The efficacy and effectiveness of online CBT
Ruwaard, J.-J.

Citation for published version (APA):
Ruwaard, J-J. (2013). The efficacy and effectiveness of online CBT

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2
Online Cognitive Behavioural Treatment of Work-Related Stress

Objective To assess the effects of a 7-week standardized Cognitive Behavioural Treatment (CBT) of work-related stress conducted via e-mail. Participants 342 people applied for treatment in reaction to a newspaper article. Initial screening reduced the sample to a heterogeneous (sub)clinical group of 239 of participants. Design participants were randomly assigned to a waiting list condition ($n = 62$), or to immediate treatment ($n = 177$). A follow-up was conducted three years after the inception of the treatment. Outcome Measures: The Depression Anxiety Stress Scales (DASS-42) and the Emotional Exhaustion scale of the Maslach Burnout Inventory - General Survey (MBI-GS). Results Fifty participants (21%) dropped out. Both groups showed statistically significant improvements. Intention-to-treat ANCOVA's revealed that participants in the treatment condition improved significantly more than the participants in the waiting control condition ($0.001 < P \leq 0.025$). In the treatment group, the effects were large to moderate ($0.9$ (stress) $\geq d \geq 0.5$ (anxiety)). The between-group effects ranged from $d = 0.6$ (stress) to $d = 0.1$ (anxiety). At follow-up, the effects were more pronounced, but this result requires replication in view of high attrition at follow-up. Conclusion The results warrant further research on Internet-driven standardized CBT for work-related stress. Such research should include the direct comparison of this treatment with face-to-face treatment, and should address the optimal level of therapist contact in Internet-driven treatment.
Chapter 2. Online CBT of Work-Related Stress

Introduction

Globalization, organizational restructuring, and the universal adoption of information technology have transformed the demands on employees in many organizations (Sparks, Faragher, & Cooper, 2001). The pressure to adapt to rapid technological and organizational demands has contributed to an increase in the prevalence of chronic stress among employees. Despite considerable efforts to counter this trend, the rates of work incapacity and sick leave due to work-related stress are high in the Netherlands compared with other European countries (Schaufeli & Kompier, 2001).

Stress is not necessarily harmful, i.e., a degree of stress is necessary to achieve desired levels of performance (van Doornen, 2001). However, stress may become a serious problem if people are exposed to high levels of stress over extended periods of time. Work-related stress is defined as excessive chronic stress caused by a mismatch between work-related demands and available coping skills (Health & Safety Executive, 2001). It is characterized by a variety of negative emotional and physiological reactions to aspects of the work itself, to the environment, and to the organization (Levi & Levi, 1995). Common psychological reactions to work-related stress are depression, anxiety, and emotional exhaustion (Cooper, Dewe, & O’Driscoll, 2001).

Cognitive Behavioural Therapy (CBT) and relaxation therapy are moderately effective in reducing work-related stress (van der Klink et al., 2001). In CBT, clients are instructed how to address the causes of their stress by changing their coping skills with regard to demands and pressures. In relaxation therapy, clients learn to lower their physiological arousal. In a meta-review, van der Klink et al. (2001) estimated the effects of CBT and relaxation therapy to be $d = .7$ and $d = .4$, respectively.

The implementation of these interventions, however, requires trained professionals, who are relatively few in number (World Health Organization, 2001). In addition, clients may be hesitant to seek help, due to limited mobility, travelling distance, or a lack of awareness concerning their condition, or the possibility of treatment. Finally, fear of stigmatization may discourage potential clients from seeking help. Given the obstacles to seeking face-to-face therapy, there is an increasing interest in methods to provide evidence-based treatment using the Internet (Barak, 1999; Emmelkamp, 2005; Newman, 2004). Online therapy may reduce the present burden on the available
mental health resources, and may provide a viable alternative for people, who face geographical, physical, psychological and/or financial barriers in seeking traditional, face-to-face care.

Critics of Internet-driven therapy have questioned its feasibility and have expressed ethical concerns (e.g. Bonger, 1988; Bauer, 2000). However, at present, the main question concerning online therapy is not whether, but rather how exactly it should be undertaken. Over the years, the accumulation of studies evaluating online therapy has resulted in a body of results. Reviews to date indicate that online therapy provides an effective alternative to existing methods for early intervention (e.g. K. M. Griffiths & Christensen, 2006; Wantland, Portillo, Holzemer, Slaughter, & McGhee, 2004). However, these reviews also show many challenging, open-ended issues that remain to be addressed. For example, the available evidence on the effectiveness of online therapy is limited to a relatively small number of disorders, and the effectiveness is found to vary considerably across studies (e.g. see Spek et al., 2007). Furthermore, dropout-rates tend to be high. Wantland et al. (2004) report an average dropout rate of 21%.

Studies vary greatly in the degree of therapist involvement. Some studies relied entirely on self-help material (e.g. Klein & Richards, 2001). Others explicitly involved a personal therapist, and included explicit structuring of the remote client-therapist interaction to facilitate the desired changes. The involvement of a therapist allows for personalization of the interventions and encourages the development of a therapeutic alliance (Knaevelsrud, Jager, & Maercker, 2004). In a meta-analysis of Internet-based interventions for anxiety and depression, Spek et al. (2007) found that the effects of interventions with therapist support were considerably larger than the effects of interventions without therapist support.

Zetterqvist, Maanmies, Ström, and Andersson (2003) tested the efficacy of e-mailed self-help for stress management with minimal therapist contact. The treatment involved relaxation training, development of problem-solving and time-management skills, cognitive restructuring, and behavioural exercise. Compared to a waiting list control group, participants in the programme displayed significantly greater improvement on the Perceived Stress Scale (PSS: S. Cohen, Kamarck, & Mermelstein, 1983) and the Hospital Anxiety and Depression Scale (HADS: Zigmond & Snaith, 1983).
The reported summary statistics suggest large to moderate (controlled) effect sizes of $d = .6$ (PSS), $d = .5$ (HADS Anxiety) and $d = .8$ (HADS Depression). Unfortunately, 26% of the participants dropped out and intention-to-treat analyses failed to reveal significant differences between the two groups.

Based on principles of online therapy developed earlier for the treatment of post-traumatic stress (Lange, Rietdijk, et al., 2003), we designed a standardized treatment of work-related stress, which can be conducted via e-mail. The treatment is based upon interventions that have proved their worth in clinical trials and clinical practice (de Jong & Emmelkamp, 2000; Lange, Richard, Gest, de Vries, & Lodder, 1998). The treatment involves psycho-education, awareness training, applied relaxation, cognitive restructuring, positive self-verbalization, social skills training, and time-management. The content of the treatment is similar to that of Zetterqvist et al. (2003). In contrast to the Zetterqvist et al. study, where client-therapist interaction was incidental and not standardized, the treatment manual includes explicit structuring of the client-therapist interaction. Specifically, it specifies the timing, nature, and frequency of the therapist instructions and feedback.

Below, we present the results of a randomized controlled trial that we conducted to assess the effects of our Internet-based standardized CBT on work-related stress. We compared the effects of treatment with those observed in a waiting list control group. We hypothesized that, in comparison to the waiting-list, the treatment would substantially reduce levels of stress. We also expected the treatment to reduce stress-related complaints, in particular depression, anxiety, and emotional exhaustion. In addition, we report the results of a three year follow-up study in which we assessed the long term effects among those who had completed treatment.

### 2.1 Treatment

The treatment comprised 7 phases, as described in detail below. Each phase required approximately 1 week to complete. The schedule allowed the participants to adjust the pace of completion within limits to their own situations. Treatment integrity was guaranteed by a manual that specified each step of the treatment in detail (e.g., the order and the nature of the assignments, the contents of the psycho-education, and
the timing of the therapist feedback). The manual also provided default text templates for feedback and instructions. The therapists tailored these texts to the specific needs of their clients. They were required to email their feedback after 1 working day. This provided the therapists with ample time to think through their feedback, or to discuss difficult cases with colleagues or a supervisor. The treatment manual specified 10 feedback-moments. Composing the feedback took about 30 minutes. Thus, a full treatment took about 5 hours of therapist time.

2.1.1 Treatment phases

Phase 1. Inducing awareness: monitoring and writing. This phase comprised one week of monitoring stress and two writing assignments to increase participants’ awareness of somatic, cognitive, emotional, and behavioural signs of stress. Participants kept records of stressful events: they described the situation, associated feelings and thoughts, and rated the degree of stress on a 10-point scale. Subsequently, they conveyed their ratings graphically and reflected on the results. In addition, the participants spend two periods of 45 minutes writing about stressful events, in the present tense and in as great a detail as possible (Lepore & Smyth, 2002; Alford, Malouff, & Osland, 2005).

Phase 2. Relaxation. First, participants were taught muscle relaxation or mental relaxation, depending on their own preference. Subsequently, they practiced 6 times a day. In addition, they were encouraged to take short breaks during their work to reduce their overall pace, and to engage in physical exercise. Participants who reported sleeping complaints received instructions how to regain healthy sleeping habits: i.e., not sleeping during the day, not sleeping too long during the night, avoiding stressful activities (e.g., stressful conversations) before going to bed, and not trying to force themselves to sleep (Morin, 1993).

Phase 3. Worrying, rumination & challenging dysfunctional thoughts. Participants were taught to recognize events that trigger worrying and ruminating. Participants were instructed to limit ruminating to predetermined, fixed moments, rather than indulging in it at all hours of the day (McKay, Davis, & Fanning, 1997). When they
started worrying and ruminating spontaneously, they were required to make a short note, and to determine the next fixed moment for ruminating. Furthermore they were required to focus on their notes and to challenge dysfunctional (black/white) thinking, overgeneralization, catastrophic thinking, self-blaming, and neglect of positive aspects (Beck, 2005). They did this three times a week for 20 minutes.

**Phase 4. Positive self-verbalization.** Participants were taught to focus on their positive qualities by using the technique of positive self-verbalization (Lange et al., 1998; Lange, Richard, Kiestra, & van Oostendorp, 1997). They wrote an essay on their positive traits and summarized this on a small card. The participants were encouraged to read the summary on the card frequently, and read it aloud.

**Phase 5. Positive assertiveness/social skills training and behavioural experiments.** Participants received a short social skills training. They were encouraged to reflect on the reciprocity of the interactions with their fellow-workers. Subsequently, they received instructions how to experiment with new strategies to improve negative reciprocal behaviours. They were, for instance, taught to communicate wishes in a positive and constructive manner, to give compliments, and to express their interest in the well-being of others. If they questioned the validity of these strategies, the therapist helped to create behavioural experiments, in which the new approaches could be put to test in a simple and realistic manner.

**Phase 6. Time management.** Participants were made aware how they managed the demands of both their work and their private life. In the evening, participants reflected on the activities of the day, after which they planned the activities for the next day. The participants learned how to set priorities, to reject requests that they found too demanding, to plan time for relaxation and self-reward, and to avoid unrealistic scheduling. They were encouraged to take their time to decide consciously whether to start an activity immediately, postpone it, or to pass it on to someone else.

**Phase 7. Future, reintegration and relapse prevention: the toolkit.** Participants reflected on symptoms that might signal relapse. They were encouraged to formulate their personal ‘relapse prevention toolkit’, using the techniques they had mastered in
2.2 Method

2.2.1 Design

The comparative study comprised a waiting-list controlled pre-post trial. Participants were randomly assigned to two groups. One group started the 7-week treatment immediately (experimental group), while the other started it after 7 weeks (waiting list control group). Three years after the completion of the pretest, a follow-up study was conducted among the clients of the experimental group who had completed the therapy to determine the long-term effects of the intervention.

2.2.2 Participants

Recruitment. A national Dutch quality newspaper published an interview with one of the authors about Internet-driven therapy. The article announced the current study, and referred interested readers to a website. This site provided psycho-education on work-related stress, explained the purpose and design of the study, and contained an application form.

Screening. Respondents were screened by means of self-report questionnaires with respect to the following exclusion criteria: age < 18 years, heightened risk of dissociation or psychosis, suicidal ideation, drug abuse, use of neuroleptic medication, and concurrent other treatment. In addition, respondents were required to download, print, and return a signed Informed Consent form. Excluded respondents were referred to other mental health institutions.

The participants completed a multiple-choice questionnaire to record age, gender, marital status, education, work contract, work years, working hours, duration of symptoms, sick leave status, and perceived causes of stress. In addition, alcohol dependency and drug abuse was registered. To assess the use of neuroleptics, participants answered open-ended questions concerning medicine brands and prescribed therapy. They were encouraged to print the toolkit, and place it in a visible place at home as a symbol of what they had mastered. Participants who were partially or fully on sick leave received suggestions for gradual reintegration.
doses. Risk of dissociation was assessed using the 5-item Somatoform Dissociation Questionnaire (SDQ-5; Nijenhuis, Spinhoven, van Dyck, van der Hart, & Vanderlinden, 1997). The internal consistency of this self-rate instrument is good (Cronbach’s $\alpha = .80$). A cut-off of 8 (used in this study) results in good discrimination between groups of patients and non-patients (sensitivity = .94; specificity = .96). Participants, who scored above the cut-off value of 13 on the Screening Device for Psychotic Disorder (SDPD; Lange, Schrieken, Blankers, van de Ven, & Slot, 2000 were excluded. The cut-off value of 13 was established in the Dutch norm group. The 7-item self-rate inventory (Cronbach’s $\alpha = .82$) is a good predictor of psychotic episodes. Agreement between self-report in a group of 33 schizophrenic patients and the ratings by their clinicians is high ($r = .85$) (Lange, Schrieken, et al., 2000). Suicidal ideation was measured by an inventory similar to that of Joiner et al. (2003). It comprises six multiple-choice questions, including “Do you currently have plans to end your life?” and “Are you currently feeling desperate?”. Participants were excluded if they had a history of suicide attempt(s) within the past two years, or if there was immediate risk of suicide.

**Randomization.** One month after the publication of the newspaper article, participants were randomly assigned to the two experimental groups by means of the random number generator procedure of SPSS 10. Assuming a correlation of $r = .5$ between a pretest and posttest measurement, a significance level of .05 would require 49 participants per group to provide an 80% probability to detect a medium ($d = .5$) effect size (J. Cohen, 1988).

**2.2.3 Procedure**

**Setting.** Communications took place entirely through e-mail, without any face-to-face session. Participants underwent the treatment at home. Incidentally, telephone contact was necessary (e.g., to inquire about the reasons for dropout when e-mails remained unanswered).

**Privacy.** Several procedures secured the privacy of the participants. An e-mail server was set up exclusively for the study and located at a professional Internet host. The
server, running Linux OS, was protected by a firewall and remotely administered through an encrypted communication channel. The therapists worked from a central computer-room that was accessible only with proper authorization. Each therapist received a private password-protected account to login to the computers and to the e-mail server. Participants were informed of mail-client extensions to secure their e-mail, and received help in configuring these extensions on request.

Materials. Participants needed a personal computer connected to the Internet and an e-mail account. For screening and outcome measurement, validated questionnaires were delivered in the body text of e-mails and in documents that were attached to these e-mails. The attachments were formatted in a common cross-platform computer format (Rich Text Format).

Therapists. The therapists were 25 doctoral students in clinical psychology and one postgraduate student, aged between 22 to 42 years old (M = 26, SD = 4.7). They had followed advanced courses in CBT. Additional training taught the therapists how to tailor the feedback and instruction templates of the manual to the needs of their clients, how to increase motivation by adopting a stimulating empathetic attitude, how to avoid the pitfalls of electronic, text-based communication (e.g. Brennan & Ohaeri, 1999), and how to make use of the asynchronous nature of the communication to enhance the quality of the feedback (e.g., by discussing cases with one another or the supervisor). Once a week, a senior specialist in Internet-driven CBT supervised the therapists. The supervisor assigned participants to therapists on the basis of the availability of the therapists.

Posttest and follow-up. Immediately after treatment, participants received the posttest measurement (the outcome questionnaires and an evaluation questionnaire). Those who did not complete treatment received a telephone call to inquire about the causes of dropout. Dropouts were also asked to complete the posttest measurements. Three years after the pretest measurements, participants, who had completed treatment, were invited by e-mail to complete the outcome questionnaires.
2.2.4 Measures

The primary outcome measure was the Stress subscale from the Depression Anxiety Stress Scales (DASS-42: Lovibond & Lovibond, 1995; Dutch version: de Beurs, van Dyck, Marquenie, Lange, & Blonk, 2001). Secondary measures were the Depression and Anxiety subscales from the DASS and the Emotional Exhaustion subscale of the Dutch version of the Maslach Burnout Inventory - General Survey (MBI-GS: Maslach, Jackson, & Leiter, 1996; Dutch version by Schaufeli & van Dierendonck, 2000).

**DASS.** The DASS is a self-report instrument that assesses depressive symptoms, physical anxiety (fear), and mental stress (nervous tension). It contains 42 items, 14 per subscale, that relate to the experience of symptoms in the past week. The items are measured on a 4-point scale ranging between 0 ("did not apply to me") to 3 ("applied to me very much, or most of the time"). Higher scores denote less favourable conditions.

All subscales of the Dutch adaptation are characterized by good internal consistencies (Cronbach’s $\alpha$ between .94 and .97, present sample: .86 to .92). Test-retest reliabilities for the Depression, Anxiety and Stress scale are $r = .75$, $r = .89$ and $r = .79$, respectively (de Beurs et al., 2001). Clinical cut-off scores of $c = 12$ and $c = 5$ for depression and anxiety are recommended by Nieuwenhuijsen, de Boer, Verbeek, Blonk, and van Dijk (2003). The corresponding percentile score of stress (14) observed in a large ($N = 1771$) non-clinical sample in the UK (Crawford & Henry, 2003) was used as the clinical cut-off in this study.

**MBI-GS Emotional Exhaustion.** The MBI-GS is a self-report questionnaire to assess burnout across professional occupations. The Emotional Exhaustion subscale assesses emotional fatigue, i.e., the feeling of being "worn-out". The scale contains five items that are measured on a 7-point scale scored from 0 to 6, where higher scores indicate higher levels of emotional exhaustion. The internal consistency of the subscale ranges between $0.84 \leq \alpha \leq 0.90$ (present sample: $\alpha = 0.84$). The 8-month test-retest reliability is satisfactory ($0.58 \leq r \leq 0.85$). Well-established clinical cut-off scores are unavailable for the Exhaustion scale. To indicate burnout, Brenninkmeijer and Van Yperen (2003) recommend a cut-off score of 2.67, which was used in the present study. With this
cut-off score, they found a false positive rate of 9.1% and a false negative rate of 13.8%.

2.2.5 Analysis

**Intention-to-treat.** The analyses were on an intention-to-treat basis and included all participants. Dropouts, who did not complete the posttest measurements, were assumed to have gained nothing. Their pretest scores also served as their posttest scores.

**Statistical significance.** One-way ANCOVA’s (using the pretest scores as a covariate) were conducted to test the difference in means of the two groups at posttest ($\alpha = .05$). The statistical significance of the within-group pre-post gain scores were assessed using paired $t$-tests.

The assumptions of ANCOVA were met. The homogeneity of the regression coefficients in the two groups was confirmed by non-significant interactions between the covariates (pretest scores) and experimental condition. The distributions of the outcome variables were approximately normal, and the variance across the experimental groups was found to be homogeneous.

**Effect size.** To express the magnitude of the effects, gain scores on the outcome measures were standardized to Cohen’s $d$’s (J. Cohen, 1988), representing the number of standard deviations separating the two means. Point estimates and 95% confidence intervals of $d$ were determined both for the within- and the between-group effects following a procedure described in detail by Robey (2004). In this procedure, between effect sizes are calculated using the pooled standard deviation and confidence intervals are approximated from the central $t$-distribution.

**Clinical relevance.** We tested the higher probability of statistically reliable improvement and recovery after treatment compared to the control group with Fisher exact tests, and expressed the difference in this probability between the treatment and the control group in terms of odds ratios (Hillis & Woolson, 2002). We used the Reliable Change Index (RCI) to test the significance of individual improvement (Jacobson & Truax, 1991; $\alpha = .05$; critical RCI = 1.96; “no change” and “detoriation” were pooled
Participants were considered recovered if they reliably improved from a pretest score above the cut-off to a posttest score below cut-off. Therefore, participants scoring below cut-off at pretest were excluded from the recovery analysis.

**Follow-up.** Participants who had completed the treatment were invited by e-mail to complete the follow-up measures on a specially constructed website. These participants include members of the treatment group and members of the control group, who followed the treatment at a later date. Pretest to follow-up data of those who participated in the follow-up were analyzed using repeated measures ANOVA’s with time of measurement as the within factor. Differences between the means at the times of measurement were tested for significance using Bonferroni adjustments to keep the family-wise Type I error at $\alpha = .05$.

2.3 Results

2.3.1 RCT

**Participants.** Of the 342 respondents who applied for treatment, 65 did not complete the screening and 38 met the exclusion criteria (See Figure 2.1 for details). The final sample ($N = 239$) comprised 143 females (60%) and 96 males (40%), aged between 22 and 60 ($M = 44$, $SD = 8$). Most of the participants were highly educated (84% completed tertiary education), and reported a wide variety of professional occupations. The majority (80%) were in full-time paid employment. Seven participants were unemployed, because of work-related stress. Ten participants suffered from stress in unpaid jobs (housewives, volunteers, and students). In 167 (70%) of the participants, the level of stress was above the clinical cut-off on the stress outcome measure (described in the measures section). On average, the duration of the reported symptoms was 30 months ($SD = 34$, Range $= 1 - 180$ months). A number of participants (38%) was on partial or full sick leave. All participants attributed their complaints to their job, although many (65%) reported additional (unspecified) personal issues.
2.3. Results

Randomization. Because we had a fairly large sample of participants, we assigned three times as many participants to the immediate treatment condition, while retaining a large enough sample-size in the control condition to ensure sufficient statistical power to detect treatment effects. Thus, 177 participants (74%) were randomly assigned to immediate treatment, and 62 (26%) to the waiting list control condition.

Figure 2.1: Flowchart of study participation.
Table 2.1: Characteristics of participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Control n = 62</th>
<th>Treatment n = 177</th>
<th>Combined N = 239</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender : female</td>
<td>68% 42</td>
<td>57% 101</td>
<td>60% 143</td>
</tr>
<tr>
<td>Age (M SD)</td>
<td>42 9</td>
<td>44 8</td>
<td>44 8</td>
</tr>
<tr>
<td>Education : tertiary</td>
<td>83% 50</td>
<td>85% 144</td>
<td>84% 194</td>
</tr>
<tr>
<td>Marital status : married</td>
<td>27% 17</td>
<td>18% 32</td>
<td>21% 49</td>
</tr>
<tr>
<td>living together</td>
<td>42% 26</td>
<td>54% 96</td>
<td>51% 122</td>
</tr>
<tr>
<td>unmarried</td>
<td>21% 13</td>
<td>21% 38</td>
<td>21% 51</td>
</tr>
<tr>
<td>divorced/widow(er)</td>
<td>10% 6</td>
<td>6% 11</td>
<td>7% 17</td>
</tr>
<tr>
<td>Work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment : fulltime</td>
<td>77% 48</td>
<td>81% 140</td>
<td>80% 188</td>
</tr>
<tr>
<td>Years in current profession (M SD)</td>
<td>9 11</td>
<td>9 10</td>
<td>9</td>
</tr>
<tr>
<td>Working hours per week (M SD)</td>
<td>35 13</td>
<td>36 13</td>
<td>36 13</td>
</tr>
<tr>
<td>Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of symptoms (months) (M SD)</td>
<td>26 29</td>
<td>32 36</td>
<td>30 34</td>
</tr>
<tr>
<td>Sickleave : no</td>
<td>56% 35</td>
<td>65% 114</td>
<td>63% 150</td>
</tr>
<tr>
<td>partial</td>
<td>23% 14</td>
<td>12% 21</td>
<td>15% 35</td>
</tr>
<tr>
<td>full</td>
<td>21% 13</td>
<td>23% 41</td>
<td>23% 54</td>
</tr>
</tbody>
</table>

Values represent subsample percentage and size unless otherwise noted.

Table 2.1 provides an overview of the characteristics of the two groups. To check the randomization, $\chi^2$ tests or $t$ tests (where appropriate) were conducted with respect to the outcome measures, gender, age, marital status, education, work contract, work years, working hours, duration of symptoms, and sick leave status. As expected, no significant difference was found [$t(238) = .26 - 1.75, P = .08 - .80; \chi^2 = .002 - 4.1, P = .13 - .96$].

**Dropout.** Fifty participants (21%) dropped out from the study, 49 from the treatment group and one from the control group. Of these 50, 25 (50%) dropped out without providing a reason (see Figure 2.1). Twenty-two dropouts (44%) completed the posttest. None of the variables used to check the randomization were predictive of dropout.

**Statistical significance.** Table 2.2 shows the results of the pretest and posttest measurements on intention-to-treat basis. As hypothesized, the treatment group improved statistically significant with respect to stress [$t(176) = 11.76, P < .001$]
2.3. Results

Table 2.2: RCT Results (Intention-to-Treat): Treatment (n = 177) vs. Waiting List Control (n = 62).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>Pre</th>
<th>Post</th>
<th>Effect size</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>d</td>
<td>CI₉⁵</td>
</tr>
<tr>
<td>DASS Stress</td>
<td>Treatment</td>
<td>19.4</td>
<td>8.0</td>
<td>.6 ± .3</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>19.7</td>
<td>8.0</td>
<td>17.6 ± 9.3</td>
<td></td>
</tr>
<tr>
<td>DASS Depression</td>
<td>Treatment</td>
<td>11.8</td>
<td>6.7</td>
<td>.4 ± .3</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>12.1</td>
<td>7.3</td>
<td>10.7 ± 7.3</td>
<td></td>
</tr>
<tr>
<td>DASS Anxiety</td>
<td>Treatment</td>
<td>8.2</td>
<td>6.2</td>
<td>.1 ± .3</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9.6</td>
<td>6.7</td>
<td>7.3 ± 5.5</td>
<td></td>
</tr>
<tr>
<td>MBI Em. Exh.</td>
<td>Treatment</td>
<td>3.1</td>
<td>1.2</td>
<td>.3 ± .2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.4</td>
<td>1.2</td>
<td>3.2 ± 1.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: RCT Results (Intention-to-Treat): Treatment (n = 177) vs. Waiting List Control (n = 62).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>Pre</th>
<th>Post</th>
<th>Effect size</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>d</td>
<td>CI₉⁵</td>
</tr>
<tr>
<td>DASS Stress</td>
<td>Treatment</td>
<td>19.4</td>
<td>8.0</td>
<td>.6 ± .3</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>19.7</td>
<td>8.0</td>
<td>17.6 ± 9.3</td>
<td></td>
</tr>
<tr>
<td>DASS Depression</td>
<td>Treatment</td>
<td>11.8</td>
<td>6.7</td>
<td>.4 ± .3</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>12.1</td>
<td>7.3</td>
<td>10.7 ± 7.3</td>
<td></td>
</tr>
<tr>
<td>DASS Anxiety</td>
<td>Treatment</td>
<td>8.2</td>
<td>6.2</td>
<td>.1 ± .3</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9.6</td>
<td>6.7</td>
<td>7.3 ± 5.5</td>
<td></td>
</tr>
<tr>
<td>MBI Em. Exh.</td>
<td>Treatment</td>
<td>3.1</td>
<td>1.2</td>
<td>.3 ± .2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.4</td>
<td>1.2</td>
<td>3.2 ± 1.2</td>
<td></td>
</tr>
</tbody>
</table>

aDASS: Depression Anxiety Stress Scales; MBI: Maslach Burnout Inventory, General Survey. DASS scores are sum scores, MBI scores are averages.

bThe means and standard deviations reflect the raw data, whereas the F-statistics reflect the results of the test of the difference in posttest means after pretest adjustments (ANCOVA).

cPretest scores were carried forward for participants, who did not complete the posttest measurement (n = 28).

dES: Effect size: Cohen’s d point estimate and 95% confidence interval.

and the secondary outcome measures (all paired t tests P < .0001). However, the improvement of the waiting list control group also reached statistical significance with respect to stress ([t(61) = 2.14, P < .036], and all secondary outcome measures, except DASS depression ([t(61) = 1.46, P = .15]). Nonetheless, the ANCOVAs showed that the treatment group improved significantly more than the control group, on both stress and the secondary measures (See Table 2.2).

Effect size. The within-group effects in the treated group were large to moderate, ranging between d = .9 (stress) and d = .5 (anxiety, exhaustion). Because the control group also improved, the effect sizes of treatment as compared to no treatment were smaller: the between effects ranged between d = .6 (stress) and d = .1 (anxiety) (see Table 2.2).

Clinical relevance. Table 2.3 shows the improvement and recovery rates in the two experimental groups. In the treatment group, with respect to stress, 53% of the participants reliably improved, and half the clinical subgroup recovered from
clinical stress. The odds of improvement and recovery were significantly higher in the treatment condition than in the waiting list. With respect to the secondary measures, the effects varied. Concerning depression, the effects were similar to the effects on stress. The effects were smaller, but statistically significant, with regard to emotional exhaustion. With respect to anxiety however, the differences were not significant. Concerning anxiety, there was substantial improvement and recovery in the control group that was not compensated by much additional improvement and recovery in the treatment group.

2.3.2 Long-term follow-up

Participants. After about three years (the mean number of months to follow-up was 34, \(SD = 2.1\)), participants who had completed treatment (\(N = 167\)) were invited to complete the follow-up questionnaires. Of these 167, 73 could not be traced. Of the remaining 94 participants, 63 (67%) completed the follow-up questionnaires.

A few differences were found between participants who completed the follow-up and those who did not. The former were more highly educated (\(\chi^2 = 4.9; P = \))
2.3. Results

Table 2.4: Results of the Three-Year Follow-up (n = 63).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Follow-up</th>
<th>Pre-FU ES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>DASS Stress</td>
<td>19.6</td>
<td>7.6</td>
<td>10.1</td>
<td>7.9</td>
</tr>
<tr>
<td>DASS Depression</td>
<td>11.6</td>
<td>6.5</td>
<td>5.6</td>
<td>5.8</td>
</tr>
<tr>
<td>DASS Anxiety</td>
<td>8.6</td>
<td>5.2</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td>MBI Em. Exhaustion</td>
<td>3.4</td>
<td>1.4</td>
<td>2.5</td>
<td>1.3</td>
</tr>
</tbody>
</table>

*DASS: Depression Anxiety Stress Scales; MBI: Maslach Burnout Inventory, General Survey.

.034). In addition, their treatment length was shorter compared to those who did not participate in the follow-up (t(117) = 3.3, P = .001). At pretest and posttest, the two groups did not differ on stress, anxiety or depression (t(164) = .27 - 1.0, P = .30 - .78). However, the mean pretest score on emotional exhaustion of participants completing the follow-up (M = 3.4) was higher than that of participants who did not (M = 3.0; t(164) = 2.47, P = .014). At posttest this difference was not statistically significant [t(157) = 1.19, P = .23].

Long-term Effects. Table 2.4 shows the means and variances of the mean scores of the participants at pretest, posttest, and the 3-year follow-up. There was no relapse: the effects became more pronounced. The 3-year follow-up means show significantly less pathology than the posttest means. In terms of effect size, the gains from pretest to follow-up were large [1.3 (anxiety) ≤ d ≤ 1.8 (stress)].

With respect to stress, the main effect of time was significant, F(2, 126) = 79.90, P < .0001. The pre-post improvement (P < .001) was maintained and more pronounced at the 3-year follow-up (pretest to follow-up: P < .001; posttest to follow-up: P = .020). The same pattern emerged with respect to depression [F(2, 126) = 60.73, P < .0001; posttest to follow-up: P= .002] and emotional exhaustion [F(2, 120) = 51.68, P < .0001; posttest to follow-up: P < .001]. However, with respect to anxiety, the effects remained constant [F(2, 126) = 49.52, P < .0001; posttest to follow-up: P = .34].

Effects of other treatments. Between the posttest and the 3-year follow-up, 41% (n = 26) of the participants received additional treatment (psychotherapy, other forms of
counselling, or medication). Those who followed additional therapy tended to display more complaints at follow-up compared to those who did not undergo additional treatment, although the difference was statistically significant only with respect to emotional exhaustion \( t(60) = 3.1, P < .001 \). At pretest and posttest, the groups did not differ on the outcome measures.

**Treatment satisfaction.** After completing treatment, participants rated the overall value of the treatment on a scale of 1 to 10 with an average of 7.6 \( (SD = 1.0; \) Range: 5 - 10). The different treatment phases were consistently rated favourably. Increasing awareness through writing received the highest mean rating \( (M = 8.1; SD = 1.4) \), relaxation was rated lowest \( (M = 7.1; SD = 1.8) \). The participants rated various aspects of the relationship with their therapists on three-point multiple-choice response scales. Overall, the relationship was rated as pleasant (88%) and personal (75%), and was perceived to have grown during treatment (57%). Sixty-eight percent indicated that they had not missed face-to-face contact.

**Mediating variables.** Comparison of the outcome in completers of the treatment group and participants of the control group who completed the treatment at a later date revealed no significant differences. Therefore, the treatment completers in both groups were pooled \( (N = 167) \) to obtain more power for the detection of significant predictors of outcome in multiple regression. However, no significant mediating variables were found. After controlling for pre-treatment stress-levels, post-treatment stress levels were not predicted by gender \( F(1,163) = .52, P = .47 \), age \( F(1,153) = .45, P = .45 \), education \( F(1,156) = .72, P = .40 \), working years \( F(1,157) = .004, P = .95 \), working hours \( F(1,157) = .12, P = .73 \), duration of symptoms \( F(1,110) = .10, P = .75 \) or length of treatment \( F(1,115) = .001, P = .97 \). Similar results were obtained with the secondary measures.

### 2.4 Discussion

Compared to no treatment, e-mailed standardized CBT moderately reduced stress and induced small to moderate improvements on depression and emotional exhaustion. Fifty percent of the participants had recovered from clinical stress following treatment.
The effects on anxiety appeared to be small. However, after three-years the (uncon-
trolled) effects were large on all outcome measures. Participants were highly satisfied
with their treatment and their therapists.

2.4.1 Evaluation of effect size

The substantial improvement of the participants in the waiting list control group con-
siderably deflated the effects between treatment and control group. The improvement
observed in the waiting-list control group is comparable to that observed in other
studies (Lange, van de Ven, & Schrieken, 2003; Zetterqvist et al., 2003). Zetterqvist
et al. (2003) provided several explanations for these improvements, including the
beneficial effects of pretesting and the prospect of treatment. In the current study,
the recruiting article and the website provided psycho-education that may have had
beneficial effects. Possibly, the participants in the control condition benefited from the
information they received, combined with the knowledge that they would start the
active treatment in the near future. However, these explanations remain speculative.
The use of different types of control groups may shed light on the source of the effects
in the control group.

Notwithstanding the improvement in the control group, the observed between
group effects are similar to those reported in the literature. The (controlled) effects on
stress ($d = .6$) and depression ($d = .4$) compare well with the effects of face-to-face
CBT ($d = .7$ and $d = .3$ respectively; van der Klink et al., 2001). In addition, the
results corroborate results obtained by Zetterqvist et al. (2003) in an RCT of an
Internet-driven stress-reduction program. In that study, the effect sizes were high,
but intention-to-treat analyses failed to reveal statistically significant between effects.
In contrast, all between effects in the present study were significant on an intention-
to-treat basis. This was perhaps due to the larger sample size and the use of a more
powerful statistical technique.

2.4.2 Maintenance

The uncontrolled three-year follow-up showed a high impact of the treatment with
large effect sizes after three years, on stress and depression, and on emotional exhaus-
tion and anxiety. This is consistent with our findings in the study on posttraumatic
stress (Lange, van de Ven, & Schrieken, 2003), in which effects also proved to be stable. We may conclude that participants, who complete this type of Internet-driven CBT, have a high chance to not relapse, and may even enjoy further improvements. These results are compelling, as the further improvements were not explained by any additional treatment, which some participants had sought in this three year period.

2.4.3 Limitations

Several aspects of this study limit the generalizability of the results. We will discuss the self-selection of the participants, the different time-frame between measurements in the experimental groups, and the dropout and attrition at follow-up.

Self-selection most clearly manifested itself in the high educational level of the participants. The high educational level is probably due to the fact that most participants were recruited through an article in a quality newspaper that is favoured by the well educated. Yet, high education seems to be rather typical of people who participate in Internet-based therapy (Andersson et al., 2006; Carlbring et al., 2001). Until the relation between education and Internet therapy is better understood, caution should be exercised in generalising the present results to less well-educated populations.

Participants in the treatment condition were posttested when they completed treatment. We expected them to complete treatment within the planned 7 weeks, but the average treatment took substantially longer (16 weeks). Consequently, the experimental group, on average, was posttested several weeks after the control condition. Therefore, the between-effects may have been confounded by (uncontrolled) effects of the mere passage of time. However, if present, this effect was probably small. Exploratory analyses did not reveal a significant effect of treatment duration on the outcome variables. Nevertheless, to avoid such interpretative issues in our current studies, we now multiply the protocol time by 1.5 to estimate the actual length of treatment.

During the study, a considerable number of participants dropped out. Although the observed dropout rate of 21% is common in online therapy (Wantland et al., 2004), it is disconcerting that most dropout occurred in the treatment group. No significant predictors of dropout were found, and many dropouts did not provide a reason for terminating their participation. However, therapists noticed that participants often
experienced problems in combining work and treatment demands. Perhaps some of the exercises required too much time for some participants or were otherwise too demanding. Thus, the treatment may have presented a practical burden in terms of time and effort that was not experienced by control group participants.

Finally, the observed long-term reduction of complaints is promising, but was uncontrolled, and suffered considerable attrition. The long-term follow-up needs replication.

2.4.4 E-mail versus website

In this study, the website was used only for basic psycho-education and recruitment. Communication between clients and therapists took place though e-mails. At present, the treatment is delivered completely through a structured, database-driven website. This has several advantages over e-mail exchanges. Privacy is ensured by transparent encryption techniques and login procedures. Further, therapists and clients now use a graphical interface to access treatment elements, which enhances access to the treatment dossier and process overview. Furthermore, the website provides real-time calculation of results and multimedia interfaces that support and facilitate the execution of the exercises. Preliminary (unpublished) findings suggest that the website-driven manual decreases the number of dropouts to a mere 15%, and that it enhances the improvement rates. It would be interesting to directly test the results of an e-mail versus website-driven treatment protocol for chronic stress.

2.4.5 Client-therapist relation and treatment integrity

Knaevelsrud and Maercker (2006) and Spek et al. (2007) observed that positive relationships between therapists and clients are quite feasible in Internet-driven therapy. This observation is supported by the positive evaluation of our participants of their therapists. Yet, this study does not permit conclusions regarding the effect of working alliance or the amount of therapist involvement in Internet-driven treatment. Given that some studies investigating Internet-based self-help with minimal therapist contact report reasonable results (e.g. Carlbring, Furmark, Steckzó, Ekselius, & Andersson, 2006), future research should focus explicitly on the role of the therapist:
How much therapist contact is optimal? What is the effect of working alliance in Internet-based treatment?

Internet-based treatment with standardized therapist contact may have distinct advantages over face-to-face treatment. In Internet-driven treatment it is easier to ensure treatment integrity than in manualized face-to-face treatment. In the present manual much attention is devoted to small details to ensure that the therapists use the manual in the intended way. The manual even specifies the motivating attitude that the therapists are required to adopt and convey to their clients. Furthermore, the weekly supervisions were intense, and exploited the full details of the exchanges between the therapists and their clients. The detailed manual provided the relatively young and inexperienced therapists with guidance and support in treating their clients. As a consequence, their results were on a par with the results of experienced therapists. Apart from the effects on clients, it is interesting to investigate the effects of Internet-driven manuals in the education and training of therapists.
2.4. Discussion