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 4
Online Cognitive Behavioural Treatment of Panic Symptoms

Background Internet-delivered treatment may reduce barriers to care in those unwilling or unable to access traditional forms of treatment. Objective To assesses the efficacy of web-based therapist-assisted cognitive behavioral treatment (web-CBT) of panic symptoms. Design A randomized waiting-list controlled trial with an uncontrolled three-year follow-up. Participants A community sample of 58 participants with chronic panic symptoms of varying severity (immediate treatment: \( n = 27 \), waiting-list control: \( n = 31 \)). Outcome Measures The primary outcome measures were a one-week Panic Diary and the Panic Disorder Severity Scale Self-Report (PDSS-SR); secondary measures were the Agoraphobic Cognitions Questionnaire (ACQ), the Body Sensations Questionnaire (BSQ), the Mobility Inventory - Alone subscale (MI-AAL), and the Depression Anxiety Stress Scales (DASS-42). Results In the RCT, 54 participants (93%) completed posttest measurements. With regard to the primary outcome measures, intention-to-treat ANCOVA's revealed that participants in the treatment condition improved more than the participants in the waiting-list control condition \( (P < .03) \), with a pooled between-group effect size of \( d = .7 \). After three years \( (n = 47; 81\% \text{ study compliance}) \), effects were more pronounced. Conclusion The results demonstrate the efficacy of therapist-assisted Web-CBT in the treatment of panic symptoms.
Chapter 4. Online CBT of Panic Symptoms

Introduction

Panic Disorder With and Without Agoraphobia (PD/A) is a debilitating condition characterized by recurrent, unexpected panic attacks accompanied by persistent concern about future attacks and possible avoidance of situations and places in which attacks are expected to occur. With pharmacotherapy or psychotherapy, especially cognitive behaviour therapy (CBT), prospects of recovery are good (Roth & Fonagy, 2005). However, too few of those affected actually receive treatment (Collins et al., 2004; W. H. O. World Mental Health Survey Consortium, 2004). Given the chronic nature of untreated PD/A and its associated individual suffering and costs to society, it is important to find ways to increase access to treatment.

Internet-delivered treatment is in general more accessible than traditional forms of treatment. In addition, randomised controlled trials (RCTs) have demonstrated the feasibility of Web-based CBT in the treatment of mood and anxiety disorders, with moderate to large effect sizes found with programs targeting PD/A (Andersson, Cuijpers, Carlbring, & Lindefors, 2007; Barak et al., 2008; Reger & Gahm, 2009). At present, a variety of web-based treatments targeting PD/A are available, from fully self-administered therapy (Farvolden, Denisoff, Selby, Bagby, & Rudy, 2005), to self-help with minimal guidance (Marks et al., 2004) and guided self-help with limited therapist support via e-mail and/or telephone (Carlbring et al., 2006; Klein et al., 2006; Shandley et al., 2008).

In web-based CBT, therapist involvement is often minimized to increase availability and reduce costs. In a review of Palmqvist et al. (2007), the mean time spent per PD/A patient is 90 minutes (range 0-6 h). This is considerably less than the time spent in the 12-15 (45 min) sessions commonly provided in face-to-face CBT (Roth & Fonagy, 2005). Although most Web-based treatments are promising in terms of established effect sizes, the effect of therapist involvement on the outcome requires further study. For instance, dropout was huge in the Farvolden et al. (2005) study of a program without therapist involvement, with as few as 12 out of 1161 participants completing the program. As reduced contact has been shown to affect outcome negatively in face-to-face CBT (Sharp, Power, & Swanson, 2000), it is possible that this is also the case in web-based treatments. Meta-reviews of Internet programs
suggest that therapist involvement reduces dropout and that greater therapist input improves outcome (Palmqvist et al., 2007; Spek et al., 2007). At present, the most extensive web-based treatment of PD/A includes 6 h of therapist involvement. The effect of increased therapist involvement on outcome of web-based CBT remains to be determined.

On the basis of principles of online therapy, which we applied in the web-based treatment of posttraumatic stress (Lange, Rietdijk, et al., 2003), work-related stress (Ruwaard et al., 2007), and depression (Ruwaard et al., 2009), we developed a therapist-guided web-based treatment of panic symptoms. As in our other studies, the primary aim was to maximize adherence and outcome, rather than to reduce therapist time. Thus, our internet program involved more therapist time (about 5-9 h) than existing programs targeting PD/A.

Earlier, Jager et al. (2004) presented a case study of a preliminary version of this treatment, in which interventions were delivered by e-mail. This treatment was found to induce large reductions in panic symptoms. Here, we present the adapted web-based version of this treatment, and the results of a RCT in which we compared the effects of this treatment with those of a waiting-list. Prevalence rates of minor forms of panic disorder are known to be high and individuals with minor forms of panic disorder suffer a similar impact on their quality of life and functioning as those with full blown panic disorder (Batelaan, de Graaf, Van Balkom, Vollebergh, & Beekman, 2006; Kessler et al., 2006; Magruder & Calderone, 2000; W. H. O. World Mental Health Survey Consortium, 2004). Therefore, to increase the external validity of the study, we aimed to ascertain a community sample with varying intensities of panic symptoms. In addition, we assessed outcome three years after the start of treatment. We hypothesised that, in comparison to the waiting-list, the treatment would reduce panic symptoms more, and that the improvements would be sustained over time.

4.1 Treatment

The treatment involves common CBT strategies for panic disorder, such as psychoeducation, awareness training, applied relaxation, cognitive restructuring and (interoceptive) exposure techniques. Similar to manualized face-to-face CBT, the treatment
Chapter 4. Online CBT of Panic Symptoms

comprises homework assignments and scheduled therapeutic sessions, in which assignments are explained and tailored to the needs of the client. In web-based CBT, the homework assignments are based on a web-based personal interactive workbook. At specific occasions indicated in the manual, therapists post feedback and further instructions on the basis of the contents of this workbook. Therapists take about 20-40 min to read a client's assignment, and to prepare feedback. The manual includes 14 of these feedback moments, so that a full treatment requires between 5 and 9 h of therapist time. Treatment integrity is guaranteed by a computerized manual that stipulates each step of treatment, including the order, the nature, and the contents of the assignments, and the timing of therapist feedback. Furthermore, the manual provides feedback templates, which the therapists adapt to the needs of their clients. These templates include suggested courses of action given various scenarios, such as problems in completing a given assignment.

The approximate duration of treatment is 11 weeks, in which clients work through seven treatment modules. The first module focuses on awareness of panic symptoms. Panic attacks are explained as the result of a cyclic process in which bodily reactions are misinterpreted and amplified (Clark, 1986). Next, clients are asked to do two writing assignments, in which they describe two past panic attacks in the light of this explanation. In the second module, clients learn to keep a Panic Diary, in which they describe, following each panic attack, the situation, their accompanying thoughts and bodily sensations, and in which they rate the intensity of fear on a 10-point scale. Clients maintain this diary throughout treatment. Next, in module 3, clients are taught a breathing retraining exercise and a progressive relaxation exercise with tension-release of the muscles. They are then encouraged to expose themselves to their feared situations. Here, the therapists add the paradoxical suggestion (Lange, 2006, chapter 10) that the occurrence of a panic attack would provide the opportunity to put the relaxation exercise into practice to counter the attack. Next, in the fourth module, clients are educated about different forms of exposure (in vitro/in vivo/interoceptive) and instructed to seek out the situations in which they expect to suffer a panic attack. The clients discuss their planned exposure assignment and their expectations with the therapist. Following the exposure assignment, clients report the outcome and the therapist provide feedback. In module 5, clients learn about the role
4.2 Method

4.2.1 Design

To assess the efficacy of the treatment, we ran a randomised waiting-list controlled pre-post trial, which was approved by the Ethics Committee of the Department of Psychology of the University of Amsterdam. The participants were randomly assigned to two groups. One group started the 11-week treatment immediately (experimental group), while the other started after 11 weeks (waiting-list control group). Three years after the start of the trial, all randomised participants were invited to complete follow-up measurements.

4.2.2 Participants

Enrollment. Radio broadcasts, newspaper articles, advertisements, and magazines interviews announced the study and referred to a website for additional information. This website provided background information on PD/A and on the purpose and design of the study. In addition, the site contained an application form. Power
analyses showed that approximately 70 participants (35 per group) were needed to detect a large, $d = .8$, between-group effect (with ANCOVA, an estimated pre-post correlation of .5 and Holm-Bonferroni corrections for comparisons on 10 outcome measures).

**Screening.** Respondents were screened through web-administered self-report questionnaires and a semi-structured 15-minute telephonic interview. Participants first completed the self-report questionnaires. Next, the results were used to prepare the clinical interview. In this interview, the presence of at least subsyndromal PD/A was established, according to the guidelines as listed in Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2000), i.e., respondents were included only if they experienced at least one full or limited symptom panic attack in the previous month. Diagnoses were made by trained interviewers and checked by supervising researchers.

The following inclusion criteria applied: age 18 years or older; no heightened risk of dissociation, psychosis, or suicide; no history of serious mental illness; no other prevailing mental illnesses; absence of a medical condition that might explain panic symptoms; no drug abuse; no use of neuroleptic medication; no use of anxiolytics or antidepressants for less than 3 months or use of unstable doses (i.e., clients took medications as prescribed and in the targeted dosage); and no concurrent other treatment. Excluded respondents were referred to their general practitioner, to other mental health institutions or to other Web-CBT programs. Eligible respondents downloaded, signed and returned an Informed Consent form.

Risk of dissociation was assessed using the Somatoform Dissociation Questionnaire (SDQ-5; citeNPNijenhuis1997). Respondents who scored above the cut-off (8), completed the more specific Dissociation Questionnaire (DIS-Q) (Vanderlinden, Van Dyck, Vandereycken, & Vertommen, 1991; cut-off: 3.0). The Screening Device for Psychotic Disorder (SDPD; Lange, Schrieken, et al., 2000) was used to assess the risk of psychotic episodes, with a cut-off value of 5 on the Hallucination scale. Hypochondriasis was assessed using the Whitely Index (Speckens, Spinhoven, Sloekers, Bolk, & van Hemert, 1996; cut-off: 15). The occurrence of a prevailing posttraumatic stress disorder was assessed using the Dutch version of the Impact of Events Scale - Revised (IES-R; Weiss & Marmar, 1996; cut-off 36: Neal et al., 1994). Obsessive compulsive
4.2. Method

disorder, social phobia and specific phobia, and bipolar disorder were signalled using seven items from the Diagnostic Interview Schedule (DIS: Helzer & Robins, 1988).

Therapists. The therapists were 7 graduate students in clinical psychology, 2 postgraduate students, and 2 psychologists. They were supervised by two senior specialists in web-CBT. All therapists had followed advanced courses in CBT, and received additional training in administering web-CBT. They were taught how to use the feedback templates of the manual, to increase motivation by adopting a stimulating emphatic attitude, to avoid the pitfalls of electronic, text-based communication (e.g., Brennan & Ohaeri, 1999) and to profit from the asynchronous nature of the communication to enhance the quality of the feedback (i.e. by discussing cases with one another or with the supervisor). Participants were assigned to therapists by the supervisor based on the availability of the therapists.

4.2.3 Outcome measures

Primary outcome measures were the self-rate version of the Panic Disorder Severity Scale (PDSS-SR; Houck et al., 2002; Shear et al., 2001) and a one-week Panic Diary. Secondary measures were the Avoidance when ALone subscale of the Mobility Inventory (MI-AAL; Chambless, Caputo, Bright, & Gallager, 1984), the Agoraphobic Cognitions Questionnaire and the Body Sensations Questionnaire (ACQ/BSQ; Chambless et al., 1984) and the Depression Anxiety Stress Scales (DASS-42; Lovibond & Lovibond, 1995). For all measures, higher scores indicate greater symptom severity.

Panic diary. In the Panic Diary, participants monitored panic attacks occurring in a one-week period. Participants were instructed to report each distinct period that was characterized by a sudden onset of intense apprehension, fearfulness, or terror, possibly associated with feelings of impending doom (American Psychiatric Association, 2000). They were asked to note the occurrence of each attack and to rate the attack severity (on a 1-10 scale) and the specific symptoms experienced during the attack. To this end, participants checked each of 13 key symptoms of panic disorder occurring during the attack, e.g., “palpitations, pounding heart or accelerated heart rate”, “trembling or sweating”. Three outcome measures were derived from the
diaries: 1) one-week attack frequency, 2) average number of experienced symptoms and 3) average attack intensity. For the three year follow-up, the burden of keeping a one-week diary was expected to have a negative impact on the response rate. Therefore, in the follow-up, we asked participants to retrospectively estimate the frequency, the average number of symptoms, and the average intensity of panic attacks in the week prior to follow-up assessment.

**PDSS-SR.** This 7-item survey is the self-report version of the more commonly used clinician-rated Panic Disorder Severity Scale (Shear et al., 2001). It assesses the severity, in the past week, of the following seven key dimensions of panic disorder, i.e., frequency of panic attacks, distress during panic attacks, anticipatory anxiety, agoraphobic fear and avoidance, body sensation fear and avoidance, and work and social impairment. The responses to the items are expressed on a 5-point Likert scale ranging from 0 (none) to 4 (extreme), and summed to create an overall severity score (range 0-28). Psychometric studies revealed excellent internal consistency (Cronbach’s α: > 9; Houck et al., 2002; Newman, Holmes, Zuellig, Kachin, & Behar, 2006), good two-week test-retest reliability ($r = .84$; Newman et al., 2006), and sensitivity to change after treatment (Houck et al., 2002). With a cut-off of 8, the clinician-rated PDSS is moderately effective in identifying patients with Panic Disorder (sensitivity: 83%; specificity: 64%; Shear et al, 2001). van der Meer and Burgerhout (2004) confirmed the usefulness of this cut-off in a psychometric study of the Dutch self-report version of the PDSS. In their study, a PDSS-SR cut-off of 8 discriminated between patients ($n = 129$) and non-patients ($n = 131$) with specificity of 100% and a sensitivity of 82%.

**MI-AAL.** This is a 27-item self-report questionnaire measuring agoraphobic avoidance in a variety of places and situations while alone. Participants rate the frequency of avoidance of these situations on a 5-point Likert scale (from 1, never avoid, to 5, always avoid). Item scores are averaged to obtain an overall measure of avoidance severity. In Dutch psychometric studies (de Beurs, 1993), the MI-AAL was shown to be reliable ($\alpha = .94$), valid and sensitive to change with treatment. Respondents scoring 2.18 or up are more likely to belong to the clinical population (de Beurs, 1993).
4.2. Method

**ACQ/BSQ.** Fear of fear was measured using the Body Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire twin-scales (ACQ/BSQ: Chambless et al., 1984; Dutch version: T. K. Bouman, 1995, 1998; de Beurs, 1993). The BSQ is a measure of the intensity of anxiety/fear provoked by 17 physical sensations. The ACQ measures the frequency of 14 beliefs about the negative consequences of anxiety. For both measures, items are rated on a 5-point Likert score ranging from 0 to 4, and item scores are averaged to obtain the overall score. Both measures have good internal consistency ($\alpha > .89$), and good test-retest reliability ($>.78$). To reduce the response burden, the ACQ and BSQ were administered at baseline and posttest, but not at follow-up.

**DASS.** The DASS measures negative affect by assessing the severity of symptoms of depression (DASS DEP), anxiety (DASS ANX), and mental stress (DASS STR). It comprises 42 items, 14 per subscale, that relate to the experience of symptoms in the past week. The items are rated on a 4-point Likert scale ranging between 0 (did not apply to me) to 3 (applied to me very much, or most of the time). All subscales of the Dutch adaptation are characterized by good internal consistencies (Cronbach’s $\alpha$ between .94 and .97), and satisfactory 1-month test-retest reliabilities (Depression: $r = .75$, Anxiety: $r = .89$; Stress: $r = .79$; de Beurs et al., 2001; Nieuwenhuijse et al., 2003).

**Impairment.** During treatment, after each treatment phase, we asked participants to rate their past-week impairment caused by panic symptoms, to gain insight in the development of panic symptoms during the course of treatment. Participants rated the impact of panic symptoms on their daily functioning on a single 10-point item ranging from 1 (no impairment) to 10 (severe impairment).

4.2.4 Analyses

**Intention-to-treat.** The RCT analysis was conducted on an intention-to-treat basis and included all participants. Participants failing to complete posttest measurements were assumed to have gained nothing. Their pretest scores served as their posttest scores. No attempt was made to correct for missing data of the three-year follow-up,
because statistical imputation was considered inappropriate with such a long time-interval. However, in the analyses of the long-term outcome data we used mixed modeling (see below for details), which is an accepted method to account for missing data.

**Statistical significance.** Two-tailed ANCOVAs (using pretest scores as a covariate) were conducted to test the difference in means of the two groups at posttest, using Holm-Bonferroni adjustments (Holland & DiPonzio Copenhaver, 1988) to maintain overall Type-1 error $\alpha$ at .05. These analyses were run using the generalized linear model function (glm) of the statistical software program ‘R’ (R Development Core Team, 2008).

The assumptions of ANCOVA were examined and found to be satisfied. The distribution of most outcome variables was approximately normal, and the variance across the groups was homogeneous. With regard to DASS Depression, normality was achieved by means of a square root transformation. Further, the distribution of attack frequency was positively skewed, as was to be expected since these are count data. For this variable, a generalized linear model with a Quasi-Poisson distribution (see Maindonald & Braun, 2007) as the link function provided a more realistic ANCOVA regression model. The homogeneity of the regression coefficients in the two groups was confirmed by non-significant interactions between the covariates (pretest scores) and experimental condition. However, a significant group by covariate interaction was found with respect to the BSQ (the effects were more pronounced for higher baseline BSQ scores). As it would be improper to use the significance of the treatment factor as an indicator of effect in this case (Enqvist, 2005), we used the significance of the group by covariate interaction term as the outcome of interest.

**Effect size.** To express the magnitude of the effects, mean gain scores on the outcome measures were standardized to Cohen’s $d$ (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). Between effect sizes were calculated using the pooled standard deviation (of the pretest scores) and confidence intervals were approximated from the central $t$-distribution.
Clinical relevance. We tested the differential probability of a clinically relevant outcome on the primary outcome measures after treatment compared to the control group with two-sided Fisher’s exact tests ($\alpha = .05$), and expressed this difference as odds ratios (OR; Hillis & Woolson, 2002). With regard to attack frequency, two clinically relevant outcomes were defined as follows: (a) a reduction in panic attack frequency of at least 50%, and (b) no full-blown panic attack at posttest, i.e. no panic attack with four or more of the key symptoms listed in DSM-IV (American Psychiatric Association, 2000). With regard to the PDSS-SR, a posttest score below the cut-off of 8 was considered clinically relevant. Participants scoring below cut-off at pretest had to remain below cut-off, and participants scoring above cut-off at pretest had to reliably improve to a score below cut-off. To account for measurement error, we used the Reliable Change Index (RCI; Jacobson & Truax, 1991) to test the significance of individual improvement. Participants had to improve at least 5 scale points before change was considered reliable.

Pooled Outcome and long-term follow-up. After the waiting period, the control group followed the web-based treatment too. To increase power of the long-term follow-up analyses, the data of the treatment and the control group were pooled, using the second (post-waiting period) assessment of the control group as the pretest. Pre-treatment to follow-up outcome data were analyzed through multilevel regression modeling (see Pinheiro & Bates, 2000) with time of measurement (at level 1) nested within participants (at level 2), and a random intercept and time-coefficient to account for individual differences between participants. Differences between the means at the times of measurement were tested for significance using simple contrasts. For these analyses, Holm-Bonferroni adjustments (Holland & DiPonzio Copenhaver, 1988) were used to maintain a familywise type I error of $\alpha = .05$ within each measure.
Applied  
\( n = 164 \)

Withdrew  
\( n = 69 \)
- 21 did not start
- 30 incomplete screening
- 18 no interview

Pre-test  
\( n = 95 \)

Excluded  
\( n = 37 \)
- 15 prevailing disorders
- 3 somatic problems
- 3 hospitalization
- 3 suicidal ideation
- 3 dissociation
- 3 alcohol dependence
- 3 no informed consent
- 2 unstable medication
- 2 concurrent treatment

Randomised (1:1)  
\( n = 58 \)

Waiting-list Control  
\( n = 31 \)
- 13-week Follow-up  
  \( n = 30 \)
  - 1 uncontactable
- Analyzed  
  \( n = 31 \)
- Web-based CBT  
  \( n = 27 \)
  - 24 completed
  - 3 discontinued
- Post-treatment Follow-up  
  \( n = 24 \)
  - 3 unmotivated
- Analyzed  
  \( n = 27 \)

Web-based CBT  
\( n = 27 \)
- Post-treatment Follow-up  
  \( n = 24 \)
  - 3 unmotivated
- Analyzed  
  \( n = 27 \)

Three Year Follow-up  
\( n = 31 \)
- 23 completed
- 5 could not be traced
- 3 refused

Three Year Follow-up  
\( n = 27 \)
- 24 completed
- 2 could not be traced
- 1 refused

Figure 4.1: Participant flow.
4.3 Results

4.3.1 RCT

**Enrollment and randomisation.** Recruitment resulted in 164 respondents who applied for treatment. Of these, 69 (42%) did not complete the screening. Of the remaining 95 respondents, 37 (39%) met the exclusion criteria (cf. Figure 4.1), and 58 were randomly assigned either to immediate treatment \((n = 27; 47\%)\) or to the waiting-list \((n = 31; 53\%)\). With these group sizes, the power of the study to detect a large \((d = .8)\) between-group effect was approximately 71% for the strongest effect. To check the randomization, \(t\)-tests and \(\chi^2\) tests were conducted with respect to the outcome measures, gender, age, marital status, education, duration of symptoms, and medication status. As implied by Table 4.1, no significant differences were found. We concluded that the randomisation procedure had been successful.

<table>
<thead>
<tr>
<th>Table 4.1: Characteristics of Participants.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Demographic</strong></td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Age ((M, SD))</td>
</tr>
<tr>
<td>Education: tertiary</td>
</tr>
<tr>
<td>With partner</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
</tr>
<tr>
<td>Years with symptoms ((M, SD))</td>
</tr>
<tr>
<td>PDSS-SR &gt; 8(^b)</td>
</tr>
<tr>
<td>MI-AAL &gt; 2.18(^c)</td>
</tr>
<tr>
<td>Panic Diary: at least 1 full-blown attack(^d)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
</tr>
<tr>
<td>Previous treatment for panic</td>
</tr>
<tr>
<td>Previous treatment for other disorder</td>
</tr>
<tr>
<td>Medication</td>
</tr>
</tbody>
</table>

\(^a\)Values represent subsample percentage and size unless otherwise noted.

\(^b\)PDSS-SR: Panic Disorder Severity Scale - Self-Report.

\(^c\)MI-AAL: Mobile Inventory - Avoidance Alone subscale.

\(^d\)A full-blown attack is an attack with four or more core symptoms of panic disorder as listed in the DSM-IV (American Psychiatric Association, 2000)
Baseline characteristics. As shown by Table 4.1, the participants, on average, were female (72%), middle-aged ($M = 38$, $SD = 10$, range: 20-69) and highly educated (47% completed tertiary education). Average duration of panic symptoms was 9 years ($SD = 8$, Range = 1 - 144 months). At pretest, symptoms levels were typical of clinical groups in the majority (52%-68%) of participants. About a third (36%, $n = 21$) of the participants had engaged in previous treatment for panic disorder, and 29% ($n = 17$) were on a stable dose of either an anxiolytic or an antidepressant.

Compliance. Trial attrition was low: 93% of the participants ($n = 54$) completed the posttest measures. Four participants dropped out: three in the treatment group and one in the control group (c.f. Figure 4.1). Given the small number of dropouts, predictors of dropout were not subject to analysis.

Statistical significance and effect size. Table 4.2 shows the results of the pretest and posttest measurements on intention-to-treat basis. With regard to the primary outcome measures, the treatment group improved significantly more than the control group ($P < .027$), with a pooled standardized mean difference in improvement between the two groups of $d = .7$. In addition, treated participants experienced significant larger reductions in fear-provoking bodily sensations (BSQ, $P < .025$) and general psychopathology (DASS) in comparison to untreated participants. However, between-group effects were moderate to small ($d < .3$), and non-significant ($P > .056$) with regard to avoidance (MI-AAL) and agoraphobic beliefs (ACQ).
4.3. Results

Table 4.2: RCT results (intention-to-treat\(^a\)): web-CBT (\(n = 27\)) vs. waiting-list control (\(n = 31\)).

<table>
<thead>
<tr>
<th>Measure(^b)</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Between ES(^c)</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>One-week attack frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>4.7</td>
<td>3.5</td>
<td>2.7</td>
<td>3.1</td>
</tr>
<tr>
<td>Control</td>
<td>3.9</td>
<td>2.6</td>
<td>3.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Symptom count (range: 1-13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>3.2</td>
<td>1.5</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Control</td>
<td>3.0</td>
<td>1.8</td>
<td>3.0</td>
<td>1.9</td>
</tr>
<tr>
<td>Attack intensity (range: 1-10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>4.5</td>
<td>2.2</td>
<td>2.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Control</td>
<td>4.1</td>
<td>2.0</td>
<td>4.2</td>
<td>2.3</td>
</tr>
<tr>
<td>PDSS-SR (range: 0-28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>9.0</td>
<td>5.5</td>
<td>5.9</td>
<td>4.4</td>
</tr>
<tr>
<td>Control</td>
<td>9.4</td>
<td>5.4</td>
<td>8.4</td>
<td>4.6</td>
</tr>
<tr>
<td>MI-AAL (range: 1-5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2.6</td>
<td>1.0</td>
<td>2.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Control</td>
<td>2.5</td>
<td>1.0</td>
<td>2.4</td>
<td>.9</td>
</tr>
<tr>
<td>BSQ (range: 1-5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2.3</td>
<td>.6</td>
<td>1.9</td>
<td>.5</td>
</tr>
<tr>
<td>Control</td>
<td>2.5</td>
<td>.7</td>
<td>2.4</td>
<td>.8</td>
</tr>
<tr>
<td>ACQ (range: 1-5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1.1</td>
<td>.6</td>
<td>.8</td>
<td>.5</td>
</tr>
<tr>
<td>Control</td>
<td>1.0</td>
<td>.6</td>
<td>.9</td>
<td>.6</td>
</tr>
<tr>
<td>DASS Dep (range: 1-42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>9.5</td>
<td>8.0</td>
<td>3.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Control</td>
<td>8.6</td>
<td>8.3</td>
<td>8.3</td>
<td>8.3</td>
</tr>
<tr>
<td>DASS Anx (range: 1-42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>11.8</td>
<td>6.7</td>
<td>6.7</td>
<td>4.9</td>
</tr>
<tr>
<td>Control</td>
<td>14.2</td>
<td>8.2</td>
<td>12.9</td>
<td>9.2</td>
</tr>
<tr>
<td>DASS Str (range: 1-42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14.2</td>
<td>7.8</td>
<td>8.5</td>
<td>4.7</td>
</tr>
<tr>
<td>Control</td>
<td>12.6</td>
<td>8.1</td>
<td>12.9</td>
<td>9.3</td>
</tr>
</tbody>
</table>

\(^a\)All randomized participants were included in the analyses. Pretest scores of dropouts were carried forward to the posttest.

\(^b\)PDSS-SR: Panic Disorder Severity Scale - Self-report; MI Alone: Mobile Inventory - Alone subscale; ACQ: Agoraphobic Cognitions Questionnaire; BSQ: Body Sensations Questionnaire; DASS: Depression Anxiety Stress Scales (Dep: Depression; Anx: Anxiety; Str: Stress). Higher scores indicate less favourable conditions.

\(^c\)ES: effect size: Cohen’s \(d\) point estimate and 95% confidence interval (\(d - CI_{95}\) to \(d + CI_{95}\)).

\(^d\)\(P\)-values were Holm-Bonferroni corrected for multiple testing.
Table 4.3: Web-CBT vs. waiting-list control: clinical relevance analysis (intention-to-treat)a.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Criterion</th>
<th>Criterion met %b</th>
<th>ORc</th>
<th>Pd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attack Frequency</td>
<td>-50%</td>
<td>52%</td>
<td>14%</td>
<td>6.3</td>
</tr>
<tr>
<td>Full-Blown Attacks</td>
<td>0</td>
<td>70%</td>
<td>47%</td>
<td>2.7</td>
</tr>
<tr>
<td>PDSS-SR</td>
<td>&lt;8</td>
<td>70%</td>
<td>42%</td>
<td>3.2</td>
</tr>
</tbody>
</table>

aPretest scores of dropouts were carried forward to the posttest.
bPDSS-SR: Panic Disorder Severity Scale. Cut-off: < 8.; A full-blown attack is an attack with four or more core symptoms of panic disorder as listed in the DSM-IV (American Psychiatric Association, 2000)
cOdds ratio (OR): the ratio of the odds of a clinically relevant outcome in the treatment group and the odds in the control group.
dListed P-values were Holm-Bonferroni adjusted for multiple testing.

Clinical relevance. Compared to the participants in the waiting list, treated participants were six times more likely to experience a 50% reduction in panic attacks (P < .02). Significant differences were not found with regard to the PDSS-SR and the number of people who experienced no full-blown panic attacks, despite a clear trend towards a more favourable outcome after treatment, as evidenced by the odds ratios in Table 4.3 (2.7 and 3.2).

4.3.2 Pooled outcome and long-term follow-up

Participants After the waiting period, 27 of the 31 participants in the control group embarked on the web-based treatment, of whom 23 (85%) completed treatment, and 20 (75%) started the posttest. Three years after the start of the trial, 47 of the 58 participants (81%) responded to the invitation to participate in the follow-up study. We ran a series of t-tests comparing those who participated in the follow-up and those who did not. These analyses failed to reveal a relation between follow-up participation and either pretest symptom severity or immediate (posttest) treatment response.

Outcome. Table 4.4 shows the results of the assessments in the pooled group at pretest, posttest, and after three years. Immediately after treatment, participants reported significant, moderate to large improvements (.4 < d < .9). Three years
4.3. Results

Table 4.4: Web-CBT: pooled outcome at posttest ($N = 58$) and after three years ($n = 47$)\textsuperscript{a}

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pre</th>
<th>Post</th>
<th>FU</th>
<th>Pre-FU ES\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Attack Frequency</td>
<td>4.2</td>
<td>3.4</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Symptom Count</td>
<td>3.1</td>
<td>1.7</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Attack intensity</td>
<td>4.3</td>
<td>2.2</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>PDSS-SR</td>
<td>8.6</td>
<td>5.0</td>
<td>5.5</td>
<td>4.0</td>
</tr>
<tr>
<td>MI-AAL</td>
<td>2.5</td>
<td>1.0</td>
<td>2.1</td>
<td>.9</td>
</tr>
<tr>
<td>DASS Dep</td>
<td>8.9</td>
<td>8.2</td>
<td>4.3</td>
<td>5.4</td>
</tr>
<tr>
<td>DASS Anx</td>
<td>12.3</td>
<td>8.2</td>
<td>6.3</td>
<td>5.6</td>
</tr>
<tr>
<td>DASS Str</td>
<td>13.5</td>
<td>8.6</td>
<td>8.4</td>
<td>6.6</td>
</tr>
</tbody>
</table>

\textsuperscript{a}To account for attrition at follow-up (19%), pre-treatment to follow-up data were analyzed through multilevel regression modeling. Missing posttest data was imputed using pretest scores.

\textsuperscript{b}PDSS-SR: Panic Disorder Severity Scale Self-Report; MI-AAL: Mobile Inventory - Alone subscale; DASS: Depression Anxiety Stress Scales (Dep: Depression; Anx: Anxiety; Str: Stress). Higher scores indicate less favourable conditions.

\textsuperscript{c}ES: Effect size: Cohen’s $d$ point estimate and 95% confidence interval ($d - CI\textsubscript{95}$ to $d + CI\textsubscript{95}$).

after the start of the trial, participants reported significant further improvements with regard to the frequency and intensity of attacks ($z = 6.2$, $P < .001$; $z = 2.5$, $p = .01$) and avoidance (MI-AAL; $z = 3.1$, $P = .002$). Overall, compared to the immediate (posttest) effects, pretest to follow-up effect sizes were substantially higher ($.7 < d < 1.4$). A majority of the follow-up participants ($n = 32$; 68%) reported no panic attacks in the week preceding follow-up measurement. With regard to the PDSS-SR, 83% of the follow-up participants ($n = 39$) scored below the clinical cut-off (8).

**Additional treatment.** In the period between the posttest and follow-up, 12 participants (26% of those completing the follow-up) received further treatment (medication and/or psychotherapy) for their panic symptoms. With respect to the primary outcome measures, those who received additional treatment did not score differently from those without additional treatment.

**Participant satisfaction.** At posttest, participants rated the overall value of the treatment with a mean score of 8.6 on a 1 to 10 point scale ($SD = 1.3$; range: 4-
and their satisfaction with their therapists with a mean score of 9.0 ($SD = 1.2$).

Eighty-one percent reported a large impact on their daily functioning, 81% thought the participant-therapist contact to be personal, and 71% indicated they had not missed face-to-face contact. At follow-up, these aspects were rated similarly. Participants were also asked to rate the degree to which they would recommend the treatment to others, on a 10-point scale ranging from 1 (No) to 10 (Yes). The average score on this item was $M = 8.9$ ($SD = 1.5$).

**Progress during treatment.** During treatment, after each treatment phase, participants rated the degree to which their panic symptoms impaired daily functioning. As shown by Figure 4.2, impairment declined over the course of treatment, with an effect size of $d = 1.2$ between the first and last measure. Noticeable improvement occurred only after the cognitive restructuring phase. Post-hoc, we tested mean impairment values before this phase with mean impairment values after this phase. This contrast was highly significant ($z = 8.2$, $P < .001$).

**Predictors of outcome.** Using post-treatment PDSS-SR scores and attack frequency as the outcome variables of interest, we tested several predictors of posttreatment symptomatology through two separate multiple regression analyses. Specifically, we tested the significance of pre-treatment symptom severity, agoraphobic avoidance (MI-AAL), gender, education level (low/high), pretest medication status (no/yes), and time of measurement (posttest/follow-up). Further, we explored interaction effects of

![Figure 4.2: Change in impairment during treatment.](image)
the time variable and the other variables, i.e. time by pre-treatment symptom severity, by gender, by education, MI-AAL, and by medication status.

With regard to the PDSS-SR, none of the interaction terms were significant. Hence, the model was simplified to include only the main effects. In this model, medication status \( t(50) = 2.0; P = .05 \) and education \( t(50) = 0.3; P = .76 \) had no effects. However, PDSS-SR pretest scores \( t(50) = 5.7; P < .001 \) and gender \( t(50) = 3.0; P = .019 \) predicted post-treatment symptoms significantly. Post treatment scores were higher for females, and for those displaying higher symptoms levels at pretest.

With regard to attack frequency, only the interaction of time and pre-treatment severity was significant \( t(50) = 3.8; P < .001 \), signalling a significant larger decline in the number of panic attacks at follow-up compared to the posttest in those experiencing more attacks at pretest. In the reduced model, in which we retained only this interaction and the main effects, none of the variables predicted post-treatment scores significantly, including gender and pre-treatment agoraphobic avoidance (MI-AAL scores).

Given the regression results, we further explored the effect of pretest symptom severity. Based upon the PDSS-SR baseline score, we split the sample into two groups (based on the PDSS-SR cut-off of 8) and determined the effect size for each group on this measure. This revealed considerable higher effects in the group with higher pretest levels (posttest \( d = 2.1 \); follow-up \( d = 2.3 \)).

4.4 Discussion

In comparison to a waiting-list, web-based therapist-assisted CBT induced moderate to large reductions in panic symptoms and general psychopathology in a heterogeneous community sample of clients suffering from chronic symptoms of panic disorder. Treated participants were six times more likely to experience a reduction in panic attacks of at least 50%. Client satisfaction was high. After three years, even though the participants never met their therapist, 80% returned for follow-up measurements. On the long term, improvements were found to be more pronounced.
4.4.1 RCT

One of the strengths of this study is that it included a community sample representative of the population that is encountered in everyday practice. Unlike most trials of internet-delivered CBT, we did not exclude respondents with subsyndromal PD/A. Subthreshold PD/A is common and constitutes a major proportion of the cases treated in routine care. Therefore, compared to most existing trials of Web-based CBT, the results of this trial are more likely to generalize to the applied setting.

The average between-effect size found in this study of $d = .7$ is considerably higher than the effects of internet-based programs without therapist support ($d = .2$, Spek et al., 2007). This lends further support to the hypothesis that therapist guidance is a critical determinant of the efficacy of internet-based psychotherapy (Palmqvist et al., 2007). The results also support the suggestion that scheduled therapist guidance increases adherence, as evidenced by the relatively low dropout rate of 13% found in this study, compared to the high dropout rates that are observed in online self-help programs (Eysenbach, 2005; Farvolden et al., 2005).

At first sight, the effect sizes appear somewhat smaller than those observed in trials with less therapist support (Carlbring et al., 2006, 2005, 2001; Klein et al., 2006). However, the larger sample variance in symptom severity in this trial precludes direct comparisons with these programs. Clearly, more research is needed concerning the relation between the amount of therapist involvement and outcome. Such research should involve the systematic manipulation of the amount of therapist input. However, given the large effect sizes that we observed among those with more severe symptom levels, we recommend that such studies include therapist involvement over and above the maximum of six hours of therapist time of existing web-based CBT of PD/A.

4.4.2 Three-year follow-up

Our follow-up study confirms and extends previous studies of web-based treatments demonstrating stable effects of online treatment (Carlbring et al., 2005, 2006; Klein et al., 2006; Shandley et al., 2008). In these trials, effects were maintained up to one year after treatment. Our three-year follow-up suggests that this also holds for the longer term. At follow-up, participants reported high client satisfaction, good
outcome characterized by large effect sizes, and further improvements compared to those observed at posttest. This, together with the fact that participants reported chronic symptoms at pretest, provides preliminary but encouraging evidence that online treatment is effective in inducing sustainable changes.

4.4.3 Limitations

One of the limitations of this study is that we did not use a formal clinical interview to assess treatment outcome. Instead, outcome was measured through self-report questionnaires. One of the key advantages of Internet-based therapy is the potential of reaching people whose access to treatment is poor. The fact that a clinical interview requires face-to-face contact with an expert could be counterproductive in that it may deter those who tend to avoid treatment because they are reticent about such contacts. Nevertheless, observer-rated diagnoses may add to the validity of the results. Telephonic diagnostic interviews may provide a future solution.

A second limitation of this trial is that the long-term effects cannot be attributed to the treatment alone, because the three-year follow-up study was uncontrolled. During the period between the posttest and the follow-up, participants were subject to many influences which we could not control. Specifically, we found that a quarter of the participants had followed additional treatment. However, there was no difference in long-term outcome between those who followed additional treatment and those who did not. We observed this before in our previous studies (Ruuwaard et al., 2007, 2009). Although some clients need additional treatment, online therapy appears to be a sufficient intervention for the majority of clients. Another limitation of the long-term follow-up is that we compared retrospective panic attack estimates at follow-up with more accurate and conservative monitoring of panic attacks at pre- and posttest. de Beurs, Lange, and van Dyck (1992) found that retrospective estimates of panic attack frequency can be substantially higher than those obtained through monitoring. Thus, the observed long-term effects may represent an underestimate. Further comparative studies of the long-term benefits of this treatment are necessary to corroborate our findings.

Third, with the absence of a formal component analyses, it is not possible to identify key ingredients of the therapy. Through repeated measurements during treatment, we
found some support in favor of combined treatment - i.e., a combination of exposure techniques, cognitive interventions, and applied relaxation - because notable change occurred relatively late, after each technique had been applied. More incisive studies would be useful to assess the necessity of individual techniques. However, repeated measurement during treatment has wider use than a measure of efficacy alone. Our therapists indicated that they found the impairment scores very useful as an additional check on the progress of their clients.

4.4.4 Conclusion

Our study provides further evidence for the efficacy and feasibility of web-based therapist-assisted CBT. Web-based CBT is a promising online alternative to traditional treatments for panic symptoms, and we recommend exploring its effectiveness through field-testing in routine practice. From a cost-benefit point of view, future research should develop criteria to guide the identification of clients who require varying intensities of (web-based) treatment. Costs, however, should come second to outcome.
4.4. Discussion