Short- and Long-Term Effectiveness of High-Density Exposure for Panic Disorder with Agoraphobia in an Inpatient Setting: A 7-Years Follow-up

Von der Medizinischen Fakultät
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
einer Doktorin der Theoretischen Medizin
genhmigte Dissertation

vorgelegt von

Dipl.-Psych. Silvia Uhle

aus

Gelsenkirchen

Berichter: Herr Universitätsprofessor
Dr.phil. Dipl.-Psych. Siegfried Gauggel

Herr Universitätsprofessor
Dr.med. Tilo Kircher

Tag der mündlichen Prüfung: 11. Dezember 2006

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.
# List of contents

0. Abstract 1

1. Introduction 2

2. Method 7
   - Participants 7
   - Treatment 9
   - Measures 9
   - Rasch analysis 11
   - Statistical analyses 16

3. Results 18
   - Long-term effects of HDE 18
   - Comparison of treatment completers and dropouts 20
   - Comparison of patients who submitted vs. not submitted questionnaires at follow-up 21

4. Discussion 22

5. References 28

6. Danksagung 36

7. Lebenslauf 37
0. Abstract

Objective: This study examined the effectiveness up to 7 years post-treatment of individual high-density exposure (3-4 weeks, all day) for panic disorder with agoraphobia (PDAG).

Method: Unselected inpatients (N = 379) with a primary diagnosis of PDAG were treated in the Christoph-Dornier-Clinic (CDK) for Psychotherapy in Germany (Muenster). Outcome measures (i.e., self-report questionnaires) were given pre-treatment, 6-weeks, 1-year and 7-years post-treatment and data were analyzed with Item Response Theory (Rasch analysis).

Results: Highly significant symptom reductions in anxiety cognitions, anxiety body sensations, agoraphobic avoidance, depressive symptoms, and general psychopathology could be found at the 6-weeks follow-up, also at 1-year follow up, and 7-years post-treatment. At the 6-weeks follow-up the effect sizes ranged from 1.08 to 1.62 with an average effect size of 1.29; at the 1-year follow-up from 1.07 to 1.48 (average effect size was 1.26). Seven years after discharge the effect sizes ranged from 1.15 to 1.33 with an average of 1.22.

Conclusions: The results suggest that high-density exposure for treatment of PDAG can be transferred to a natural clinic-setting without reducing its effectiveness. There is convincing evidence for long-lasting effectiveness up to seven years after treatment.

Keywords: Agoraphobia, panic disorder, behavior therapy, exposure, effectiveness study, Rasch analysis
1. Introduction

Efficacy vs. effectiveness studies

For the past three decades many investigators have been searching for a beneficial effect of psychotherapy methods. They conducted so called efficacy studies which are mostly performed in a structured university laboratory setting and imply randomized clinical trials (Grawe, Donati, & Bernauer, 1994; Smith, Glas, & Miller, 1980). Their high internal validity is reached by homogenous samples, treated by well-trained clinical investigators, manualized treatments and a fixed number of sessions. Patients with comorbid disorders are excluded. The target outcome of each patient is systematically assessed by “blind raters” (Seligman, 1995; Fishman, 2000a).

Chambless & Hollon (1998) pointed out that there is a need to identify empirically supported treatments (EST’s), which are defined as “clearly specified psychological treatments shown to be efficacious in controlled research with delineated populations”. Guidelines were published to establish these EST’s in the clinical practice (APA, 1996): In Germany Dengler & Selbmann (2000) published a guideline for diagnostic and treatment for anxiety disorders.

But the way in which psychotherapy is studied in efficacy studies is not comparable to treatment in clinical practice (Kadzin, 2003). Patients in clinic settings have multiple problems and comorbid diagnoses. Furthermore, the duration of their treatments is linked to treatment progress or insurance limits. Clinical treatment means that psychotherapy is self-correcting; it is well adjusted to the individual aims of patients and their improvement. The context of a clinic setting is influenced by a myriad of uncontrollable factors which do not affect the context of laboratory research. All these characteristics lead to the question whether the positive findings achieved in efficacy studies can be replicated under less well controlled but more realistic clinical conditions (Russel & Orlinsky, 1996).

Since the mid 90th psychotherapy researchers keep themselves busy by transferring results of efficacy studies (research therapy) to the daily psychotherapeutic practice (Hoagwood, Hibbs, Jensen, & Brent, 1995; Seligman, 1995; Weisz, Donenberg, Han, & Weiss, 1995). In the meta-analysis of Weisz, Weiss, & Donenberg (1992) six studies with child and adolescent therapies were listed in which psychotherapy was ineffective under clinical conditions. However, there are several other effectiveness studies with evidence that empirically supported psychotherapies work well under clinical conditions.

Shadish et al. (1997) and Shadish, Navarro, Matt, & Phillips (2000), for example, conducted a secondary analysis of past meta-analytic data of psychotherapies. Only few studies were found which fulfilled the criteria for a clinical therapy: they showed the same
effects than efficacy studies under laboratory settings. Some of these effectiveness studies were published in the last five years. Wetzel, Bents, & Florin (1999) found evidence for the application of high-density exposure (HDE) with response prevention for obsessive-compulsive disorder in the Christoph-Dornier-Clinic for Psychotherapy (CDK). The results of the 85 unselected treated patients are comparable with findings of controlled studies, effect sizes at 6-weeks and 1-year follow-up were greater than 1.0 for all measures. Another group of investigators has replicated success rates for patients with obsessive-compulsive disorder treated with exposure and response prevention (Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000). Tuschen-Caffier, Pook, & Frank (2001) evaluated the effectiveness for CBT in a sample of 73 unselected patients with bulimia nervosa from an outpatient clinic and found comparable results to efficacy studies. Lincoln et al. (2003) examined the effectiveness for CBT for social phobia for 217 unselected patients treated in four outpatient clinics. The results are also comparable to average effect-sizes reported by meta-analytic studies.

Panic disorder with agoraphobia (PDAG)

PDAG is a common (Jacobi et al., 2004; Weissman et al., 1997) and disabling disease with high costs for the health systems (Mendlowitz & Stein, 2000; Pollack & Marzol, 2000). The lifetime prevalence in the United States is 3.5% for panic disorder and 5.3% for agoraphobia (Kessler et al., 1994). In Germany prevalence rate is 2.4 % and 5.7%, respectively (Wittchen, Essau, von Zerssen, Krieg, & Zaudig, 1992). In the first nationwide German mental health survey (the German Health Interview and Examination Survey, GHS-MHS) a lifetime prevalence of 3.9% for PDAG was found (Jacobi et al., 2004).

Patients with a panic disorder experience recurrent unexpected panic attacks with physical symptoms like palpitations, dizziness, shortness of breath, sweating, and trembling or shaking. There is fear that a panic attack may reoccur (fear of fear), concern about implications or consequences of the attacks (e.g., fear of loosing control or going crazy, fear of dying) and/or an impressive change of behavior (e. g. avoidance of physical strain). Patients with agoraphobia avoid situations in which panic attacks could be triggered and from which it would be difficult to escape or no help could be achieved in the event of a panic attack (e.g. underground train, bus, elevators, crowded shopping malls or restaurants, movie theaters etc.) (APA, 1994; Barlow, Raffa, & Cohen, 2002).

The possibility of incidence of comorbid disorders is quite high. In the German Health Interview and Examination Survey (GHS-MHS) about 88.3% patients with PDAG had at least one comorbid mental disorder (Jacobi et al., 2004). In the National Comorbidity Survey (USA) comorbidity rate was 87.6% (Magee, Eaton, Wittchen, McGonagle, & Kessler, 1996). If patients with panic disorders are not treated there is a high risk for the
development of a chronic course (Noyes et al., 1990; Yonkers, Bruce, Dyck, & Keller, 2003; Wittchen, 1991) often with substance abuse, suicidal tendencies and cuts in social and private (Markowitz, Weissman, Quellette, Lish, & Klerman, 1989). Patients with a panic disorder, especially in combination with agoraphobia, have a chronic course with a high rate of relapse after remission (Keller et al., 1994).

**Efficacy studies for the treatment of patients with PDAG**

Several efficacy studies provide evidence for efficacy of cognitive-behavioral therapy (CBT) for patients with PDAG (e.g., Barlow, Gorman, Shear, & Woods, 2000; Bouchard et al., 1995; Brown & Barlow, 1995; De Beurs, van Balkom, Lange, Koele, & van Dyck, 1995; Fava, Zielezny, Savron, & Grandi, 1995; Michelson, Marchione, Greenwald, Testa, & Marchione, 1996; Öst, & Westling, 1995; Sharp & Power, 1996). Additional effects for exposure in vivo can be found in several meta-analytic studies (Bakker, van Balkom, Spinhoven, Blaauw, & van Dyck, 1998; Chambless & Gillis, 1993; Clum, Clum, & Surls, 1993; Cox, Endler, Lee, & Swinson, 1992; Gould, Otto, & Pollack, 1995; Mattick, Andrews, Hadzi-Pavlovic, & Christensen, 1990; Ruhmland & Margraf, 2001; Trull, Nietzel, & Main, 1988; van Balkom et al., 1997; Westen & Morrison, 2001). The effect sizes for panic symptoms in these studies varied from 0.62 to 2.11, for avoidance behavior from 1.03 to 3.23, for anxiety from 0.55 to 1.57, for depression from -0.78 to 1.44 and for global severity (Symptom-Checklist-90-Revised) from 0.99 to 1.45.

Treatments that show statistically effects in several efficacy studies can be declared as “empirically supported treatment” or “EST” (Chambless & Hollon, 1998; Kendall, 1998). So there is broad evidence for the exposure in vivo as an empirically supported treatment for PDAG. However, only few efficacy studies with patients in an inpatient setting are reported. For example, in the meta-analytic study by Ruhmland and Margraf (2001) only 4 studies from over 45 studies were conducted with an inpatient sample. Furthermore, also the question arises whether the findings of the efficacy studies can be replicated in a clinical setting.

**Effectiveness studies for the treatment of patients with PDAG**

During the last six years a few effectiveness studies were conducted for PDAG. Wade, Treat, and Stuart (1998) compared the therapy results from 110 outpatients with panic disorder with results from two controlled efficacy studies (Barlow, Craske, Cerny, & Klosko, 1989; Telch et al., 1993) and found comparable results at discharge and at a 1-year follow-up (Stuart, Wade, & Treat, 2000). In a naturalistic study Martinsen, Olsen, Tønset, Nyland, and Aarre (1998) found, that for an unselected group of 83 inpatients with PDAG, treated with cognitive-behavioral group therapy, CBT appeared to be effective in a
general clinic setting up to the 1-year follow-up. Chambless, Renneberg, Gracely, Goldstein, and Fydrich (2000) studied a sample of 51 agoraphobic outpatients in a clinical setting, participating in a 2-week intensive group-treatment program with a focus on exposure in vivo and found comparable results as those in efficacy studies.

In Germany to our knowledge three naturalistic studies have been published so far with patients with PDAG. Rief, Trenkamp, Auer, and Fichter (2000) found evidence for the effectiveness of CBT-group therapy in a sample of 80 inpatients, also in a further study 2003 with 165 inpatients, (Rief, Auer, Wambach, & Fichter, 2003), treated in a routine clinical care setting. Results at 1-year follow-up were comparable to those from efficacy studies. Peikert, Wagner, Tauber, Gruhn, and Sobanski (2004) conducted an effectiveness study with an inpatient sample with PDAG. Eighty patients were treated with individual exposure in vivo therapy and showed comparable improvements with patients in efficacy studies. Hahlweg, Fiegenbaum, Frank, Schroeder, and von Witzleben (2001, 2004) examined the effectiveness of individual high-density exposure therapy (HDE) for 416 unselected patients with PDAG, treated in three outpatient clinics of the Christoph-Dornier-Foundation (CDS). They found effect sizes at 6-weeks and 1-year follow-up comparable to those reported from efficacy studies.

Taken together, there is some evidence for transferring the results from efficacy studies for CBT to daily clinical practice in patients with PDAG. However, there is lack of evidence for an empirical based treatment for PDAG in an inpatients sample. Therefore, the first aim of the present study is to replicate previous findings of effectiveness studies which found comparable effect sizes to efficacy studies for cognitive behavioral treatment of PDAG. A large sample size of patients treated in an inpatient naturalistic setting will be included for a better generalization of the CBT-treatment program.

A second purpose for the present study is the assessment of the long-term stability of treatment effects. Many researchers call for studies that show long-term effects of a treatment first before generalizing the results. Additionally, there is need for long-term naturalistic studies to estimate the cost-effectiveness of a treatment program. Up to now there are only few efficacy studies for treatment of panic disorders with follow-up periods longer than two years (Burns, Thorpe, & Cavallaro, 1986; Emmelkamp & Kuipers, 1979; Fava et al., 2001, 1995; Fiegenbaum, 1988; Kendall & Southam-Gerow, 1996; Lelliott, Marks, Monteiro, Tsakiris, & Noshirvani, 1987; McPherson, Broughtman, & McLaren, 1980; Munby & Johnston, 1980; Swoboda, Amering, Windhaber, & Katsching, 2003).

The few available effectiveness studies have shown that for panic disorders treatment effects are still visible 1 year after the end of treatment (Chambless et al., 2000; Hahlweg et al., 2001; 2004, Martinsen et al., 1998; Rief et al., 2000; Stuart et al., 2000). For clinical practice this is an important finding. However, the question remains how long treatment
effects of a naturalistic treatment last. Therefore, the current study investigates the long-term (> 1 year) effects of the naturalistic treatment. For the one year follow-up we expect to find the same effect sizes as in the other naturalistic studies as described previously. Furthermore, we expect that improvements remain stable over follow-up.

Another important aspect of the current study is that outcome measures (i.e., self-report questionnaires) will be analyzed with Item Response Theory (IRT, for an introduction see Bond & Fox, 2001; Embretson & Reise, 2000). IRT offers several advantages over Classical Test Theory (CTT). The estimation of items and person parameters is independent from normative data (sample free test calibration) and used items (item free person measurement). This so called specific objectivity of comparisons is needed when different measurements and repeated measures are used. Also a transformation from ordinal-scaled raw data to interval-scaled scores will be possible. This will be especially useful when comparing several ratings made in the course of treatment.

To summarize, the current study attempts to replicate and extend previous findings by effectiveness studies that found comparable effect sizes to efficacy studies for cognitive behavioral treatment of PDAG in a naturalistic setting. A large sample size and repeated measurement up to seven years post-treatment are used. In addition, IRT is used for analyzing questionnaire data.
2. Method

**Participants**

Participants were 379 inpatients with a primary diagnosis of panic disorder agoraphobia (PDAG) according to the DSM-III-R (APA, 1987), who completed treatment in the Christoph-Dornier-Clinic for Psychotherapy (CDK) in Muenster (Germany) in the years 1993-1995. The diagnosis was verified by the Diagnostic Interview for Mental Disorders (DIPS), a German modified version of the Anxiety Disorders Interview Schedule-Revised (ADIS-R; DiNardo & Barlow, 1988; German version: Diagnostisches Interview bei Psychischen Stoerungen DIPS Diagnostic Interview for Psychological Disorders, Margraf, Schneider, & Ehlers, 1991). The DIPS, which was administered by trained clinical psychologists, is a reliable and valid structured clinical interview for the DSM-III-R. Each diagnosis was agreed on in supervision by the senior clinical psychologist or senior psychiatrist of the CDK. Exclusionary criteria were psychosis, drug or alcohol dependency or an anxiety disorder due to a medical condition.

Mean age of the sample was 34.7 (SD = 8.6, range = 16 – 64); 70% were women; the mean years in school (university education was not considered) were 11.7 (SD = 3.46, range = 9-19). Fifty-seven percent were employed, 8% in job training or students; 16% were housewives, 11% were unemployed, 2% were not able to work and 1% were retired. Fifty-seven percent were married, 36% unmarried, and 5% divorced. The mean duration of the disorder was 7.8 years (SD=6.8) with a range of 0 – 45; the mean age at the time of onset was 26.8 years (SD = 8.3, range 15-61). Three hundred and five patients (80%) had been in out-patient treatment at least once before (mean 1.4; SD = 1.2; range = 0-9), 184 (49%) of the patients had been in in-patient treatment at least once with a mean of 0.9 (SD = 1.5; range 0-21); Seventeen percent were taking antidepressive, and 24% anxiolytic medication before treatment in the CDK. The mean duration of treatment in the CDK was 16 days (SD = 5.4; range = 3 - 50); the patients were treated by 37 different therapists (73.3% of therapist were female). Eighty-seven percent of patients had none comorbid diagnosis, 15% had one and 3% had more than two. Documentation of comorbidity is limited because Axis I comorbidity was not assessed thoroughly from the beginning of data collection, also there was the lack of assessing Axis II diagnosis. Details of the sample are listed in Table 1.

Clinical characteristics and demographic variables of our sample are comparable to those from other efficacy outcome studies (Fava et al., 1995; van Balkom et al., 1997) and
effectiveness studies (Hahlweg et al., 2001, 2004; Martinsen et al., 1998; Rief et al., 2003, 2000).

Of the 379 patients who completed treatment we were able to contact 340 patients (90%) for the follow-up assessment. 314 patients sent back the 6-weeks follow-up (83%), 274 sent back the 1-year follow-up (72%) and 190 patients could be contacted successfully after a period of 7 years (50%).

Table 1. Sociodemographic and clinical characteristics of the sample (N=379)

<table>
<thead>
<tr>
<th>M</th>
<th>(range)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>34.7</td>
<td>(16-64)</td>
</tr>
<tr>
<td>Years in school</td>
<td>11.7</td>
<td>(9-19)</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>26.6</td>
<td>(0-61)</td>
</tr>
<tr>
<td>Duration of disorder (years)</td>
<td>8.1</td>
<td>(0-60)</td>
</tr>
<tr>
<td>Duration of being in the CDK (days)</td>
<td>16.0</td>
<td>(3-50)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>265</td>
</tr>
<tr>
<td>Male</td>
<td>114</td>
</tr>
<tr>
<td>Occupational Status</td>
<td></td>
</tr>
<tr>
<td>Full employment</td>
<td>158</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>59</td>
</tr>
<tr>
<td>Unemployed</td>
<td>41</td>
</tr>
<tr>
<td>In apprenticeship or student</td>
<td>32</td>
</tr>
<tr>
<td>Housewife</td>
<td>60</td>
</tr>
<tr>
<td>Incapacity to work</td>
<td>8</td>
</tr>
<tr>
<td>Pension</td>
<td>4</td>
</tr>
<tr>
<td>Others</td>
<td>16</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>137</td>
</tr>
<tr>
<td>Married (living together)</td>
<td>201</td>
</tr>
<tr>
<td>Married (not living together)</td>
<td>15</td>
</tr>
<tr>
<td>Divorced</td>
<td>20</td>
</tr>
<tr>
<td>Comorbid diagnosis (Axis I)</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>311</td>
</tr>
<tr>
<td>One</td>
<td>57</td>
</tr>
<tr>
<td>&gt; One</td>
<td>11</td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation.
Treatment

In the CDK patients with anxiety disorders are treated on an inpatient basis with “high-density” cognitive-behavioral in vivo exposure (HDE). This program is characterized by symptom-focused and highly individualized interventions (Bartling, Fiegenbaum, & Krause, 1980; Fiegenbaum, 1986, 1988; Marks, 1986, 1993; Tuschen & Fiegenbaum, 1997). Average length of stay is 3-4 weeks. Each inpatient is treated by one therapist only, up to 8 hours a day.

HDE typically consists of four phases: (1) Psychological assessment (see below) and a medical check-up with a diagnostic feedback and information about the individual therapy-program (duration 6-8 sessions within one or two days). (2) Cognitive preparation for therapy which is necessary to enhance the patient’s motivation for the exposure in vivo phase. The patient’s core assumptions of the etiology of his/her disorder is integrated into the disorder-formulation. Due to that model the therapist explains the implications for therapy, and elaborates further steps in treatment. Also, the patient receives detailed information about HDE. (3) Phase of the in vivo exposure (HDE). A large number of in vivo exposures are performed within 2-3 weeks (Foa, Jameson, Turner, & Payne, 1980). In the CDK patients are confronted (up to 8 h daily) with anxiety stimuli that frighten them most. (4) Self-management: confrontation without a therapist has to be performed after several days of HDE. At the beginning of the self-management phase, the tasks are planned and organized with the help of the therapist. During the course of therapy the patient is asked to find out and ‘create’ his/her own most difficult tasks. After therapist-independent in vivo exposures difficulties and problems of exposures are evaluated. Once the patients are back home they stay in contact with the therapist for at least 6 weeks until the telephone contacts are reduced or put to an end.

Measures

A battery of self-report questionnaires was administered before treatment and 6 weeks, 1 year and 7 years after the treatment.

Agoraphobic Cognition Questionnaire (ACQ). The 14-item questionnaire ACQ (Chambless, Caputo, Bright, & Gallagher, 1984; German version: Ehlers, Margraf, & Chambless, 2001) was used to assess cognitions related to the negative social consequences and/or the threat of the anxiety disorder. The ACQ consists of 14 items and assesses cognitive beliefs associated with panic disorder (e.g., “I’m going to die, throw-up,
pass out, lose control," etc.). Items may be scored as a total scale, or according to its two subscales: Loss of Control (consists of 7 items) and Physical Concerns (consists of 5 items). The subscale or total scores are calculated by averaging the responses to the individual items composing that score. The ACQ scores from 1 to 5, higher scores reflect more encroachment. Persons without a diagnosis of anxiety disorder reached a mean score of 1.32 (SD = 0.32) (Ehlers et al., 2001).

**Body Sensations Questionnaire (BSQ).** The BSQ evaluates the degree to which individuals fear somatic symptoms commonly associated with panic. The BSQ total score consists of the average of responses to the 17 items, or of responses to whatever items the client rated if items were skipped (Chambless et al., 1984; German version: Ehlers et al., 2001). With a score from 1 to 5 the higher scores indicate more fear of somatic symptoms. Mean score for people without an anxiety disorder is 1.65 (SD = 0.51) (Ehlers et al., 2001).

**Mobility Inventory for Agoraphobia (MI).** The MI was used to assess phobic avoidance (Chambless, Caputo, Jasin, Gracely, & Williams, 1985; German version: Ehlers et al., 2001). The MI includes two subscales for determining the level of phobic avoidance when the patient is alone (MIA, 26 items) and when accompanied (MIB, 27 items). The subscales are scored separately and have been found to possess good psychometric properties in clinical samples (Chambless et al., 1985). The avoidance is rated on a 1 to 5 scale, high scores show high avoidance. The subscales scores are an average of the item-response. For control groups without a diagnosis of anxiety disorder the mean score for the MIA is 1.45 (SD = 0.48) and for the MIB 1.22 (SD = 0.35) (Ehlers et al., 2001).

**Beck Depression Inventory (BDI).** The level of depressive symptoms was assessed by the BDI, which is a 21-item questionnaire presented in multiple choice format (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; German version: Hautzinger, Bailer, Worall, & Keller, 1995). The BDI is a reliable and well-validated measure of depressive symptomatology. Potential scores range from 0 to 63. Higher scores correspond to higher levels of depression; a cut off point of 11 is regarded as a good indicator of depressive symptoms. People without any psychiatric diagnosis show a mean raw score of 6.45 (SD = 5.2) (Hautzinger et al., 1995).

**Symptom Checklist-90-Revised (SCL-90-R).** The SCL-90-R is a brief, multidimensional self-report inventory designed to screen for a broad range of psychological problems and symptoms of psychopathology (Derogatis, 1986; German version: Franke, 1995). Patients rate each symptom on a 5-point scale (from 0 to 4) for intensity and symptom domains. The total number of symptoms is calculated as an item average for all subscales, high scores indicating a great strain on psychopathology (Global Severity Index GSI). A normal control group showed a mean GSI-score of 0.33 (SD = 0.25) (Franke, 1995).
**Self-rating of improvement.** Patients estimated their improvement on a 7-point rating scale at the end of the treatment program (6-weeks follow-up), at the 1-year follow-up and the 7-years follow-up. The rating-points are 1 = very much better, 2 = much better, 3 = better, 4 = no change, 5 = worse, 6 = much worse, and 7 = very much worse.

The reported questionnaires were given to the participants at the first appointment in the clinic. The same questionnaires were sent 6 weeks, 1 year and 7 years after discharge. The computer input of data was done by students throughout the years 2002 to 2003.

For the present sample the results of the questionnaires ACQ, BSQ, MIA, MIB, BDI and the SCL-90-R (GSI) show a high symptomatology at pre-treatment and are comparable with other clinical groups from previous published effectiveness studies (for details see: Hahlweg et al., 2001, 2004; Martinsen et al., 1998; Rief et al., 2003, 2000).

**Rasch analysis**

The items of the six self-report questionnaires ACQ, BSQ, MIA, MIB, BDI and SCL-90-R were analyzed using the Rasch model to construct linear measures and to evaluate changes in these measures over time. The Winsteps computer program for the analysis of rating scale data was used (Linacre & Wright, 1999). The Rasch model is based on the work of the Danish mathematician George Rasch (for an introduction to Rasch analysis see Bond & Fox, 2001; Embretson & Reise, 2000; Wright & Stone, 1979). Rasch analysis helps to define the underlying construct by revealing the hierarchy of item difficulties and the properties of the rating scale. If the data meet the requirements of the Rasch model sufficiently, one obtains estimates of interval-level measures from the ordinal scores. This feature is especially useful when comparing ratings, such as those between staff and patient or comparing several ratings made in the course of a treatment (Wright & Linacre, 1989).

A common metric (logits or log-odd units) makes it possible to compare the individual items with regard to their difficulty. Besides estimating item difficulties, the ability of the tested persons (e.g. level of anxiety) can also be depicted in the same metric on an interval scale. The mathematical properties of the Rasch model make it possible to estimate the item (item difficulties) and person (person abilities) parameters.

---

1 For the present sample there was at admission a mean raw ACQ score of 2.3 (SD = 0.7), the mean raw BSQ score was 2.8 (SD = 0.7), the mean raw MIA score was 3.3 (SD = 1.1), the mean raw MIB score was 2.4 (SD = 0.9), the raw BDI score was 16.5 (SD = 8.5) and the mean raw GSI score (SCL-90-R) was 1.2 (SD = 0.6).
independently. Therefore, if the Rasch model holds in the sample in question, test construction and test evaluation are independent from the sample and items. Several criteria can be used to judge the quality of an instrument with the Rasch model (for a detailed introduction of the interpretation of Rasch indexes see Bode, Heinemann, & Semik, 2000; Bond & Fox, 2001). These criteria include (a) person separation (estimate of spread or separation of persons on the measured variable) and item separation (estimate of spread or separation of items on the measured variable); and (b) scale structure (the extent to which raters are using the steps in the scale correctly and consistently). A clinically useful set of items should define at least three strata of patients and items (e.g., high, moderate, and low levels of impairment) which are reflected in a separation index of 2.0 and an associated separation reliability of .80. Evidence of the intended use of the rating scale is found in monotonically increasing average measures and step calibrations across categories.

Results from Rasch Analysis

Figure 1 and Figure 2 represent exemplarily Rasch item-person maps which report the relations between the person ability estimates and item difficulty estimates for one of the used questionnaires, the BSQ (Body Sensations Questionnaire), for two time points (Figure 1 pre-treatment and Figure 2 six-weeks follow-up). On the left side of the maps the distribution of person measures (ability/level of anxiety) is represented. Person measures being on top of the scale represent patients with high anxiety symptoms, on bottom of the scale patients with low anxiety symptoms. On the right side of the maps the hierarchy of items from most difficult/rarely used (top) items to easiest difficult/frequently used (bottom) items is represented. Regarding to the anxiety symptom severity, items with “high difficulty” were more likely to be responded by patients with more severe anxiety symptoms, whereas items with “less difficulty” were likely to be answered also by patients with less severe anxiety symptoms in the direction of anxiety. The item with the lowest difficulty was “Dizziness”, it was the most frequent responded item which was used, the most difficult item was “Tingling in finger tips” and was rarely used, i. e. only from persons with a high anxiety level. The item hierarchy ties up with the clinical observation that for patients with PDAG the items “Dizziness”, “Heart Palpitations” and “Feeling short of breath” are the most frequently reported symptoms, whereas the items “Tingling in finger tips”, “Knot in stomach” and “Numbness in another part of body/arms/legs” are rarely reported and if they are reported it’s only by patients with severe PDAG symptomatology. This indicates good construct validity for the BSQ. Moreover the order of items was consistent for both time points.
At pre-treatment for the BSQ mean person ability was $-0.19$ (SD = 0.76) and the difference to the mean item estimates was one standard deviation, which indicates that the items are well targeted for the sample.

**Reliability.** Rating scale analysis of the ACQ yielded person separations ranging from 1.52 to 1.82 (reliability: .70 - .77) for the four time points (pre-treatment, 6-weeks, 1-year and 7-years follow-up). The item separations ranged from 5.10 to 10.68 (reliability: .96 - .99). For the BSQ the person separations varied from 2.08 to 2.32 (reliability from .81 to .84), item separations were 4.94 to 9.05 (reliability from .96 to .99). The person separations for the MIA ranged from 2.19 to 2.32 (reliability from .83 to .90), the items separations reached from 4.63 to 7.15 (reliability from .96 to .98), for the MIB person separations from 1.74 to 2.93 (reliability from .75 to .90) and item separations 3.70 to 7.10 (reliability from .93 to .98). For the GSI scale of the SCL-90-R questionnaire the analysis yielded person separations from 4.24 to 5.38 (reliability from .95 to .99) and item separations from 4.60 to 10.12 (reliability from .95 to .99). The BDI shows person separations from 1.81 to 2.31 (reliability from .77 to .84) and item separations from 3.93 to 6.88 (reliability from .94 to .98).

The item separation statistics for all questionnaires and time points indicated that the potential breadth of item difficulties was large enough to clearly distinguish more than three strata of items. The item reliabilities of the used questionnaires indicated excellent data reproducibility for every time point. The person separation statistics were good for nearly all questionnaires. A low person separation and low reliability was found for the ACQ. This depended on the difference between mean person ability and item difficulty estimates, which was two standard deviations and indicated that the items of ACQ questionnaire were too tough for the persons. So for this questionnaire only two strata of patients with different abilities (level of anxiety) could be distinguished.

**Rating scale use.** Across the different rating categories of the six questionnaires ACQ, BSQ, MIA, MIB, BDI and SCL-90-R (GSI), the average ratings increased monotonically. Step disorder was found for the ACQ, MIA and MIB. This result reveals that patients used rating scale categories atypically. Step disorder does not necessarily indicate a misuse of the rating scale but suggests that more categories exist in the scale than are needed to describe the construct (Bode et al., 2000).
Figure 1. Comparison of the person and BSQ-item measures and standard deviations at pre-treatment. Person abilities and item difficulties are expressed in logits. Each “#” on the left side means two, each “.” one person(s). M = mean; S = one standard deviation; T= two standard deviations.
Figure 2. Comparison of the person and BSQ-item measures and standard deviations at 6-weeks follow-up. Person abilities and item difficulties are expressed in logits. Each “#” on the left side means two, each “.” one person(s). M = mean; S = one standard deviation; T = two standard deviations.
Statistical analyses

All data were screened for deviation from normality, outliers and homogeneity of variance, and assumptions for statistical analysis were examined (Tabachnik & Fidell, 1996). To investigate whether patients have earned benefit from the HDE, the measures (logits) of all questionnaires at all four assessment-points of measurements were submitted to a multivariate analysis of variance (MANOVA) with time of measurement (pre-treatment, 6-weeks, 1-year and 7-years post-treatment) as a within-group measure/factor. Treatment completers were compared with patients dropping out during treatment with t-tests for independent groups. Not all contacted patients resubmitted questionnaires for follow-up. Therefore comparisons between contacted vs. not contacted patients were carried out, t-tests for independent groups and chi-square tests (for the categorical variables) were used. As numbers of resubmitted questionnaires differed for every assessment-point, analyses were performed for every measurement point. An alpha level of .05 was used for all statistical tests. To indicate the size of improvement effect size was calculated according to the formula \[
\frac{M_{\text{pretest}} - M_{\text{posttest (6-weeks, or 1-year, or 7-years follow-up)}}}{\sqrt{SD_{\text{pretest}} + SD_{\text{posttest (6-weeks, or 1-year, or 7-years follow-up)}}}}.
\]
According to Cohen (1988), effect sizes are categorized as follows: low: ≤ .40, moderate: .41 to .79, and high: ≥ .80.

The Reliable Change Index (RCI; Jacobson & Truax, 1991) was applied to show reliable change from pretest to posttest score for every individual patient. For a change to be clinically significant, first a patient must change enough so that one can be confident, that the change exceeds measurement error, calculated by a statistic titled the RCI. The RCI is a calculation of a difference score (post-treatment minus pre-treatment) divided by the standard error of measurement (calculated based on the reliability of the measurement). Second a patient must move from one distribution (dysfunctional) into another (functional). With these two criteria there is the possibility to distinguish between three groups of patients: first patients with reliable changes and posttest scores belonging to the functional distribution rather than to the distribution of the clinical sample (clinically significant change, “recovered”), second patients with reliable change (“improved”) and third patients who have not improved or have deteriorated.

For our sample, to assess the move from a dysfunctional population to a functional population after treatment, a cutoff score for every patient was calculated depending on the criterion C for the questionnaires ACQ, BSQ, MIA, MIB and BDI, which based on information from both, functional and dysfunctional populations, and means that data for a
normative sample are required and both distributions are symmetrical (Jacobson & Truax, 1991). As proposed by Schmitz and Davies-Osterkamp (1997) criterion B for the SCL-90-R: GSI had to be used because they could show, that for the SCL-90-R the distributions from dysfunctional and functional patients are not symmetrical. For each outcome measure the percentage of patients with reliable improvement or deterioration was calculated.
3. Results

Results are presented in three sections: First the data for changes were analyzed from pre-treatment to the 6-weeks, 1-year, and 7-years follow-up. Effect sizes and percentages of reliably improved and clinically significant improved patients (Jacobson & Truax, 1991) were calculated. Second, treatment completers were compared with patients dropping out during treatment. Third contacted persons, who submitted follow-up questionnaires, were compared to patients who did not submit their questionnaires.

Please note that in the following sections reported measures are the Rasch-transformed data (logits).

**Long-term effects of HDE**

In Table 2 the person measures, expressed in logits, are listed for each assessment-point, in Table 3 the effect sizes. There was an overall and highly significant decrease of symptomatology in all measures from pre to every follow-up assessment-point, \( F (18;90) = 24.98, p<0.001 \).

**Table 2. Means (M) and standard deviations (SD) of questionnaires**

<table>
<thead>
<tr>
<th></th>
<th>pre-treatment</th>
<th>6-weeks follow-up</th>
<th>1-year follow-up</th>
<th>7-years follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>ACQ</td>
<td>379</td>
<td>-0.74 0.79</td>
<td>297</td>
<td>-1.18 1.18</td>
</tr>
<tr>
<td>BSQ</td>
<td>365</td>
<td>-0.15 0.71</td>
<td>303</td>
<td>-1.34 1.14</td>
</tr>
<tr>
<td>MIB</td>
<td>371</td>
<td>-0.63 1.25</td>
<td>284</td>
<td>-3.02 1.70</td>
</tr>
<tr>
<td>MIA</td>
<td>371</td>
<td>0.28 1.18</td>
<td>290</td>
<td>-1.86 1.65</td>
</tr>
<tr>
<td>SCL-GSI</td>
<td>368</td>
<td>-0.80 0.73</td>
<td>312</td>
<td>-1.74 1.01</td>
</tr>
<tr>
<td>BDI</td>
<td>370</td>
<td>-1.21 0.95</td>
<td>307</td>
<td>-2.63 1.53</td>
</tr>
</tbody>
</table>

Note. Please note that logits are reported in this Table.
ACQ = Anxiety Cognition Questionnaire; BSQ = Body Sensation Questionnaire; MIA = Mobility Inventory, alone; MIB = Mobility Inventory, accompanied; SCL-GSI = Symptom Checklist-90-Revised, Global Severity Index; BDI = Beck Depression Inventory.
Table 3 indicates a noticeable improvement, high effect sizes for all measures from admission to follow-up assessment-points were found. At the 6-weeks follow-up the effect sizes ranged from 1.08 (SCL-90-R: GSI) to 1.62 (MIB) with an average effect size of 1.29; at the 1-year follow-up from 1.07 (ACQ) to 1.48 (MIB) (average effect size was 1.26). 7-years after discharge the effect sizes ranged from 1.15 (BDI) to 1.33 (BSQ) with an average of 1.22.

Table 3. Effect sizes

<table>
<thead>
<tr>
<th></th>
<th>pre vs. 6-w</th>
<th>pre vs. 1-y</th>
<th>pre vs. 7-y</th>
<th>6-w vs. 1-y</th>
<th>6-w vs. 7-y</th>
<th>1-y vs. 7-y</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACQ</td>
<td>1.11</td>
<td>1.07</td>
<td>1.17</td>
<td>0.02</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>BSQ</td>
<td>1.28</td>
<td>1.33</td>
<td>1.33</td>
<td>0.08</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>MIB</td>
<td>1.62</td>
<td>1.48</td>
<td>1.22</td>
<td>-0.12</td>
<td>-0.30</td>
<td>-0.19</td>
</tr>
<tr>
<td>MIA</td>
<td>1.51</td>
<td>1.37</td>
<td>1.26</td>
<td>-0.17</td>
<td>-0.23</td>
<td>-0.07</td>
</tr>
<tr>
<td>SCL-GSI</td>
<td>1.08</td>
<td>1.12</td>
<td>1.20</td>
<td>0.13</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>BDI</td>
<td>1.14</td>
<td>1.19</td>
<td>1.15</td>
<td>0.11</td>
<td>0.05</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

Note. pre = pre-treatment, 6-w = 6-weeks follow-up, 1-y = 1-year follow-up, 7-y = 7-years follow-up. ACQ = Anxiety Cognition Questionnaire; BSQ = Body Sensation Questionnaire; MIA = Mobility Inventory, alone; MIB = Mobility Inventory, accompanied; SCL-GSI = Symptom Checklist-90-Revised, Global Severity Index; BDI = Beck Depression Inventory.

The average percentage of patients with reliable improvement over all questionnaires was 75% at 6-weeks follow-up, 73% at 1-year follow-up, and 72% at 7-years follow-up. The average percentage of reliable deterioration was 1%, 3% and 3%. On average 57% of patients showed clinically significant change at 6-weeks follow-up, 59% at 1-year follow-up and 59% at 7-year follow-up (for details see Table 4).
### Table 4. Percent of patients with reliable change, improvement or deterioration; and clinical significance (‘recovered’) for clinical variables

<table>
<thead>
<tr>
<th>Reliable Change</th>
<th>Sdiff</th>
<th>Improvement 6-w 1-y 7-y</th>
<th>Deterioration 6-w 1-y 7-y</th>
<th>Clinical Significance Cutoff 6-w 1-y 7-y</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACQ</td>
<td>0.42</td>
<td>45 43 45 2 1 1</td>
<td>1.6 32 32 32</td>
<td></td>
</tr>
<tr>
<td>BSQ</td>
<td>0.35</td>
<td>72 74 79 0 2 3</td>
<td>2.2 56 63 63</td>
<td></td>
</tr>
<tr>
<td>MIB</td>
<td>0.25</td>
<td>93 89 84 0 1 5</td>
<td>1.6 75 75 68</td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>0.30</td>
<td>93 91 83 0 1 2</td>
<td>1.8 65 63 59</td>
<td></td>
</tr>
<tr>
<td>SCL-GSI</td>
<td>0.16</td>
<td>82 79 80 1 7 5</td>
<td>0.8 60 63 66</td>
<td></td>
</tr>
<tr>
<td>BDI</td>
<td>4.30</td>
<td>63 63 63 1 4 2</td>
<td>10.3 51 55 63</td>
<td></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td>75 73 72 1 3 3</td>
<td>57 59 59</td>
<td></td>
</tr>
</tbody>
</table>

Note. pre = pre-treatment, 6-w = 6-weeks follow-up, 1-y = 1-year follow-up, 7-y = 7-years follow-up.
ACQ = Anxiety Cognition Questionnaire; BSQ = Body Sensation Questionnaire; MIA = Mobility Inventory, alone; MIB = Mobility Inventory, accompanied; SCL-GSI = Symptom Checklist-90-Revised, Global Severity Index; BDI = Beck Depression Inventory.

In addition to an overall improvement in symptomatology, ratings of subjective improvement also increased. Six weeks after discharge 93% of the patients rated their improvement better to very much better. One year after post-treatment 89% and 7-years after post-treatment 87% reported a better or very much better improvement. No benefit was reported by 4% at 6-weeks follow-up (5% at 1-year follow-up, 7% 7-years follow-up) and deterioration was reported by 4% at 6-weeks follow-up (6% at 1-year follow-up, 6% 7-years follow-up) of the patients.

**Comparison of treatment completers and dropouts**

If patients who not competed treatment are included, 446 patients with PDAG started therapy in the years 1993-1995, but there were some dropouts during treatment because of problems with payment of treatment and the patients’ doubts regarding the rationale for the treatment model of exposure in vivo. Fifteen percent dropped out within treatment (n = 67), 85% completed (n = 379) the treatment. When the dropout patients were compared to the treatment completers only one significant difference emerged. Treatment completers had higher score on the ACQ then dropouts. No differences were found in the other clinical scores and variables.
Comparison of patients who submitted vs. not submitted questionnaires at follow-up

Of the 379 patients who completed treatment we were able to contact only 340 patients (90%) who sent back the follow-up assessment differently. So there is need to look for differences between the patients who sent back the follow-ups at different follow-up assessment-points to ensure that the results are not distorted by group differences. The two groups of participants (the ones who took part in the follow-up evaluations and the ones who did not) were compared by different parameters to see whether there is a group difference in important clinical and sociodemographic variables. Variables were compared between the patients who submitted versus patients who not submitted questionnaires for each follow-up assessment-point. In comparison with the patients who had not sent back any questionnaires, there were significant differences in the GSI of the SCL-90-R, t (367) = -2.01, p = 0.05 and number of other pre-treatments, t (376) = -2.07, p = 0.04. The not submitted patients showed higher scores on these two variables at the beginning of the therapy. But in comparison with the patients who had sent back questionnaires at 6-weeks follow-up, the number of pre-treatments of the not submitted patients was lower than in the other group, t (376) = 2.09, p = 0.04. Within the 1-year follow-up, there was no significant difference and at the time of the 7-years follow-up there was a significant difference in the marital status, (submitted = 61.0% married, not submitted = 45.3% married), $\chi^2(5, n = 187) = 13.34, p = 0.02$. Taken together, there is no consistent significant group difference between patients who submitted vs. patients who not submitted questionnaires. Therefore, we can conclude that both groups are quite comparable on clinical and sociodemographic variables.
4. Discussion

The purpose of the present study was to investigate whether a large number of inpatients with PDAG, treated in a naturalistic clinical setting with high-density exposure in vivo (HDE), show comparable treatment improvements with those of previous published effectiveness studies. Another purpose was to investigate the long-term effects of the applied treatment. In all clinical variables patients showed a highly significant and enduring decrease of both anxiety symptoms and global symptoms. Our results are either comparable or better than results from other naturalistic studies (Chambless et al., 2000; Hahlweg et al., 2001, 2004; Martinsen et al., 1998; Peikert et al., 2004; Rief et al., 2000, 2003, Stuart et al., 2000, Wade et al., 1998). Particularly, it is remarkable that the improvement is continuing up to seven years post-treatment.

Comparisons to other naturalistic studies: The intra-group effect sizes in our study were very high, 1.38 for anxiety symptoms (depression 1.14, global severity index 1.08) at 6-weeks follow-up, 1.31 (1.19 and 1.12) at 1-year follow-up and 1.25 (1.20 and 1.15) at seven years post-treatment. In comparison Hahlweg et al. (2001, 2004) found average effect sizes of 1.23 at 6-weeks follow-up and 1.24 at 1-year follow-up in an unselected sample of outpatients with PDAG. So no differences in effect sizes were found between outpatients from the Hahlweg et al. study and our inpatient sample at 6-weeks and 1-year follow-up. Chambless et al. (2000) published for their outpatient group of 51 patients with PDAG lower average effect sizes of 1.13 one week post-treatment and 1.07 at 6-month follow-up for anxiety symptoms.

In Peikert’s et al. (2004) effectiveness study with an inpatient sample with PDAG (N=80) they found an average effect sizes of 1.42 for anxiety symptoms (measured by the questionnaires ACQ, BSQ, MIA, and MIB), of 1.13 for depression and of 1.20 for global severity index. But their results concern only the discharge measurement, before there was any possibility for patients to evaluate the effectiveness of treatment improvement under natural conditions at home, so their effect sizes seem to be overestimated.

In comparison with the effectiveness study of Rief et al. (2000, 2003), who found an average effect size of 0.85 for anxiety symptoms (measured by the questionnaires ACQ, BSQ, and MIA) the effect sizes in the present study are much higher. This could be due to the fact that in Rief’s et al. study patients were treated with group therapy only, whereas our patients were treated with individual high density exposure in vivo up to several hours per day. So, one could assume that individual exposure in vivo treatment for PDAG is more effective than group therapy. Many studies give prove of efficacy for exposure in
vivo for panic disorder with agoraphobia. In their meta-analysis Ruhmland & Margraf (2001) found for example highest effect sizes for anxiety symptoms treated with exposure in vivo of 1.64, for cognitive-behavioral therapy effect size was 1.19 and for cognitive therapy alone effect size was only 0.92. So for panic disorder with agoraphobia the focuses of treatment should be on exposure in vivo.

**Long-term effectiveness:** In our sample we found a long-term effectiveness of empirically validated exposure treatment for patients with panic disorder and agoraphobia with remarkable effect sizes. Our patients, treated with HDE, showed highly significant decreases in symptoms up to seven years post-treatment, so treatment improvement extents for several years. Many researchers call for studies that show long-term effects of a treatment first before generalizing the results. Naturalistic studies, conducted in the past years, measured the effectiveness of CBT for panic disorder only up to one year post-treatment (Hahlweg et al., 2001, 2004; Martinsen et al., 1998; Rief et al., 2000; Stuart et al., 2000). In the present study we could demonstrate a long-lasting effect for CBT for inpatients with PDAG treated in a naturalistic clinical setting.

Like the previously described effectiveness studies we could demonstrate that cognitive-behavioral treatment for PDAG, especially high-density exposure as individual treatment, can be transferred to a natural setting without reducing its effectiveness on both outpatients and inpatients. Additionally, in our study we could demonstrate for the first time, that there is evidence for long-lasting effectiveness up to seven years after treatment for a large sample size of inpatients with PDAG.

**Agoraphobia as comorbid diagnosis:** In epidemiology studies a lower rate of agoraphobia as comorbid diagnosis is found than in clinical routine, where 80-90% of patients with panic disorder have an additional diagnosis of agoraphobia (Rief et al., 2000). This is an important point because in several studies patients with a comorbid diagnosis of agoraphobia showed lower treatment outcome (Fava et al., 1995; Katschnig et al., 1995). In most efficacy studies under controlled conditions patients with severe agoraphobic symptoms were excluded (e.g. Öst & Westling, 1995), so one could assume, that under naturalistic conditions with a high rate of comorbid agoraphobia symptoms the effectiveness of treatment is much lower than in efficacy studies. In our sample most patients (91%) had a diagnosis of panic disorder with agoraphobia, 6% had a diagnosis of agoraphobia without panic disorder and only 3% had only a panic disorder without agoraphobia. Nevertheless, as already described, we found very high effect sizes comparable to those from meta-analyses. So this could be evidence for the need to
realize special treatment programs like HDE, especially for the agoraphobic symptoms, in clinic routines.

Clinical significance: The clinical relevance of the results found in psychotherapy outcome studies is frequently questioned, because statistically significant improvement does not necessarily mean practically important improvements for the patient. Some methods were developed to investigate clinical relevance in outcome studies of changes in psychotherapy treatment additional to the customary use of only statistical significance (for detailed descriptions of various methods see: Jacobson & Truax, 1991; Kadzin, 1999; Kendall & Sheldrick, 2000; Martinovich, Saunders, & Howard, 1996; Saunders Howard, & Newman, 1988; Wise, 2004). For our sample the common approach of clinical significance (Jacobson & Truax, 1991, 1984), used in several psychotherapy outcome studies, was used. It has to be taken into account that this criterion is very conservative. At pre-treatment 79% of patient scores of measurement (ACQ, BSQ, MIA and MIB, SCL90-R GSI and BDI) were in a dysfunctional distribution. For these patients in average 57% recovered at 6-weeks follow-up, 59% at 1-year and 59% at 7-years follow-up. Seventy-five percent improved at 6-weeks follow-up, 73% at 1-year and 72% at 7-years follow-up, 1% (3% and 3%) of patients showed reliable deterioration at 6-weeks follow-up (1-year and 7-years follow-up). Especially our results at 7-years follow-up are remarkable, nearly sixty percent of our patients recovered, that means they are closer to the mean of the functional population than to the mean of the dysfunctional population. Also over seventy percent showed endurable improvement up to seven years post-treatment. These data are much better than the results of a reanalysis of 11 outcome data from a series of studies investigating the efficacy of exposure-based treatments for agoraphobia (Jacobson, Wilson, & Tupper, 1988). On average only 27% of patients recovered (range 0-54%) at post-treatment, 34% at follow-up (range 0-64%) three to six month after treatment. Fifty-eight percent improved at post-treatment (range 27-82%), 60% at follow-up (range 33-93%) and 5% (range 0-13%) of patients showed reliable deterioration posttest to follow-up. In Chambless’s et al. effectiveness study (2000), 32% of all patients recovered at post-treatment, 40% at 6-month follow-up, 58% improved at post-treatment and 60% at 6-month follow-up. In comparison with Hahlweg’s et al. effectiveness study (2001, 2004), where the same treatment-program was used (HDE) as in our study, the results are very similar. In their sample of 416 unselected outpatients on average 55% recovered at 6-weeks follow-up, 59% at 1-year follow-up. Eighty-one percent improved at 6-weeks follow-up and 79% at 1-year follow-up. Only 5% (6%) of the patients deteriorated at 6-weeks (1-year) follow-up. So there is a strong support for an enduring clinical effectiveness for high-density
exposure (HDE) for unselected outpatients. It is remarkable that our results of clinically significance are much better than these in the reanalysis of Jacobson (1988) and an effectiveness study with agoraphobics (Chambless et al., 2000). One reason could be that there is a better effect of individual therapy which was applied during our treatment than of group therapy used in most studies. Another reason could be the specialized therapy program, the individual high-density exposure (HDE), realized in the Christoph-Dornier-Clinic (CDK) and the Christoph-Dornier-Foundation (CDS).

**Self-rating of improvement:** In addition to the encouraging results of the employed questionnaires the patients global assessments of improvement after treatment were high: over 90% of patients rated themselves being better to very much better at 6-weeks follow-up, 89% at 1-year and 87% at 7-year follow-up.

**Rasch analysis:** In our study the Rasch analysis was used for data analyses. We found good psychometric properties for most of the questionnaires. However, there is one exception. For the ACQ (Agoraphobic Cognition Questionnaire) the person separation statistics indicates that fewer than three strata of differently impaired patients (e.g., low-, moderate and severe symptoms) could be distinguished at every time point (pretreatment, 6-weeks, 1-year and 7-years follow-up), also the reliability was not satisfying. Reasonable person separation and reliability require both ability estimates well targeted by a suitable pool of items and a sufficient spread of ability across the sample. For the ACQ the moderate separation statistics and reliability for the patients are due to the poorer targeting of the items on patients ability estimates for every time point. The items are too difficult for the patients; the least anxious persons have no items to distinguish between, whereas the most difficult questions (e.g., “I am going blind”, “I will hurt someone”) are not sufficiently able to provide good information about persons. For every post-treatment time point for all questionnaires the items are too difficult for the person sample (e.g., see Figure 2). But this result has nothing to do with bad psychometric properties of the used questionnaires; it indicates the patients’ improvement at post-treatment, when items are too difficult for patients because they are less anxious now.
Limitations: There are several limitations of the present study. First, there was a treatment dropout rate of 15%, so the validity of our results must be interpreted carefully. The dropouts were not considered in follow up and could decrease treatment effects. However, when comparing treatment completers and dropouts only one significant difference emerged (treatment completers showed higher scores on ACQ pre-treatment) indicating that treated and not-treated patients are quite similar. Furthermore, our dropout rate is comparable with dropouts in other effectiveness studies, where dropout rates range between 8% (Peikert et al., 2004) and 26% (Wade et al., 1998). In meta-analyses comparable dropout rates are reported, e.g., van Balkom et al. (1997) reported a mean dropout of 16%, Ruhmland and Margraf (2001) reported a mean dropout rate of 17% for patients with panic disorders with agoraphobia treated with exposure in vivo. Another limitation is that not all contacted patients submitted questionnaires. Of the 379 patients who completed treatment we were able to contact 340 patients (90%) for any follow-up assessment. 314 patients sent back the 6-weeks follow-up (83%), 274 sent back the 1-year follow-up (72%) and 190 patients submitted the follow-up questionnaires after a period of 7 years (50%). So this could threaten the validity and generalizability of our findings as, for example, more improved patients may be more likely to send back the follow-ups (Chambless & Hollon, 1998). We cannot rule out that our effect sizes are positively biased although group comparisons between patients who submitted vs. patients who not submitted questionnaires did not indicate differences in any clinical variables.

Unfortunately, Axis I and Axis II comorbidity were not assessed thoroughly from the beginning of data collection, so the real rate of comorbidity is probably much higher than described. As Hahlweg et al. (2001) pointed out, this issue was a problem in the CDS as well as in the CDK and has received greater attention. Another shortcoming is the lack of information about treatment received between discharge and follow-ups. Since we do not know whether patients received treatment during the follow-ups, the enduring effect of our treatment is questionable. It could be that our positive results depend on further treatments in other clinics. At least only a battery of self-report measurement was used. Independent blind assessor ratings are missing and can limit the internal validity of our study, but in a meta-analysis van Balkom and colleagues (1997) found comparable effect sizes calculated for blind assessors to those from self-report measurement.
Conclusions: With his famous question “What treatments work and for whom?” Kiesler (Kiesler, 1966) laid the foundation for the psychotherapy research. With randomized controlled trials (efficacy studies) the "what" question can be answered. However, the answer of the "whom"-question needs an additional step: effectiveness studies in real-world service setting to generalize the findings from internally valid efficacy studies. Putting the findings together there is the possibility to find answers to Kieslers critical questions. Overall several meta-analyses showed high effect sizes for exposure in vivo treatment in a research context with selected samples of patients with panic disorders. The current study indicates that the high-density exposure in vivo treatment routinely offered in an unselected inpatient sample clinic yields comparable results. By considering the limitations of effectiveness studies, there is evidence that HDE can be transported into a natural field setting. With the present study we could demonstrate that inpatients with PDAG, treated in a naturalistic setting of a clinic with a special treatment of HDE, showed a high decrease of symptoms for a long-term up to seven years post-treatment. Further research is needed to bridge the gap between research and practice.
5. References


6. Danksagung

Für die Bereitstellung der zur Durchführung meiner Promotion erforderlichen Daten danke ich den Mitgliedern der Therapeutischen Leitung und der Geschäftsführung der Christoph-Dornier-Klinik für Psychotherapie in Münster.


### 7. Lebenslauf

**Geburtsdatum und -ort**  
11.01.1968 Gelsenkirchen

**Schulabschluss**  
1987 Allgemeine Hochschulreife: Albert-Schweitzer-Gymnasium, Kassel

**Berufsausbildung**  

**Studium**  
1991-1997 Diplom-Psychologie: Georg-August Universität, Göttingen

**Berufstätigkeit**  
1998-1999 Christoph-Dornier-Klinik für Psychotherapie, Münster  
1999-2000 St.-Rochus-Hospital, Telgte  
seit 2000 Christoph-Dornier-Klinik für Psychotherapie, Münster (CDK)  

2000 **Leiterin der Qualitätssicherung**

2004 **Psychologische Leitung der Abt. für Jugendlichenpsychotherapie**

2005 **Stellvertr. Leitende Psychologin der CDK**

2006 **Leitende Psychologin der CDK**

**Weiterbildung**  
1998-2002 Berufsbegleitende Weiterbildung zur Psychologischen Psychotherapeutin (VT), APV, Münster  
2004-2005 Zusatzqualifikation zur Kinder- und Jugendlichenpsychotherapeutin (VT), AKJP, Osnabrück

** Approbation **  
12.2002 Approbation als Psychologische Psychotherapeutin

**Publikation**  