Immediate and long-term effectiveness of adding an Internet intervention for depression to routine outpatient psychotherapy: Subgroup Analysis of the Evident Trial

Raphael Schuster, PhD 1*; Anton-Rupert Laireiter, Prof. 1 2; Thomas Berger, Prof. 3; Steffen Moritz, Prof. 4; Björn Meyer, Prof. 5; Fritz Hohagen, Prof. 6, Jan Philipp Klein, PhD 6

1 Department of Psychology, University of Salzburg, Austria

2 Faculty of Psychology, University of Vienna, Austria

3 Department of Clinical Psychology and Psychotherapy, University of Berne, Switzerland

4 Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Germany

5 Research Department, Gaia AG, Hamburg, and Department of Psychology, City, University of London, London, United Kingdom

6 Department of Psychiatry and Psychotherapy, Lübeck University, Lübeck, Germany

*Corresponding author:
Raphael Schuster, PhD.
Department of Psychology, University of Salzburg
Hellbrunnerstraße 34, 5020 Salzburg, Austria
raphael.schuster@sbg.ac.at

Key words: Blended therapy; computer-supported interventions; combined treatment; depression; psychotherapy
Public Health Significance Statement:

- Findings suggest that routine psychotherapy can be augmented by means of concomitant Internet treatment. Sustained gains were independent from therapeutic orientation of face-to-face psychotherapy, indicating general feasibility of adjunctive Internet treatment.

Running title: Combined Internet and face-to-face psychotherapy results in augmented effectiveness of routine outpatient psychotherapy

Highlights:

- This is the currently largest study on blended (face-to-face plus Internet-based) treatment for depression.
- Combined treatment resulted in augmented effectiveness of routine psychotherapy.
- Increases are comparable to those found in combined pharmaco- and psychotherapy studies.
- Treatment effects lasted over a 6-month follow-up period and were independent from therapeutic orientation of face-to-face psychotherapy.
- Limitations exist due to secondary data analysis and self-report assessment.
Objective: To examine immediate and long-term effectiveness of an adjunctive Internet intervention for depression in a large sample of patients undergoing routine psychotherapy.

Method: The current study evaluated a subgroup of patients from the Evident trial, a randomized investigation of a 12-week minimally guided Internet intervention (Deprexis) in the treatment of mild to moderate depression. 340 adults (mean age = 43.3 years; 71.7 % female) of the original sample received routine outpatient psychotherapy during the trial period, resulting in a standard psychotherapy group (n = 174) and an augmented therapy group (n = 166). Outcomes were assessed at baseline, post-treatment and 6-month follow-up.

Results: Intention-to-treat analyses indicated that combined treatment led to a greater reduction in symptoms of depression (effect size $d = 0.32; p = .002$), improved therapeutic progress ($d = 0.36; p = .003$), and higher mental health-related quality of life ($d = 0.34; p = .004$). There was no intervention effect on physical health-related quality of life. The same pattern was found at 6-month follow-up. Treatment success was independent from therapeutic orientation of combined face-to-face therapy. Median usage time of the adjunctive online intervention was 7.08 hours, indicating high patient engagement.

Conclusion: Results indicate that the adjunctive use of the investigated intervention can produce additional and lasting effects in routine outpatient psychotherapy for mild to moderate levels of depression. The study adds to the ongoing literature on augmented effects of blended treatment. Future studies should investigate different types of blends in diverse populations by means of change-sensitive assessment strategies.
Introduction

Over the last two decades, the evidence base for Internet interventions has grown steadily and first routine online clinics have emerged (Hedman et al., 2014; El Alaoui et al., 2015; Titov et al., 2018; Wells et al., 2018). While unguided Internet interventions provide wide reach at potentially low cost (Karyotaki et al., 2017), programs that include guidance have been shown to be associated with higher effectiveness and adherence rates (Baumeister, Reichler, Munzinger & Lin; 2014; Wells et al., 2018). For this reason, many Internet interventions entail some form of written guidance, or they are blended with occasional personal sessions (e.g. diagnostic assessment, or treatment introduction).

Recent research has shown that Internet interventions might also help to improve face-to-face therapy. In a first review on so-called blended cognitive behavioural therapy (bCBT), Erbe and colleagues (2017) identified 44 studies on blended treatment. Since then additional high quality studies have been carried out (Kleiboer et al., 2016; Thase et al., 2017; Vernmark et al., 2018), and several of the above mentioned online clinics offer routine blended care (Titov et al., 2018).

The term “blended therapy” usually refers to face-to-face treatment augmented with one or more digital intervention elements (e.g. Internet-, app-, or VR-based interventions). Erbe and colleagues (2017) distinguish between integrated and sequential interventions, with the latter including Internet-based aftercare. Within integrated formats, in which Internet interventions and face-to-face therapy are applied simultaneously, a further distinction can be made between adjunctive (Berger et al., 2018; Zwerenz et al., 2017) and highly integrated interventions. In many adjunctive treatments, both treatment strands (face-to-face and digital elements) remain largely independent. Highly integrated treatments, by contrast, are designed as new treatment rationales (cf. Kooistra et al., 2016; Schuster et al., 2019) or as a repository of applications to be selected according to the individual course of face-to-face therapy.
(Månsson, Ruiz, Gervind, Dahlin, & Andersson, 2013; Månsson, Klintmalm, Nordqvist, & Andersson, 2017). Furthermore, it is likely that digital assessment and digital treatment will merge into new approaches (e.g. EMI, ecological momentary intervention). A short overview of empiric formats is provided in Figure 1.

****** Figure 1 about here ******

Usually blended interventions are based on cognitive-behavioural therapy (CBT). Postulated advantages are, for example, the improved self-management of patients, beneficial effects on treatment transfer, repetition of CBT materials, or reduced travel times (Titzler, Saruhanjan, Berking, Riper, & Ebert, 2018; Urech et al., 2018; van der Vaart et al., 2014), as well as comparably high acceptance amongst stakeholders (Topooco et al., 2017; Schuster, Topooco, Keller, Radvogin, & Laireiter, 2020). Risks and limitations can exist with regard to possible over-structuring of therapy (Schuster, Sigl, Berger, & Laireiter, 2018), problems with non-verbal communication (van der Vaart et al., 2014), or additional workload for therapists (Titzler et al., 2018).

Blended treatments aim at increasing efficiency in terms of reduced therapist time, or in terms of enhanced treatment effects. Non-inferiority studies consistently found comparable outcomes in time-reduced treatments (Marks, Kenwright, McDonough, Whittaker, & Mataix-Cols, 2004; Wright et al., 2005). Most recent studies support these findings. For example, Thase and colleagues (2017) found comparable outcomes for a treatment in which face-to-face sessions were reduced by two thirds. The study included 154 medication-free patients suffering from major depression (MDD). In addition, a large European multicentre study on depression (eCompared, www.ecompared.eu) is currently investigating the cost-effectiveness of time-reduced bCBT over standard CBT (Kleiboer et al., 2016). Preliminary results of a pilot study (N = 102) indicate potential savings on provider- but not on societal-level (Kooistra et al., 2019). As an efficient alternative to individual therapy, blending might also
improve the delivery of group treatments, which have been tested for depression (Schuster et al., 2018; Schuster et al., 2019), anxiety (Gruber, Moran, Roth, & Taylor, 2001; Newman, Przeworski, Consoli, & Taylor, 2014), or hoarding (Fitzpatrick et al., 2018; Ivanov et al., 2018).

With regard to enhanced treatment effects, several superiority studies with different intervention designs have been carried out in the past. A relatively large number of studies investigated sequential aftercare programs, which frequently led to enhanced effects (Bauer, Wolf, Haug, & Kordy, 2011; Ebert, Tarnowski, Gollwitzer, Sieland, & Berking, 2013; Hennemann, Farnsteiner, & Sander, 2018). For simultaneous bCBT (with digital support during face-to-face therapy), Carroll and colleagues (2009; 2014) published a series of articles, indicating that relapse rates and abstinence times for substance-related addictions could persistently be improved by the use of individual- and group-based bCBT. This finding is supported by comparable studies (cf. Christensen et al., 2014). For depression, however, only few studies exist (cf. Berger, Krieger, Sude, Meyer, & Maercker, 2018; Sethi, Campbell, & Ellis, 2010; Zwerenz et al., 2017). Berger and colleagues (2018) investigated the effects of outpatient psychotherapy supplemented with an adjunctive Internet program ($N = 98$).

Enhanced effects ($d = 0.51$) were observed immediately after treatment, but during follow up ($N = 44$) these effects diminished ($d = 0.28$) and failed to reach statistical significance. In a second study, Zwerenz and colleagues (2017) found comparable results using the same intervention in addition to routine inpatient treatment ($N = 229$). Compared to the active control group (computer-assisted psychoeducation over 12 weeks) the combined treatment again resulted in enhanced effects ($d = 0.44$). The corresponding follow-up results have been published recently (Zwerenz et al., 2019).

To sum up, the evidence base for Internet interventions is growing rapidly, and the developed interventions might also serve to enhance face-to-face therapy. Apart from considerable
evidence for sequential bCBT (e.g. in aftercare), the concomitant treatment of depression by face-to-face therapy together with Internet interventions remains rarely investigated. At this point, ongoing evidence indicates that such interventions could actually lead to augmented effects in the treatment of depression, but more evidence building on preferably larger sample sizes is required.

The purpose of the present study is to investigate the effects of an adjunctive treatment for mild to moderate depression in routine care. It therefore reports on re-analyses of a sufficiently large subsample \((N = 340)\) of the Evident trial \((N = 1013;\) Klein et al., 2016), which received outpatient psychotherapy during the trial period. According to existing literature, we hypothesized that combined treatment would result in enhanced immediate effects, and that those effects would be maintained during the follow-up period. Furthermore, we were interested in potential predictors of therapy success and intervention usage. A considerable number of studies on Internet interventions tested relevant predictors, such as age, attitudes towards digital treatment, or initial depression level (cf. Beintner et al., 2019). Regarding simultaneous bCBT for depression, evidence is clearly limited by the small number of conducted studies. However, two previous studies investigated potential predictors of treatment utilization and effects in the context of the Evident study (Fuhr et al., 2018; Schneider et al., 2018). Findings revealed, for example, a positive relation between higher age or higher depressive symptoms and use of the intervention. Relations with treatment outcome, however, were less clear (cf. Schneider et al., 2018; Fuhr et al., 2018). We therefore investigated if those variables would exhibit a comparable pattern in the selected subsample of the Evident trial, which received the Internet intervention as an adjunct to routine psychotherapy.

Method
This study analysed a subsample which was taken from the EVIDENT trial (N = 1013; Klein et al., 2016). Adhering to the CONSORT and the CONSORT-EHEALTH statement, this large randomized multicentre study investigated the effectiveness of the Internet intervention Deprexis. The trial was conducted at five sites in Germany after ethical approval by the Ethics Committee of the German Psychological Association (SM042012), and after registration at ClinicalTrials.gov (NCT01636752). In the principal study, participants were either randomised to care as usual (CAU) or to CAU plus Deprexis. CAU of the original trial comprised any form of treatment participants actively sought, including antidepressant medication (50.2 %), and/or psychotherapy (33.6 %), and/or outpatient psychiatric treatment (27.1 %), and/or treatment by general practitioner, or no formal treatment.

The present subsample (N = 340), however, received concomitant routine psychotherapy during the trial period. Selection of the subsample was based on patients indicating to having undergone routine psychotherapy at post-assessment of the EVIDENT trial. Of these patients, n = 174 (17.2 % of the original sample) received psychotherapy and n = 166 (16.4 %) received psychotherapy plus the adjunctive Internet intervention Deprexis. Thus, the first group constituted the control condition, while the latter constituted the blended treatment condition of the present study. Due to secondary data analysis, the exact onset of psychotherapy did not coincide with time of randomization to both groups.

Participants

Recruitment of the original study began in 2012 and ended in 2015 (see the CONSORT flow chart in Klein et al. 2016, p. 222). The current sub-sample included all patients undergoing psychotherapeutic treatment at the time of the study (N = 340). Patients of the original study
were either recruited via outpatient medical or psychological clinics, or via a multimodal recruitment strategy, including health insurance companies and public media (e.g. newspaper or radio). The main inclusion criterion was presence of self-reported mild to moderate levels of depression, defined as a score of 5 to 14 scale points on the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001). Eligible patients were able to communicate in German, and their age ranged from 18 to 65 years. Acute suicidality, or a lifetime diagnosis of schizophrenia or bipolar disorder constituted further exclusion criteria. Written informed consent was obtained online prior to baseline assessment. Psychiatric diagnoses were based on the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), and acute suicidality was assessed by status of current suicidal ideation together with past suicide attempts.

Interventions

Patients in both groups received full routine outpatient psychotherapy in Germany. Psychotherapy in Germany is part of public health care, and, therefore, widely accessible and free of charge. All included patients of both study conditions underwent routine psychotherapy by trained psychotherapists from a low number of officially recognized therapeutic orientations. The assignment of therapeutic orientation was based on self-report, and half of the sample received one of the two common forms of psychotherapy (CBT: 23.7%; or psychodynamic therapy (PDT): 22.5%). A minor proportion of 2.7% received systemic therapy. The remaining patients either indicated to have received talk therapy or client-centred therapy (20.1%), or were not able to specify the respective treatment orientation (26.0%).

Patients in the bCBT study condition ($n = 166$) additionally received Deprexis as adjunctive Internet intervention together with the above described forms of routine psychotherapy. Deprexis is a 12-week CBT-based online programme which has been evaluated for mild to
moderate levels of depression (Berger, Hämmerli, Gubser, Andersson, & Caspar, 2011; Meyer et al., 2009), as well as for severe depression (Meyer et al., 2015). A comprehensive overview of studies is presented by Twomey, O’Reilly, & Meyer (2017). The programme consists of ten modules (e.g. cognitive restructuring, behavioural activation, acceptance and mindfulness, problem solving) and a summary module, presented by an algorithm to increase tailoring of text- or audio-based elements (including automatized shifts between modules). Automatized daily messages, delivered either by SMS or e-mail, supported the programme, and all patients with moderate depressive symptoms (PHQ-9 > 9 scale points; \( n = 109 \)) received weekly asynchronous feedback by a trained, remotely operating supporter. The generic feedback was restricted in time (approximately 5 minutes per message) and did not apply specific therapeutic techniques, but was rather aimed to motivate patients to engage with the provided online modules. Further information on the intervention and the provided guidance is presented in the study protocol of the original trial (Klein et al., 2013).

Measures

The primary outcome was the digitalized version (Erbe, Eichert, Rietz, & Ebert, 2016) of the 9 item form of the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), a common scale to assess severity of self-rated depression on a 4 step Likert-scale. The observed internal consistency of the PHQ-9 can be classified as good (Cronbach’s alpha = .821). According to the authors, mild depression severity can be classified by scorings of 5 to 9 scale points, whereas moderate depression severity is represented by scorings between 10 to 14 scale points (Kroenke et al., 2001).

Furthermore, a measure of health-related quality of life (Short-Form Health Survey, SF12; Ware, Kosinski, & Keller, 1996), and a measure for the Evaluation of Psychotherapeutic Progress (Lutz et al., 2009) were employed in the study. The FEP-2 is a generic process questionnaire with four latent factors (symptom severity, interpersonal problems,
psychological incongruence, and personal wellbeing). With Cronbach’s alpha = .901 internal consistency can be classified as excellent. The SF-12 is composed of the subscales mental health (MH), and physical health (PH). As the scale includes three different types of Likert-scales we did not calculate internal consistency. All measures were administrated online and took place at pre-, post-, and 6-month follow-up.

Statistical analyses

Statistical analyses were based on the intent-to-treat principle (ITT), which included all randomized patients of the selected subsample. The analyses were performed using SPSS 25, and linear mixed models (LMM) were applied to test treatment by time interactions of the 2 x 3 factorial design. Under the assumption of random missing data, LMM yields the advantage of providing unbiased estimates of treatment effects (Bell, Fiero, Horton, & Hsu, 2014). In order to account for dependence of repeated measures, restricted maximum likelihood estimation (REML) with compound symmetry (CS) was chosen in the model specifications. To optimize statistical power, baseline assessments entered the analysis as covariates (van Breukelen, 2006). Rates of clinically significant improvement for PHQ-9 were defined as less than nine scale points during post-assessment (McMillan, Gilbody, & Richards, 2010), combined with a 50 % symptom reduction. We used 50 % change from baseline as this criterion classifies cases independently from their baseline level of depression (Karin, Dear, Heller, Gandy, & Titov, 2018). For reliable deterioration (RD) two criteria were defined. The standard RD criterion was defined as a 50 % increase from baseline. As this criterion should identify a low number of subjects (in our case 7 % of the sample), statistical power to detect differences in deterioration rates between groups is low. We therefore defined suboptimal to reliably deteriorated cases as any deterioration above 25 % increase from pre- to post-assessment, resulting in 12.6 % of critical cases in the sample.
Depending on their distribution, demographic variables of conditions were tested with \( t \) tests for independent samples or with corresponding chi-squared tests to estimate success of randomisation. Potential predictors of overall therapy success and of adherence with the online part of the intervention were investigated by means of multiple linear regression, in which baseline severity entered as independent predictor variable. Further predictor variables from previous Deprexis studies (Fuhr et al., 2018; Schneider et al., 2018) constituted demographic variables (age, gender, education, and relation status), as well as important clinical variables (therapeutic orientation of concomitant psychotherapy, attitudes towards Internet interventions, medication status, and level of baseline depression). Additionally, we tested if adherence with the Internet intervention (e.g. time spent on platform) influenced treatment success in a meaningful way. Adherence was assessed objectively by tracking number of chats and minutes spent with the intervention.

All presented effect sizes were calculated according to Cohens’\( d \) (Cohen, 1988), by subtracting the estimated mean of the experimental group from the mean of the control group. Power calculation was based on G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) and revealed statistical power of 80 % to detect between group differences in a factorial study design with a treatment effect of \( d = 0.25 \), and correlation of repeated measures of \( r = 0.3 \). Note that the baseline covariate is not implemented into the analysis and we do not report the interaction term here, making the power calculation a conservative estimate.

Results

Patient characteristics

There were neither significant differences in baseline characteristics (Table 1), nor in psychotherapeutic orientation of face-to-face treatment between both groups (\( p = .119 – \)
and the study exhibited excellent retention of 88.8% for the follow-up period. If not corrected for alpha inflation, however, marital status ($\chi^2 = 2.32, p = .021$) and number of depressed episodes ($\chi^2 = 2.05, p = .040$) would have deviated significantly. Overall, women constituted the majority of the sample (71%) and level of education can be considered representative for Germany (54% high school graduate or higher). Participants’ age ranged from 19 to 65 years ($M = 43.3, SD = 10.9$). About 84% of the sub-sample fulfilled the criteria for full MDD with a high number of self-reported recurring depressive episodes. Regarding time of randomization, patients on average received 12 therapy sessions ($SD = 7.6$) before pre-assessment, resulting in expectably smaller pre-to-post treatment effects for the present subgroup analysis. A higher number of sessions was received by psychodynamic patients ($M = 18, SD = 17$) and a lower number by patients undergoing systemic therapy ($M = 6.1, SD = 1.4$) Details on both groups’ demographic and clinical characteristics are presented in Table 1.

****** Table 1 about here ******

Intervention usage

A total of 166 participants was randomized to the 12 week blended format, with a small fraction not using the program (0.9% less than two sessions or 60 minutes). A somewhat larger fraction only exhibited moderate program usage (4.1% less than four sessions or 120 minutes). The median usage time was 425 minutes ($Q_{25-75} = 288 – 600$ minutes) or 7.08 hours. The median number of therapist messages was 9 ($Q_{25-75} = 6 – 11$).

Main outcome

Based on hierarchical analyses (LMM) of the intention-to-treat sample, bCBT resulted in higher treatment effects on the primary study outcome (PHQ-9: $F(2, 646.56) = 6.21, p = .002; d$
= 0.32) compared to psychotherapy alone. Those differences remained stable at 6-month follow-up (PHQ-9: \( F_{(1, 342.86)} = 7.67, p = .006; d = 0.31 \)). Detailed information on estimated scale means, standard deviations and effect sizes is provided in Table 2. Clinically significant improvement was observed in 20.7 % of cases in the standard condition and 30.1 % in the augmented group, indicating more clinically significant improvement for the latter (\( x^2 = 1.99, p = .046 \)). Rates of reliable deterioration (> 50 % from baseline) were 8 % for the standard group and 6 % for the augmented group, indicating no significant differences (\( x^2 = 0.73, p = .468 \)). Finally, rates for suboptimal to reliable deterioration (deteriorations > 25 %) were 15 % in the standard group and 10.2 % in the augmented group, indicating significant differences in favour of the augmented treatment (\( x^2 = 2.40, p = .017 \)).

Secondary outcomes

Most self-rated secondary outcomes indicated superior effects at termination of treatment. While both psychological constructs, the SF-12 for mental health related quality of life (\( F_{(2, 610.83)} = 5.63, p = .004, d = 0.34 \)), and the FEP-2 for overall therapy progress (\( F_{(2, 635.89)} = 5.85, p = .003, d = 0.36 \)), indicated augmented effects, no increases have been found for physical health-related quality of life (SF-12 PH: \( F_{(2, 606.28)} = 0.72, p = .488; d = 0.11 \)).

Regarding stability of secondary outcomes, a consistent pattern emerged. Together with the primary outcome, overall therapy progress (FEP-2: \( F_{(1, 315.53)} = 7.42, p = .007, d = 0.35 \)), as well as mental health related quality of life (SF-12 MH: \( F_{(1, 307.85)} = 3.59, p = .033, d = 0.22 \)) indicated superiority of bCBT. Again, physical health-related quality of life was comparable in both groups (SF-12 PH: \( F_{(1, 295.42)} = 0.38, p = .536, d = -0.05 \)). Detailed information on scale means, standard deviations and effect sizes is presented in Table 2.

***** Table 2 about here *****
Predictors of therapy success and intervention usage

Depression severity at post-treatment was neither predicted by demographic variables (age, gender, level of education, or relation status) nor by parameters of program utilization. Neither of the $z$ standardized utilization variables predicted the primary outcome PHQ-9 at post-treatment ($p = .101 - .901$), nor at 6-month follow-up ($p = .027 - .910$), after Bonferroni adjustment for alpha inflation had been applied. It was only predicted by baseline depression severity ($\beta = .268, p = .002$). Regarding prediction of intervention usage, baseline depression severity again emerged as clear predictor ($\beta = .221, p = .004$), indicating a positive relation between severity of depression and utilization of the Internet intervention. Age was also positively related to increased usage time ($\beta = .208, p = .008$). As a last aspect, therapeutic orientation influenced intervention usage negatively ($\beta = -.280, p = .002$) if patients indicated to have undergone client-centred or talk therapy.

Discussion

This study analysed the effects of combined Internet-based and face-to-face treatment for mild to moderate depression. In a subsample ($N = 340$) of the Evident trial (Klein et al., 2016) the Internet intervention Deprexis was randomly added to outpatient psychotherapy, making the study the presently largest investigation on long-term effects of bCBT. We found that combined treatment was associated with immediate and long-term benefits for all assessed psychological parameters (depression, mental health, and therapeutic progress), but not for physical-health related quality of life (SF-12 PH). During 6-month follow-up those increases remained stable, even though mental-health related quality of life diminished slightly ($d_{\text{diff}} = \ldots$)
Thus, the present results indicate that the effectiveness of routine psychotherapy can be augmented by adjunctive Internet treatment. As Berger and colleagues (2018) suggest, a simple explanation for this effect is the additional time patients invested in the treatment. The median patient spent about 7 hours with the minimally supported intervention, which can be considered a reasonable amount of extra therapy time. The argument of treatment intensification seems particularly plausible, as the number of weekly sessions predicted the outcome of psychotherapy for depression in previous studies (Cuijpers et al., 2013). Furthermore, homework engagement generally relates to success of CBT (Conklin & Strunk, 2015; Kazantzis, Whittington, & Dattilio, 2010), and bCBT has been described as treatment facilitator, fostering the reconsolidation of CBT materials, as well as the transfer into daily life (Etzelmueller, Radkovsky, Hannig, Berking, & Ebert, 2018; Schuster et al., 2018; Urech et al., 2018). However, the assumption that more online activity (e.g. time spend with the intervention, or number of logins) would result in pronounced decreases of depression was not supported by the present study ($\beta = 0.018$ to $-0.043$) - even though tracked usage patterns exhibited sufficient variability, and results from the principal study indicated a weak but statistically significant relation (Klein et al., 2016).

Two possible explanations are that the relation between intervention usage and therapy outcome might be more complex, or that intervention usage (formalized by online log files) is not synonymous with effective engagement or actual therapy adherence in blended interventions (Sieverink, Kelders, & van Gemert-Pijnen, 2017; Yardley et al., 2016). Regarding immediate effectiveness, reported findings are in line with results from first bCBT trials on depression (Berger et al., 2018; Sethi et al., 2010; Zwerenz et al., 2017). At this, present effects appear a bit smaller ($d_{diff} = 0.1$) than in two current studies (Berger et al., 2018; Zwerenz et al., 2017). While the use of different questionnaires for assessing depression symptoms limits the interpretability of these marginal differences, smaller effects might also
relate to differences in sampling. Both studies included patients with severe levels of depression (rather than mild to moderate levels), and meta-analyses suggest greater symptom reductions with increased levels of depression (Bower et al., 2013). Moreover, the delayed onset of face-to-face psychotherapy could have limited observable effects. On average, 12 therapy sessions were consumed before start of face-to-face treatment. To provide a reference, German psychotherapy is free of charge and the average treatment (across therapeutic orientations) endures 49.8 sessions (Rabe-Menssen et al., 2011). It is therefore probable, that carefully designed studies in unrestricted samples would result in even higher gains in routine care. Still, even the comparably moderate effects of the present study should be considered as meaningful as they approach effects from combined antidepressant and psychotherapy treatment (Cuijpers et al., 2014; Hedge’s $g = 0.43$). As a last aspect, clinically significant improvement was significantly higher in the blended treatment group, with comparable proportions as reported in the German outpatient study by Berger and Colleagues (2018). Regarding reliable deterioration, present findings suggest no significant difference between standard and augmented treatment. As reliable deteriorations typically constitute comparably rare events our study was underpowered to detect potential deviations, limiting interpretability of present findings. When cases of mild deterioration are added (> 25 % from baseline), however, the augmented treatment resulted in improved outcomes. This pattern also fits the reported improvements of therapeutic progress (assessed by FEP-2). Taken together, results indicate that adding an adjunctive Internet intervention can lead to statistically and clinically relevant improvements of routine psychotherapy for depression. Furthermore, the blended format could contribute to reductions of undesired effects, at least when clinically mild deteriorations are included. Concerning long-term effectiveness of bCBT, the magnitude of observed improvement ($d = 0.22$ to 0.35) is consistent with first findings from one comparable study ($N = 98$) (Berger et
Due to the considerably larger sample size, the present study possessed statistical power to detect the effect of adjunctive treatment more reliably. Supporting evidence can be found for other mental disorders, such as substance-related addictions (Carroll et al., 2009; Carroll et al., 2014), as well as for sequential bCBT in aftercare (Bauer et al., 2011; Ebert et al., 2013; Hennemann et al., 2018), and the wider field of Internet interventions (Andersson, Rozental, Shafran, & Carlbring, 2018). As potential improvement, long-term depression was only assessed by means of self-report questionnaires. Future studies should therefore include clinician ratings over a follow-up period of one year, which is in accordance with current mental health research priorities (Wykes et al., 2015). Taken together, the present study can be regarded as ongoing evidence for augmented and enduring effects of combined Internet-based and face-to-face depression treatment.

Further strategies to optimize statistical power and to reduce required sample size are advisable for future trials, as the improvement of current gold-standard treatments probably will lead to moderate effects and conducting sufficiently large RCTs will impose high financial costs. In this context, digital psychiatry could facilitate the implementation of intense assessment, for example by weekly assessment or by short pre-post ecological momentary assessment (EMA). The magnitude of expectable gains in statistical power is considerable and could help to mitigate the current situation without necessarily increasing sample sizes. For example, a hypothetical RCT with point assessments of psychopathology (pre-post assessment by questionnaire) would need 90 patients, but 8 (bi-) weekly assessments during the active trial phase reduce the required sample size to 42 - 50 patients. Additionally, short EMA has previously been tested to substitute point assessments of psychopathology in randomized controlled trials (Moore, Depp, Wetherell, & Lenze, 2016; Vork et al., 2019; Schuster et al., 2020). As app-based assessments can easily be implemented into bCBT (Newman et al., 2014; Kleiboer et al., 2016; van de Ven et al., 2017; Colombo et al., 2018),
digital mental health interventions could facilitate both, the provision and evaluation of psychological treatments.

As a last aspect, we investigated potential predictors of bCBT. At this, baseline depression severity emerged as the only significant predictor, suggesting effectiveness of bCBT independently from basic patient characteristics. The present study also explored the impact of therapeutic orientation on combined therapy, with results indicating that adjunctive online treatment can improve face-to-face psychotherapy regardless of its orientation. This finding fits both previous bCBT superiority trials, as one was conducted in a psychodynamic inpatient setting (Zwerenz et al., 2017), while the other investigated outpatient CBT (Berger et al., 2018). Regarding utilization of the Internet intervention, engagement was higher in more depressed individuals and it was positively related to age. This result is in accordance with prior findings (Fuhr et al., 2018; Schneider et al., 2018), and may be related to increased motivation or conscientiousness of those groups. Furthermore, patients indicating to have participated in client-centred (CP) or talk therapy exhibited less compliance with tracked online tasks. This result, however, needs to be interpreted with caution as therapy orientation was based on self-report, and client-centred and (more general) talk therapy can be confounded due to similarity in German language. Taken together, the present findings suggest no relation between demographic characteristics or type of face-to-face treatment and effectiveness of bCBT. Additionally, some predictors of intervention usage have been found, but the literature in the field for now remains inconclusive (Beatty & Binnion, 2016).

This study has noteworthy strengths and limitations. Amongst its most important strengths, the principal study (Klein et al., 2016) was based on a multiple recruitment strategy and those involved in the current study received publicly funded high quality routine psychotherapy from several different clinics, settings, and therapeutic orientations in Germany. Furthermore, the adjunctive use of the Internet intervention reduced potential interferences with the
respective routine treatment. Regarding its evaluation, the study was based on assessments of psychopathology, therapeutic progress and mental health. During the evaluation, a consistent pattern of improvements in psychological but not in physical scales emerged, indicating low risk of overall response bias. As a last aspect, the study exhibited very low attrition during the follow-up period.

Regarding the study limitations, issues of intervention and trial design need to be considered when interpreting its findings. Firstly, due to secondary data analysis the onset of psychotherapy did not coincide with the time of randomization to the study groups. Instead, randomization to the actual trial frequently lagged behind the onset of psychotherapy. Therefore, the reported pre-to-post effects could underestimate the true impact of routine psychotherapy, while the influence on reported between-group effect sizes appears less impactful (cf. paragraph three of the discussion section). Secondly, the investigated blended format was based on an adjunctive algorithm-driven online intervention which originally was not designed to operate together with face-to-face therapy (Berger et al., 2011; Meyer et al., 2009). It is therefore not specifically adjusted to face-to-face treatment, but rather works in parallel on a relatively low level of integration. Thus, the effects of alternative designs, for example with digital CBT-elements on-demand (Månsson et al., 2013; Månsson et al., 2017) or with more manualized blends (Koositra et al., 2016; Schuster et al., 2019), might present differently. Thirdly, the standard treatment condition only received routine psychotherapy, but lacked an active online condition (e.g. online psychoeducation), which potentially could have reduced the observed effects (Munder et al., 2018) or the risk for introducing bias. Contrary to this limitation, however, the first study with active online control condition found superior effects of comparable magnitude (Zwerenz et al., 2018), and we found a consistent pattern of differential effects for psychological and physical variables in the present study. Still, future studies should implement active controls in order to reduce potential sources of bias (e.g.
expectancy bias). Fourthly, regardless of the comparably large overall sample size certain study findings should be interpreted with caution due to limited statistical power. According to the provided power estimation, the study can be regarded sufficiently sensitive to detect proposed effects of blended treatment. Furthermore, power for multiple linear regression (deviation from zero in a fixed model, with medium effect size of $f^2 = 0.15$) for 10 predictors was 99 %; and 96 % for a small effect of $f^2 = 0.08$. However, power to detect a medium correlation of $r = 0.3$ for one unevenly distributed variable (low program utilization) was only 50 %. As a last aspect, the classification of face-to-face treatment was based on self-report. While the proportions of CBT and PDT appear reliable, the CP group might have confounded the subcategory client-centred therapy with talk therapy (as a non-specific expression for psychotherapy in general). Therefore, this groups’ reduced intervention usage cannot be attributed with safety, and, thus, needs to be clarified in further studies.

Conclusion

Taken together, the present findings add to the ongoing literature on augmented effects of combined Internet-based and face-to-face treatment for MDD. The effects of combined treatment correspond to ongoing evidence from previous studies, with clinically meaningful increases which persisted over time. Potential predictors of intervention usage were identified, but treatment effectiveness was largely independent from such variables. Thus, bCBT should be further investigated in the striving to optimize current gold-standard treatments. Future studies should test diverse interventions in active control designs, together with more sensitive assessment strategies of psychopathology.

References


Beintner, I., Görlich, D., Berger, T., Ebert, D. D., Zeiler, M., Camarano, R. H., ... & ICare Consortium. (2019). Interrelations between participant and intervention characteristics, process variables and outcomes in online interventions: A protocol for overarching analyses within and across seven clinical trials in ICare. *Internet Interventions, 16*, 86-97.


Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. Journal of Clinical Epidemiology, 59(9), 920-925.


<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
<th>Journal</th>
<th>Pages</th>
</tr>
</thead>
</table>
Figure 1. Types of blended treatment.

\(^a\) cf. Ebert et al., 2013. \(^b\) cf. Berger et al., 2018. \(^c\) cf. Kooistra et al., 2016. \(^d\) cf. Månsson et al., 2017
Table 1. Demographic and clinical variables presented by treatment condition. Frequencies, means and standard deviations (in brackets).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Blended treatment</th>
<th>Standard treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 166$</td>
<td>$n = 174$</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>$%$</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>119</td>
<td>71.7</td>
</tr>
<tr>
<td>Male</td>
<td>47</td>
<td>28.3</td>
</tr>
<tr>
<td>Mean age ($SD$)</td>
<td>43.3 (11.3)</td>
<td>43.3 (10.4)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>37</td>
<td>22.3</td>
</tr>
<tr>
<td>Partnership</td>
<td>41</td>
<td>24.7</td>
</tr>
<tr>
<td>Married</td>
<td>65</td>
<td>39.2</td>
</tr>
<tr>
<td>Divorced</td>
<td>20</td>
<td>12.0</td>
</tr>
<tr>
<td>Widowed</td>
<td>3</td>
<td>1.8</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compulsory or technical school</td>
<td>80</td>
<td>48.2</td>
</tr>
<tr>
<td>High school graduate or higher</td>
<td>86</td>
<td>51.8</td>
</tr>
<tr>
<td>Diagnosis (self-reported diagnoses)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression (MDD)</td>
<td>139</td>
<td>83.7</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>61</td>
<td>36.7</td>
</tr>
<tr>
<td>PTSD</td>
<td>24</td>
<td>14.5</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>18</td>
<td>10.8</td>
</tr>
<tr>
<td>History of depression (self-report)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 2 depressive episodes</td>
<td>23</td>
<td>13.9</td>
</tr>
<tr>
<td>2-5 depressive episodes</td>
<td>65</td>
<td>39.2</td>
</tr>
<tr>
<td>6-10 depressive episodes</td>
<td>39</td>
<td>23.5</td>
</tr>
<tr>
<td>&gt; 10 depressive episodes</td>
<td>39</td>
<td>23.5</td>
</tr>
<tr>
<td>Medication</td>
<td>139</td>
<td>83.7</td>
</tr>
<tr>
<td>Combined treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td>49</td>
<td>29.5</td>
</tr>
<tr>
<td>PDT</td>
<td>40</td>
<td>24.1</td>
</tr>
<tr>
<td>CP</td>
<td>48</td>
<td>28.9</td>
</tr>
<tr>
<td>Other</td>
<td>29</td>
<td>17.5</td>
</tr>
</tbody>
</table>
Note. MDD = major depressive disorder; PTSD = post-traumatic stress disorder; SD = standard deviation; CBT = cognitive behavioural therapy; PDT = psychodynamic therapy; CP = client-centred or talk therapy.

Table 2. Descriptive statistics and effect sizes of main and secondary outcomes.

<table>
<thead>
<tr>
<th></th>
<th>Estimated means</th>
<th>Between effect sizes (estimated means)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre (SE)</td>
<td>Post (SE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blended therapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>($n=166$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard therapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>($n=174$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>10.51 (0.298)</td>
<td>10.24 (0.292)</td>
</tr>
<tr>
<td>FEP-2</td>
<td>2.92 (0.047)</td>
<td>2.95 (0.046)</td>
</tr>
<tr>
<td>SF-12 MH</td>
<td>30.90 (0.798)</td>
<td>31.18 (0.777)</td>
</tr>
<tr>
<td>SF-12 PH</td>
<td>47.08 (0.749)</td>
<td>47.04 (0.729)</td>
</tr>
</tbody>
</table>

Note. SE = standard error; PHQ-9 = Patient Health Questionnaire; FEP-2 = Questionnaire for the Evaluation of Psychotherapeutic Progress, SF-12 = Short-Form Health Survey; CI = 95% confidence interval.
Table 3. Predictors of therapy success and online usage time

<table>
<thead>
<tr>
<th>Predictor</th>
<th>PHQ-9 post score</th>
<th>Usage time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predator</td>
<td>Beta</td>
<td>p</td>
</tr>
<tr>
<td>Age</td>
<td>-.036</td>
<td>.668</td>
</tr>
<tr>
<td>Gender</td>
<td>.078</td>
<td>.353</td>
</tr>
<tr>
<td>Education</td>
<td>.124</td>
<td>.131</td>
</tr>
<tr>
<td>Relation status (in relation)</td>
<td>-.063</td>
<td>.434</td>
</tr>
<tr>
<td>PHQ-9 baseline score</td>
<td>.268</td>
<td>.002 **</td>
</tr>
<tr>
<td>MASKED [Attitudes towards Internet-based treatment]</td>
<td>-.049</td>
<td>.555</td>
</tr>
<tr>
<td>Medication</td>
<td>.103</td>
<td>.211</td>
</tr>
<tr>
<td>Combined treatment (CBT)</td>
<td>-.101</td>
<td>.279</td>
</tr>
<tr>
<td>Combined treatment (DP)</td>
<td>-.028</td>
<td>.732</td>
</tr>
<tr>
<td>Combined treatment (CP)</td>
<td>-.105</td>
<td>.280</td>
</tr>
<tr>
<td>No moderate usage (120 mins. or more)</td>
<td>-.159</td>
<td>.101</td>
</tr>
<tr>
<td>Usage (in minutes)</td>
<td>-.043</td>
<td>.901</td>
</tr>
<tr>
<td>Number of chats</td>
<td>.018</td>
<td>.750</td>
</tr>
</tbody>
</table>

Note. PHQ-9 = Patient Health Questionnaire; MASKED = Attitudes towards Internet-based treatment (MASKED); CBT = cognitive behavioural therapy; PDT = psychodynamic therapy; CP = client-centred or talk therapy; ** = significant after Bonferroni correction (p < .01).