielasek, J., 2009. Mental health quality, outcome measurement, and improvement in Germany. Curr. Opin. Psychiatr. 22, 636–642.

Gandek, B., Ware, J.E., Aaronson, N.K., Apolone, G., Bjorner, J.B., Brazier, J.E., Bullinger, M., Kaasa, S., Leplege, A., Prieto, L., 1998. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. J. Clin. Epidemiol. 51, 1171–1178.

German Federal Statistical Office (Statistisches Bundesamt), 2014. 79% der Personen ab zehn Jahren nutzen das Internet (79% of persons from age 10 are using the internet). German Federal Statistical Office (Statistisches Bundesamt).

Gorwood, P., 2004. Generalized anxiety disorder and major depressive disorder comorbidity: an example of genetic pleiotropy? Eur. Psychiatry 19, 27–33.

Gräfe, K., Zipfel, S., Herzog, W., Löwe, B., 2004. Screening psychischer Störungen mit dem Gesundheitsfragebogen für Patienten (PHQ-D). Diagnostica 50, 171–181.

Gueorguieva, R., Krystal, J.H., 2004. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch. Gen. Psychiatry 61, 310–317.

Ivarsson, D., Blom, M., Hesser, H., Carlbring, P., Enderby, P., Nordberg, R., Andersson, G., 2014. Guided internet-delivered cognitive behavior therapy for post-traumatic stress disorder: a randomized controlled trial. Internet Interv. 1, 33–40.

Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59, 12.

Jenkinson, C., Layte, R., Jenkinson, D., Lawrence, K., Petersen, S., Paice, C., Stradling, J., 1997. A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? J. Pub. Health 19, 179–186.

Johansson, R., Andersson, G., 2012. Internet-based psychological treatments for depression. Expert. Rev. Neurother. 12, 861–870. http://dx.doi.org/10.1586/ern.12.63.

Johansson, R., Ekbladh, S., Hebert, A., Lindström, M., Möller, S., Petitt, E., Poysti, S., Larsson, M.H., Rousseau, A., Carlbring, P., 2012. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. PLoS ONE 7, e38021.

Kazdin, A.E., Blase, S.L., 2011. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect. Psychol. Sci. 6, 21–37.

Klein, D.N., Schwartz, J.E., Santiago, N.J., Vivian, D., Vocisano, C., Castonguay, L.G., Arnow, B., Blalock, J.A., Manber, R., Markowitz, J.C., Riso, L.P., Rothbaum, B., McCullough, J.P., Thase, M.E., Borian, F.E., Miller, I.W., Keller, M.B., 2003. Therapeutic alliance in depression treatment: controlling for prior change and patient characteristics. J. Consult. Clin. Psychol. 71, 997–1006. http://dx.doi.org/10.1037/0022-006X.71.6.997.

Klein, J.P., Berger, T., Schröder, J., Späth, C., Meyer, B., Caspar, F., Lutz, W., Greiner, W., Hautzinger, M., Rose, M., 2013. The EVIDENT-trial: protocol and rationale of a multicenter randomized controlled trial testing the effectiveness of an online-based psychological intervention. BMC Psychiatry 13, 1–10.

Knaevelsrud, C., Maercker, A., 2006. Does the quality of the working alliance predict treatment outcome in online psychotherapy for traumatized patients? J. Med. Internet Res. 8.

Kriz, D., Nübling, R., Steffanowski, A., Wittmann, W.W., Schmidt, J., 2008. Patientenzufriedenheit in der stationären Rehabilitation: Psychometrische Reanalyse des ZUF-8 auf der Basis multizentrischer Stichproben verschiedener Indikation (Patients' satisfaction in inpatient rehabilitation: psychometrical evaluation of the ZUF-8 based on a multicenter sample of different indications). Z. Med. Psychol. 17, 67–79.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16, 606–613.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2002. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom. Med. 64, 258–266.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B., 2010. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen. Hosp. Psychiatry 32, 345–359. http://dx.doi.org/10.1016/j.genhosppsych.2010.03.006.

Lecrubier, Y., Sheehan, D.V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., Janavs, J., Dunbar, G.C., 1997. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur. Psychiatry 12, 224–231.

Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med. Care 42.

Mackinnon, A., Griffiths, K.M., Christensen, H., 2008. Comparative randomised trial of online cognitive–behavioural therapy and an information website for depression: 12-month outcomes. Br. J. Psychiatry 192, 130–134.

Martin, D.J., Garske, J.P., Davis, M.K., 2000. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J. Consult. Clin. Psychol. 68, 438–450. http://dx.doi.org/10.1037/0022-006X.68.3.438.

Martin, A., Rief, W., Klaiberg, A., Braehler, E., 2006. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. Gen. Hosp. Psychiatry 28, 71–77.

McCracken, C., Dalgard, O.S., Ayuso-Mateos, J.L., Casey, P., Wilkinson, G., Lehtinen, V., Dowrick, C., 2006. Health service use by adults with depression: community survey in five European countries: evidence from the ODIN study. Br. J. Psychiatry 189, 161–167.

McMillan, D., Gilbody, S., Richards, D., 2010. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J. Affect. Disord. 127, 122–129.

Meyer, B., Berger, T., Caspar, F., Beevers, C.G., Andersson, G., Weiss, M., 2009. Effectiveness of a novel integrative online treatment for depression (deprexis): randomized controlled trial. J. Med. Internet Res. 11, e15. http://dx.doi.org/10.2196/jmir.1151.

Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374, 609–619.

Moritz, S., Schilling, L., Hauschildt, M., Schröder, J., Treszl, A., 2012. A randomized controlled trial of internet-based therapy in depression. Behav. Res. Ther. 50, 513–521. http://dx.doi.org/10.1016/j.brat.2012.04.006.

National Institute for Health and Clinical Excellence, 2009. Depression: The Treatment and Management of Depression in Adults (partial update of NICE clinical guideline 23).

Pinninti, N.R., Madison, H., Musser, E., Rissmiller, D., 2003. MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. Eur. Psychiatry 18, 361–364.
Richards, D., Richardson, T., 2012. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin. Psychol. Rev. 32, 329–342. http://dx.doi.org/10.1016/j.cpr.2012.02.004.
Salyers, M.P., Bosworth, H.B., Swanson, J.W., Lamb-Pagone, J., Osher, F.C., 2000. Reliability and validity of the SF-12 health survey among people with severe mental illness. Med. Care 38, 1141–1150.
Schmidt, J., Lamprecht, F., Wittmann, W.W., 1989. Satisfaction with inpatient management. Development of a questionnaire and initial validity studies. Psychotherapie, Psychosomatik. Med. Psychol. 39, 248–255.
Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166, 1092–1097.
Sugar, C., Sturm, R., Lee, T.T., Sherbourne, C.D., Olshen, R.A., Wells, K.B., Lenert, L.A., 1998. Empirically defined health states for depression from the SF-12. Health Serv. Res. 33, 911.
Titov, N., 2011. Internet-delivered psychotherapy for depression in adults. Curr. Opin. Psychiatr. 24.
Titov, N., Dear, B.F., Schwencke, G., Andrews, G., Johnston, L., Craske, M.G., McEvoy, P., 2011. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav. Res. Ther. 49, 441–452.
Titov, N., Dear, B.F., Johnston, L., Lorian, C., Zou, J., Wootton, B., Spence, J., McEvoy, P.M., Rapee, R.M., 2013. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS ONE 8, e62873.
van Gelder, M.M.H.J., Bretveld, R.W., Roeleveld, N., 2010. Web-based Questionnaires: the future in epidemiology? Am. J. Epidemiol. 172, 1292–1298.
Ware Jr., J.E., Kosinski, M., Keller, S.D., 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med. Care 34, 220–233.
Webb, C.A., DeRubeis, R.J., Amsterdam, J.D., Shelton, R.C., Hollon, S.D., Dimidjian, S., 2011. Two aspects of the therapeutic alliance: differential relations with depressive symptom change. J. Consult. Clin. Psychol. 79, 279–283. http://dx.doi.org/10.1037/a0023252.
World Medical Association, 2001. World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. 79. Bulletin of the World Health Organization, p. 373.
Zwarenstein, M., Treweek, S., Gagnier, J.J., Altman, D.G., Tunis, S., Haynes, B., Oxman, A.D., Moher, D., 2008. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 337.