Journal of Consulting and Clinical Psychology
2001, Vol. 69, No. 3, 375–382


0022-006X/01/$5.00 DOI: 10.1037//0022-006X.69.3.375

# Short- and Long-Term Effectiveness of an Empirically Supported Treatment for Agoraphobia

Kurt Hahlweg, Wolfgang Fiegenbaum, Monika Frank, Brigitte Schroeder, and Ines von Witzleben
Christoph-Dornier-Stiftung fuer Klinische Psychologie

This study examined the effectiveness of individual high-density exposure (2–3 weeks, all day) for panic disorder with agoraphobia (PDAG). Participants were 416 unselected patients with a primary diagnosis of PDAG who were treated by 52 therapists in 3 outpatient clinics of the Christoph-Dornier Foundation of Clinical Psychology in Germany. Results 6 weeks after the end of therapy and at the 1-year follow-up showed highly significant reductions in anxiety symptoms, anxious cognition, agoraphobic avoidance, general symptomatology, and depressive symptoms. Results did not differ significantly between the 3 outpatient clinics and are comparable with the average effect sizes reported by meta-analytic studies of controlled efficacy research, using selected patients and specifically trained therapists. Effectiveness was not dependent on duration of disorder, number of treatment sessions, and therapist experience. The study suggests that high-density exposure can be transported from research settings to the mental health field.

How well do the results of empirically supported treatments hold up in natural settings (Chambless & Hollon, 1998; Kendall & Chambless, 1998)? Recently, writers have begun to differentiate between psychotherapy efficacy and effectiveness (Weisz, Donenberg, Han, & Weiss, 1995). Efficacy (or research therapy) refers to the effects of psychotherapy in randomized, controlled trials usually conducted in university settings, with the aim of trying to establish a high degree of internal validity. Effectiveness (or clinic therapy) refers to the effects of "natural" clinical psychotherapy conducted in the field (e.g., in private practice or in mental health centers) using quasiexperimental designs to try to establish a high degree of external validity or generalizability of results to various settings. Although the efficacy of psychotherapy is established, Weisz et al. (1995) reported for child and adolescent therapies modest or nonsignificant effectiveness, challenging the generalizability of the efficacy findings. The first clinically representative controlled study in the child psychotherapy area by Weiss, Catron, Harris, and Phung (1999) reported nonsignificant results and negative effect sizes for clinic therapy in contrast to a control group. Kendall and Southham-Gerow (1996) reviewed the various factors that may contribute to the gap between research and practice outcomes.

Kurt Hahlweg and Brigitte Schroeder, Christoph-Dornier Foundation, Institut Braunschweig, Braunschweig, Germany; Wolfgang Fiegenbaum and Monika Frank, Christoph-Dornier Foundation, Institut Marburg, Marburg, Germany; Ines von Witzleben, Christoph-Dornier Foundation, Institut Dresden, Dresden, Germany.

We thank Anke Ehlers, Juergen Margraf, and Silvia Schneider for their help in developing and implementing the standardized assessment battery. We are grateful to Christine Kroeger for her statistical skills, and we thank Don H. Baucom for his invaluable helpful comments and editorial assistance.

Correspondence concerning this article should be addressed to Kurt Hahlweg, Christoph-Dornier-Stiftung fuer Klinische Psychologie, Institut Braunschweig, Konstantin-Uhde Str. 4, D-38106 Braunschweig, Germany. Electronic mail may be sent to k.hahlweg@tu-bs.de.

Shadish et al. (1997) conducted a secondary analysis of past meta-analysis and found very few studies that were at least somewhat clinically representative and only one that fulfilled the complete set of criteria for clinic therapy. Generalizability studies are therefore needed to explore the transportability of empirically supported treatments to the field of outpatient psychotherapy (Wilson, 1996). Recently, some generalizability studies have been conducted.

In a study using a benchmarking strategy, Wade, Treat, and Stuart (1998) examined the transportability of cognitive–behavioral therapy for panic disorder to a community mental health center (CMHC). The CMHC outcome data for 110 patients were compared with the results of two controlled efficacy studies (Barlow, Craske, Cerny, & Klosko, 1989; Telch et al., 1993). Patients were self-referred or were referred by physicians and mental health professionals and treated by a manualized 15-session panic control intervention (Barlow & Craske, 1994). Unlike the Barlow et al. and Telch et al. studies, no exclusions were made on age, presence of severe agoraphobia, severity of panic attacks, or use of psychotropic medications. Despite differences in settings, patients, and therapists, the CMHC outcomes were comparable with the controlled studies: 87% of patients were panic free at the end of treatment, and patients showed significant reductions in anticipatory anxiety, agoraphobic avoidance, anxiety, and depression. Panic control treatment seems, therefore, transportable to a CMHC.

Two recent effectiveness studies were conducted in Germany. Wetzel, Bents, and Florin (1999) examined the long-term effects of high-density exposure (HDE) therapy with response prevention for obsessive–compulsive disorder (OCD). A sample of 85 unselected inpatients were treated by 28 therapist practitioners. Results showed comparable effects with controlled studies: Success rates were 68% at the 1-year follow-up, and effect sizes were greater than 1.0 for all measures. Tuschen-Caffier, Pook, and Frank (2001) evaluated the effectiveness of cognitive–behavioral therapy for bulimia nervosa. A sample of 73 unselected patients were treated by 16 therapists in an outpatient clinic of the

Christoph-Dornier Foundation for Clinical Psychology (see below). At the 1-year follow-up, significant improvements were found in all outcome variables. The effect sizes were in the range of those found in controlled research.

Another empirically supported treatment for panic disorder with agoraphobia (PDAG) exists: situational in vivo exposure (Barlow, Esler, & Vitali, 1998; Trull, Nietzel, & Main, 1988). However, the transportability of this treatment to outpatient psychotherapy has not been tested as yet and is the focus of this article.

Patients with panic disorder report experiencing recurrent unexpected panic attacks with physical symptoms (racing heart, dizziness, or sweating) leading to continued anxiety focused on experiencing future panic attacks (fear of fear). Those patients with agoraphobia avoid situations that trigger panic attacks and for which escape would be difficult in the event of a panic attack, such as crowded restaurants, department stores, buses, movie theaters, or trains. Lifetime prevalence rates for the disorder range from 3.5% to 5.3% (Kessler et al., 1994). Panic disorder is associated with extensive social and health consequences, for example, increased risk of substance abuse; suicide attempts; impaired social, marital, and vocational functioning; and health problems resulting in increased use of medical care and psychotropic medication (Markowitz, Weissman, Quellete, Lish, & Klerman, 1989).

Several meta-analyses have consistently shown the efficacy of cognitive–behavioral situational in vivo exposure treatment (e.g., Barlow et al., 1998; Chambless & Gillis, 1994; Clum, Clum, & Surls, 1993; Gould, Otto, & Pollack, 1995; Rumland, 2000; van Balkom et al., 1997). The average treatment dropout rate was 16% (range = 3%–25%; Chambless & Gillis, 1994; van Balkom et al., 1997). Average effect sizes varied from 0.74 for depression, 1.18 for anxiety, 0.99 for global severity (Symptom Checklist–90–Revised) to 1.30 for avoidance behavior. Between 60% and 75% of the patients showed evidence of clinical improvement, which was maintained at long-term follow up, ranging from 6 to 15 months (Barlow et al., 1998). Rumland (2000) calculated the intragroup effect sizes for the control groups and found low effect size ranging from .19 (depression) to .02 (avoidance behavior).

In a study by Fiegenbaum (1988), HDE was contrasted with graded exposure in PDAG patients and was significantly superior. Over 75% of HDE-treated patients were found to be symptom free at the 5-year follow-up.

In this article, we examined the transportability and the generalizability of HDE to three outpatient clinics in the community using unselected patients and a large number of experienced and inexperienced therapists. The goals of this study were as follows:

1. Examine the transportability of HDE by comparing the effectiveness and clinical significance of HDE after treatment and at the 1-year follow-up with efficacy studies.
2. Investigate the generalizability of results by examining the differences between the three outpatient clinics and therapists with different levels of experience.
3. Predict outcome using age, duration of disorder, number of treatment sessions, and depression as independent variables.
4. Compare treatment dropouts with treatment completers.

## Method

The Christoph-Dornier Foundation for Clinical Psychology (CDS) was founded in 1989 with the aims of promoting research and clinical practice in clinical psychology. The CDS runs five outpatient clinics in Germany, in which a variety of disorders are treated, in particular patients with anxiety disorders. Patients are referred from different sources, for example, general practitioners, psychotherapists, or psychiatric clinics. Therapists are doctoral students of the CDS, and treatment is supervised extensively by the directors of the respective CDS outpatient clinic. Treatment is paid by the patient's insurance company. Patients with a primary *Diagnostic and Statistic Manual of Mental Disorders* (3rd ed., Rev., *DSM–III–R;* American Psychiatric Association, 1987) diagnosis of PDAG were included in the study. Exclusion criteria were acute alcohol or drug dependency, psychosis, or medical condition not allowing HDE treatment (e.g., myocardial infarct). Other selection criteria were not used.

### Participants

Participants were 416 patients (67% women) with at least pre–post data, treated in three CDS outpatient clinics in the cities of Marburg (MB; founded in 1989), Dresden (DD; founded in 1994), and Braunschweig (BS; founded in 1995). Marburg contributed 62%, Braunschweig 21%, and Dresden 18% of the participants. Data from BS and DD patients were from consecutive admissions and were entered into the clinics' computerized databases within 1 week. Although MB patients were also assessed consecutively, data were infrequently entered into the computer database because of financial restrictions for data management personal. Consequently, not all patients treated in MB were recorded in the database. Therefore, MB patients can be regarded as a representative sample from the total treated population in MB (about 800 PDAG patients).

The mean age of the sample was 35.6 years ($SD$ = 8.9, range = 17–72). Eighteen percent completed secondary school, 26% high school, and 27% had a university degree; 65% were employed, 14% were students or in an apprenticeship, and 14% were housewives. Fifty-two percent were married, 25% lived together with a partner, and 23% were single; 46% were childless.

Mean age at onset was 27.8 years ($SD$ = 8.8), and mean duration of disorder was 8.4 years ($SD$ = 7.1, range = 1–51). Ninety-five percent of the patients had undergone some form of psychotherapy or medical treatment for PDAG; 19% were using antidepressive medication, 37% used anxiolytic medication, and only 35% used no medication.

Characteristics for 677 patients out of 55 studies undergoing exposure in vivo treatment were given in van Balkom et al. (1997): Mean age was 37.4 years ($SD$ = 4.0), the female–male ratio was 2:1, mean duration of disorder was 8.0 years ($SD$ = 3.2), and mean age of onset 29.6 years ($SD$ = 3.6). The socioeconomic and disorder-related data reported here are similar to those from efficacy studies reported by van Balkom et al.

### Treatment

Typically, PDAG patients were treated with high-density cognitive–behavioral in vivo exposure (HDE), typically lasting 4 to 10 days, during which patients are expected to confront the feared situations for several hours per day. Treatment is highly individualized and consists of three phases (see Tuschen & Fiegenbaum, 1997).

1. Psychological assessment (four to six sessions) and a medical checkup, which is particularly important in the context of HDE because this treatment is stressful and may be contraindicated (e.g., for patients with coronary heart disease).

2. Diagnostic feedback and cognitive preparation (about 1 week later, lasting for two to three sessions). Cognitive preparation for therapy is necessary to enhance the patient's motivation for treatment, integrate the patient's core assumptions about the etiology of PDAG into a scientific model, and delineate implications for therapy on the basis of this model. Detailed information on the strategies of HDE is provided. No pressure is exerted on the patient to undergo the treatment. The patient is given 1 or 2 weeks to decide whether to participate in the HDE treatment.

3. When the patient decides to participate, HDE begins (duration is variable and depends on the individual patient's needs). Patients are exposed to the feared situations, starting with one of the most difficult situations. The therapist is in close contact with the patient during the first days, and it is not unusual for treatment to last for 12 hr per day during the first week. Exposure is extended over prolonged periods of time until the anxiety has decreased to a level necessary to achieve habituation. For example, on the first day, a patient is accompanied by his or her therapist and is asked to use public transportation in a large city, go to a popular shopping mall, eat in a crowded restaurant, use the train to another city, and stay overnight in a hotel while the therapist is in a different hotel. During the next day, the patient is asked to use public transportation again, drive to a public park, walk alone for an hour, go to a movie theater, sit in the middle of the row, and go to a disco at night. This program is continued until the patient is able to expose himself or herself to most situations and experience habituation. Already early in treatment, patients may be asked to perform some of the exposure trials by themselves. However, close supervision by the therapist is provided. As therapy continues, the focus is more and more on self-initiated exposure.

HDE was the treatment of choice. However, some patients refused (at the beginning or during treatment) exposure to the most feared situations. Therapists were then flexible to use graded exposure, or in a few cases, panic control treatment in the therapist's office. However, even with these patients the ultimate goal was to use HDE as often as possible in the process of treatment. Therapists were also flexible to use a variety of cognitive interventions to motivate the patient to undergo exposure.

For the present sample, the mean duration of direct therapist–patient contact was 36.2 sessions of 50-min duration each ($SD$ = 17.6). In a U.S. survey of the practice of behavior therapy by Turner, Beidel, Spaulding, and Brown (1995), the average number of 50-min sessions was 32 for the treatment of PDAG patients.

Treatment was conducted by 52 diploma psychologists (roughly equivalent to a master's degree; 72% female, 28% male) with training in behavior therapy. Training in HDE was not delivered in a standardized way and was comparable with the procedures as described by Wade et al. (1998). Training for novice therapists consisted of reading the relevant literature, viewing videotapes of treatment sessions, participating in supervision sessions, and participating as a cotherapist to the clinic director in the treatment of at least 2 patients. Therapists differed in experience: $n$ = 9 were inexperienced (total number of treated patients with any disorder: 1–10), $n$ = 17 had medium experience (11–20), and $n$ = 26 were experienced (>21; range = 21–61). Fifteen of the therapists treated 1 or 2 patients; 16 therapists treated 3 to 7 patients, 14 therapists treated 8 to 14 patients, and 8 therapists treated 15 patients or more (range = 15–24).

### Measures

Patients are assessed before therapy (pre), 6 weeks after the end of treatment (post), and 1 and 5 years thereafter (data not yet available), using an extensive self-report assessment battery that comprised the following instruments.

*Diagnostic interview.* *DSM–III–R* diagnoses are based on the German 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 ADIS–R/DIPS is a semistructured interview with well-established psychometric properties. DIPS interviews were conducted by the therapists, all of whom received intensive training in the use and scoring of the instrument. Each case was reviewed by the clinical director of the respective outpatient clinic. In difficult cases, a consensus diagnosis was derived jointly.

In the beginning of standardized data collection, little emphasis was placed on assessing comorbidity for Axis I disorders. For clinical purposes, it seemed sufficient to establish the primary diagnosis to develop a treatment plan. Similarly, Axis II personality disorders were not assessed, partly because insurance companies will not pay for the extra time needed to assess personality disorders.

*Beck Anxiety Inventory (BAI).* The BAI (Beck & Steer, 1993; German version by Ehlers & Margraf, in press) is a 21-item scale that measures the severity of anxiety symptoms (e.g., numbness or tingling, heart pounding or racing, shaky, feelings of choking). The BAI has a high internal consistency of .92 and sufficient validity data.

*Agoraphobic Cognition Questionnaire (ACQ).* The ACQ (Chambless, Caputo, Bright, & Gallagher, 1984; German version: Ehlers, Margraf, & Chambless, 1993) is a 14-item questionnaire to assess anxiety and agoraphobic cognitions (e.g., fear of dying by myocardial infarction or loss of control). The internal consistency of the German version is .75.

*Body Sensations Questionnaire (BSQ).* The BSQ (Chambless et al., 1984; German version: Ehlers et al., 1993) is a 17-item questionnaire to assess the anxiety with regard to bodily symptoms and has an internal consistency of .85 (German version).

*Mobility Inventory (MI).* The MI (Chambless, Caputo, Jasin, Gracely, & Williams, 1985; German version: Ehlers et al., 1993) consists of 27 items assessing the patient's avoidance behavior with regard to the most common agoraphobic situations. Patients rate the avoidance of the situations twice, when being alone (MIA) or accompanied by another person (MIB). Internal consistencies for the German version are .97 and .96, respectively.

*Beck Depression Inventory (BDI).* The BDI (Beck & Steer, 1987; German version: Hautzinger, Bailer, Worall, & Keller, 1995) is a 21-item self-report questionnaire used to assess the severity of depression. Common depressive symptoms and attitudes are assessed. The BDI is the most frequently used measure of patient improvement in psychotherapy outcome studies. Internal consistency for the German version is .88, and sufficient validity data are provided.

*Symptom Checklist–90–Revised (SCL-90-R).* The SCL-90-R (Derogatis, 1983; German version: Franke, 1995) is a 90-item questionnaire assessing nine primary symptom dimensions and a Global Severity Index (GSI), based on all 90 items. The GSI is used to measure the intensity of the perceived distress. Internal consistency for the German version's GSI is .97. It is most frequently used as part of psychotherapy evaluation.

*Rating of improvement.* A 7-point rating scale (1 = *very much better,* 2 = *much better,* 3 = *better,* 4 = *no change,* 5 = *worse,* 6 = *much worse,* and 7 = *very much worse*) was used to assess the subjective improvement due to the therapy. Patients and therapists rated the degree of improvement 6 weeks after therapy (post) and at the 1-year follow-up.

## Results

Data analyses were performed in several steps. First, treatment completers were compared with patients dropping out after the cognitive preparation phase or during treatment. Second, pre–post and pre–1-year follow-up comparisons were calculated using paired sample $t$ tests with Bonferroni adjustment for each time comparison separately ($p$ = .05/7 = .007). Third, effect sizes were computed, and the percentages of reliably improved and clinically significant improved patients (Jacobson, Follette, & Revenstorf, 1984) were calculated. Fourth, differences between the outcomes of the three outpatient clinics and between inexperienced and experienced therapist were analyzed. Fifth, treatment outcome was predicted using multiple regression analysis, with age, duration of disorder, number of treatment sessions, and depression as independent variables.

### Comparison of Treatment Completers and Dropouts

Of 692 patients who applied for treatment in one of the outpatient clinics, 13% ($n$ = 90) dropped out after the cognitive prep-

aration phase and 8.5% ($n$ = 59) during HDE treatment. Van Balkom et al. (1997) reported a mean dropout of 16% for patients undergoing exposure treatment. However, they did not differentiate between various dropout phases. Therefore, it is somewhat difficult to compare their average rate with ours. When looking only at the treatment dropouts, our rate is comparable with the literature.

The following were the most important reasons for not undergoing treatment after the cognitive preparation phase (typically after six to seven sessions, including three to four diagnostic sessions): (a) The patient had doubts regarding the rationale for the treatment model (35%), (b) there was improvement in symptoms during the cognitive preparation phase (16%), and (c) the treatment seemed too difficult to endure (9%). The following were the most important reasons for dropping out of exposure therapy: (a) There was uncertainty about the rationale for exposure treatment (34%), (b) the treatment was seen as too difficult (10%), and (c) there were organizational difficulties (e.g., the clinic was too far away from home; 9%).

In Table 1, the pretreatment variables for dropouts and treatment completers are shown. Univariate analysis of variance (ANOVA) with least-significant-difference post hoc tests for continuous variables and chi-square tests for categorical variables were used to examine differences between the groups.[1]

When the dropout patients were compared (separately for those dropping out after cognitive preparation and those dropping out during exposure) with treatment completers, three significant differences emerged. First, exposure treatment dropouts had a significantly longer duration of distress than did treatment completers. Second, those dropping out during exposure had significantly higher depression scores than those dropping out after the cognitive preparation phase or treatment completers. Third, treatment completers had a significantly higher education (high school or university degree) than dropouts (52% vs. 38%; $\chi^2(2, N = 270) = 7.3, p = .026$). Dropouts and treatment completers did not differ significantly with regard to age, gender, number of children, or marital status or in the other clinical scales.

## *Transportability of Treatment*

*Pre–post comparisons.* In Table 2, the means and standard deviations of the clinical variables for pre, post, and 1-year follow-up are presented. In paired sample $t$ tests with Bonferroni adjustment ($p < .007$), patient scores on all variables decreased highly significantly from pre to post (see Table 2). From post to the 1-year follow-up, patients improved significantly on the ACQ. In the other variables, there were no significant changes.

*Intragroup effect sizes, reliable change, and clinical significance.* Intragroup effect sizes were calculated using the formula $(M_{\text{pretest}} - M_{\text{postest}})/SD_{\text{pretest}}$ or $SD_{\text{pooled pre/post}}$. According to Cohen (1988), effect sizes are categorized as follows: low: ≤.40, moderate: .41 to .79, and high: ≥.80.

In a recent special section of this journal (Vol. 67, No. 3) on clinical significance (Kendall, 1999), several new methods for conducting normative comparisons were proposed by Kendall, Marrs-Garcia, Nath, and Sheldrick (1999). However, to compare our findings with the literature, we used the two criteria for clinical significance proposed by Jacobson et al. (1984). First, patients receiving psychological intervention should move from a dysfunctional (Dysf) population to a functional (Func) population as the result of treatment. Cutoff scores can be calculated using different methods. In the present study, criterion C (see Jacobson et al., 1984, p. 340) was applied $(M_{\text{Dysf}} * SD_{\text{Func}} + M_{\text{Func}} * SD_{\text{Dysf}})/(SD_{\text{Dysf}} + SD_{\text{Func}})$. Second, change for a patient must be reliable. Here, the Reliable Change Index was applied, based on the difference of the pretreatment score minus the posttreatment score divided by the standard error of the difference. The results for the outcome variables for the different criteria are shown in Table 3.

At postassessment and at the 1-year follow-up, the intragroup effect sizes were high for all the outcome variables, ranging at post from 0.93 (BDI) to 1.82 (MIA), with an average of 1.23. At the 1-year follow-up, effect sizes ranged from 0.92 (BDI) to 1.70 (MIA), with an average of 1.24. Results are comparable with those average effect sizes reported in the meta-analyses: anxiety = 1.18 (1.16, BAI), avoidance behavior = 1.30 (1.82, MIA), and depression = 0.74 (0.93, BDI).

Next, using each specific outcome measure, we calculated the percentage of patients demonstrating reliable improvement or deterioration (see Table 3). Then the average percentage of improvement and deterioration was calculated across these measures. The average percentages of patients with reliable improvement or deterioration were 81% and 5% at post and 79% and 6% at the 1-year follow-up, respectively. At post, on average 55% of patients showed clinically significant changes; at the 1-year follow-up, on average 59% of patients. On the basis of avoidance behavior, at post 68% and at follow-up 70% of patients showed clinically significant improvement. Again, these percentages compare well with the 60% to 75% rate reported in the literature (Barlow et al., 1998).

*Consumer satisfaction.* Improvement after treatment was rated by the patient (and the therapist) on a 7-point rating scale at post and the 1-year follow-up. At post, 84% of the patients (80% of the therapists) rated themselves as being much better or very much better. At follow-up, the rates were 78% and 77%, respectively. Only 2.9% and 2.8%, respectively, reported no change (at follow-up: 4.4% and 5.6%), and 1.6% and 0.6%, respectively, reported deterioration (at follow-up: 4% and 1.8%).

## *Generalizability*

*Differences among the outpatient clinics.* Analysis of covariance with preassessment variables as covariates yielded nonsignificant results for the outcome measures, indicating that the treatment was delivered with the same effectiveness despite the differences in setting, therapists, and supervision.

*Differences among therapists.* ANOVAs with the average effect sizes at postassessment and follow-up showed no significant differences between inexperienced and experienced therapists.

## *Prediction of Outcome*

The last analysis addressed the predictive relationship between (a) age, duration of distress, number of treatment sessions, and BDI preassessment score and (b) the average intragroup effect size

[1]Bonferroni adjustment was not applied because only a few studies examined dropout treatment completer differences previously. The sample size differs across variables because of missing data.

Table 1
*Pretreatment Comparisons of Dropouts After Cognitive Preparation (CP; n = 90) and During Treatment (T; n = 59) and of Treatment Completers (TC; n = 439)*

| | CP | | T | | TC | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *M* | *SD* | *M* | *SD* | *M* | *SD* | *F* | *df* | *p* |
| Age (years) | 37.5 | 9.0 | 35.6 | 9.1 | 35.6 | 8.9 | 1.65 | 2, 555 | .194 |
| Duration (years) | 9.4 | 8.3 | 11.7 | 9.1 | 8.2 | 7.1 | 4.46 | 2, 449 | .012[a] |
| BAI | 23.9 | 11.8 | 26.9 | 12.8 | 26.9 | 12.3 | 2.23 | 2, 579 | .109 |
| ACQ | 2.19 | 0.62 | 2.35 | 0.61 | 2.28 | 0.61 | 1.46 | 2, 581 | .233 |
| BSQ | 2.72 | 0.77 | 2.79 | 0.66 | 2.87 | 0.71 | 1.73 | 2, 573 | .178 |
| MIA | 3.24 | 1.08 | 3.45 | 1.01 | 3.25 | 1.01 | 1.41 | 2, 559 | .246 |
| MIB | 2.33 | 0.98 | 2.47 | 0.89 | 2.35 | 0.89 | 0.47 | 2, 552 | .624 |
| SCL–GSI | 1.04 | 0.63 | 1.25 | 0.61 | 1.10 | 0.59 | 2.31 | 2, 585 | .100 |
| BDI | 16.6 | 10.3 | 20.0 | 9.7 | 15.6 | 8.2 | 6.78 | 2, 582 | .001[b] |

*Note.* Because 104 patients were still in treatment at the time of data analysis, postassessment data were not available. BAI = Beck Anxiety Inventory; 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.
[a] Significant differences between T and TC (post hoc least-significant difference [LSD] test). [b] Significant differences between T and CP and between T and TC (post hoc LSD test).

at postassessment. BDI was included because it was a significant predictor of treatment dropout. The standardized beta weights, explained variance, and zero-order correlations are presented in Table 4. The regression equation was significant, $F(2, 194) = 10.7$, $p < .000$, and explained 13% of the variance of the average effect size at postassessment. More specifically, the higher the depression score at preassessment and the younger the patient, the better the overall outcome at postassessment. Duration of disorder and number of treatment sessions did not enter the regression equation.

## Discussion

The present study provides evidence that the empirically validated situational exposure treatment for PDAG patients can be transported into clinical settings. The results were achieved using a large number of patients and therapists and underscore the generalizability of the results. There is now cumulative evidence that cognitive therapy for patients with bulimia nervosa (Tuschen-Caffier et al., 2001), exposure with response prevention for OCD patients (Wetzel et al., 1999), panic control treatment for panic disorder patients (Wade et al., 1998), and situational exposure for PDAG patients are not only empirically supported interventions by controlled research standards but also very effective treatments in the mental health field. Nevertheless, these conclusions have to be qualified: It is most likely that these results require the frequent and maintained supervision of the therapists and that the institution has to be empirically oriented in supporting the collection of basic patient and clinical outcome data on an ongoing basis.

The present study fulfills most of the criteria for a clinically representative study as defined by Shadish et al. (1997): (a) Treatment was conducted in a nonuniversity setting, (b) it involved patients referred through usual clinical routes rather than solicited by the experimenter, (c) it used patients heterogeneous in personal characteristics, (d) therapists did not use a treatment manual, (e) therapists were free to use a variety of procedures and were not

Table 2
*Means, Standard Deviations, and Paired t Tests for Clinical Outcome Measures at Preassessment (Pre), Postassessment (Post), and 1-Year Follow-Up for Patients With Panic Disorder With Agoraphobia*

| | | Pre | | Post | | Follow-up | | | *t* test | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *N* | *M* | *SD* | *M* | *SD* | *N* | *M* | *SD* | PrPo[a] | PoFU[b] |
| BAI | 399 | 27.0 | 12.4 | 13.8 | 10.3 | 292 | 12.8 | 10.5 | 20.9** | 1.6 |
| ACQ | 400 | 2.3 | 0.6 | 1.7 | 0.5 | 293 | 1.6 | 0.5 | 22.5** | 3.2* |
| BSQ | 390 | 2.9 | 0.7 | 2.0 | 0.7 | 284 | 2.0 | 0.7 | 23.1** | 2.4 |
| MIA | 365 | 3.2 | 1.0 | 1.7 | 0.7 | 269 | 1.7 | 0.8 | 23.4** | 1.1 |
| MIB | 354 | 2.3 | 0.9 | 1.3 | 0.5 | 263 | 1.4 | 0.6 | 23.0** | 1.5 |
| SCL–GSI | 401 | 1.1 | 0.6 | 0.6 | 0.5 | 296 | 0.5 | 0.5 | 20.7** | 1.9 |
| BDI | 400 | 15.6 | 8.1 | 8.1 | 7.7 | 293 | 7.7 | 7.4 | 20.0** | 0.4 |

*Note.* PrPo = pre–post comparison, PoFU = post–follow-up comparison; BAI = Beck Anxiety Inventory; 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.
[a] $df = 401$. [b] $df = 295$.
* $p < .002$. ** $p < .000$.

Table 3
*Intragroup Effect Sizes (IGES); Percentage of Patients With Reliable Change (RC), Deterioration (D), or Improvement (I); and Clinical Significance (CS) for Clinical Variables*

| | IGES | | RC | | | | | CS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Post | | FU | | | | |
| Variable | Post | FU | C | D | I | D | I | Cutoff | Post | FU |
| BAI | 1.16 | 1.16 | 2.5 | 8 | 82 | 10 | 72 | 11.5 | 46 | 49 |
| ACQ | 1.09 | 1.18 | 0.3 | 2 | 70 | 4 | 72 | 1.6 | 38 | 44 |
| BSQ | 1.22 | 1.33 | 0.3 | 4 | 79 | 8 | 77 | 2.0 | 53 | 70 |
| MIA | 1.82 | 1.70 | 0.2 | 3 | 95 | 4 | 92 | 1.9 | 68 | 70 |
| MIB | 1.43 | 1.30 | 0.2 | 3 | 86 | 5 | 83 | 1.4 | 71 | 64 |
| SCL–GSI | 0.99 | 1.06 | 0.1 | 6 | 83 | 6 | 86 | 0.5 | 53 | 55 |
| BDI | 0.93 | 0.92 | 3.4 | 6 | 73 | 7 | 70 | 12.5 | 60 | 61 |
| Average | 1.23 | 1.24 | | 5 | 81 | 6 | 79 | | 55 | 59 |

*Note.* Post = postassessment; FU = follow-up; C = Reliable Change Index; BAI = Beck Anxiety Inventory; 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.

restricted to a fixed number of sessions, and (e) implementation of the treatment was not monitored.

Two criteria were not met: (f) Homogeneous patients with regard to the primary diagnosis (PDAG) were included instead of patients heterogeneous in focal presenting problems, and (g) only about 50% of the therapists can be regarded as experienced, professional therapists with regular caseloads rather than therapists in training or receiving training specifically for that research study. However, one can question the heterogeneous criterion in the context of a behavior therapy transportability study for situational exposure, because exposure is the treatment of choice for phobias only. Using only experienced therapists may not be a valid criterion for a clinically representative study, because there are varying levels of expertise among therapists working in institutions such as community mental health centers or psychiatric inpatient facilities. Therefore, from our point of view, the present study can be regarded as a clinically representative study in the context of the practice of behavior therapy. Most importantly for clinical generalizability is the fact that the treatment was paid by the patient's health insurance, whereas in many research therapies, treatment is delivered free of charge.

The outcome results for patients completing the intervention 6 weeks after the end of treatment (postassessment) and at the 1-year follow-up provided strong support for the clinical effectiveness of HDE for patients with PDAG. In all clinical variables, highly significant decreases in symptoms resulted at postassessment, which stayed stable up to 1 year later. At postassessment and at the 1-year follow-up, the intragroup effect sizes were high for all the outcome variables, in particular for avoidance behavior (effect size = 1.82), with an effect size of 1.23, averaged over all outcome variables. Results are comparable with those average effect sizes reported in the meta-analyses for anxiety, avoidance behavior, and depression (Barlow et al., 1998; Chambless & Gillis, 1994; Clum et al., 1993; Rumland, 2000; van Balkom et al., 1997).

Table 4
*Multiple Regression Analysis Predicting Average Intragroup Effect Size at Postassessment*

| Independent variable | *r* | $\beta$ |
|---|---|---|
| BDI | .30*** | .30*** |
| Age | −.21*** | .20** |
| No. of sessions | .12* | .08 |
| Duration of disorder | −.13* | −.02 |

*Note.* $R = .36$; $R^2 = .13$. BDI = Beck Depression Inventory.
* $p < .05$. ** $p < .01$. *** $p < .000$.

The average percentage of patients with reliable improvement at both assessment points was 80%; however, 6% deteriorated after the treatment. At postassessment, on average 55% of patients showed clinically significant changes, as did 59% at the 1-year follow-up. On the basis of avoidance behavior (MIA), at postassessment, 69% of patients (70% at follow-up) showed clinically significant improvement. Again, these percentages compare well with the rates reported in the literature (Barlow et al., 1998). The consumer satisfaction after treatment and at follow-up were high, with about 80% of patients rating themselves being much better or very much better.

The present study did not use a strict benchmarking approach as Wade et al. (1998) used, because we did not compare our results with specific studies but with results from meta-analyses. Nevertheless, given the large sample size and the number of therapists involved, it seems justified to conclude that HDE can be transported to natural settings without reducing its effectiveness. Given that we did not find any difference between the three settings or therapist experience level, there is also evidence for the generalizability of the results to other settings.

It is interesting to note that for treatment completers, duration of disorder and number of treatment session did not predict outcome. Level of depression and education predicted dropout from treatment, findings similar for panic control treatment (Wade et al., 1998). The greater percentage of patients with lower education among the dropouts call for intensified efforts to tailor the treatment for that population. Certainly more research is needed in that area.

Even when patients accept treatment, high levels of depression are correlated with dropout during exposure treatment. It seems

that HDE may be too difficult for these clinically depressed patients and that it may be advised to treat the depression first before actually engaging the patient in exposure treatment, which requires some activity from the patients—activity that clients with depression often are not able to do. For those who complete treatment, a somewhat higher depression score at preassessment, although not in the clinical range, is predicting better outcome. Therefore, elevated depression scores are not necessarily bad for the treatment process.

One shortcoming of the present investigation is that is entirely based on self-report measures; for example, independent blind assessor ratings are missing and should be included from a methodological point of view. Unfortunately, in the current setting with no extramural funding, it seems impractical to provide such ratings in clinical practice. It is too expensive to hire experienced raters necessary for that purpose, and health insurance companies will not pay for such a measure of quality control. Furthermore, results from a meta-analysis conducted by van Balkom et al. (1994) showed that effect size calculated for blind assessor ratings typically exceed those for self-report. In the light of these findings, self-report data may give a realistic picture, if not a somewhat more negative view of treatment outcome than assessor ratings (Wetzel et al., 1999).

Unfortunately, Axis I comorbidity was not assessed thoroughly from the beginning of data collection. More recently, this issue has received greater attention at the CDS and is included in the current assessment. The assessment of Axis II—personality disorders—is more difficult to achieve on a regular basis in clinical practice because of time constraints and is only performed with specific patients to rule out comorbidity.

Another shortcoming of the present investigation is the lack of treatment integrity data. Apart from the forementioned lack of financial support for collecting and rating integrity data, it is difficult to assess integrity data in situational exposure because most of the treatment happens outside of the office, preventing audio- or videotaping of sessions. It is clear that smaller effectiveness studies investigating the specific methodological issues mentioned above are a way to overcome the practical problems.

## References


American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., Rev.). Washington, DC: Author.

Barlow, D. H., & Craske, M. G. (1994). *Mastery of your anxiety and panic II.* Albany, NY: Graywind.

Barlow, D. H., Craske, M. G., Cerny, J. A., & Klosko, J. S. (1989). Behavioral treatment for panic disorder. *Behavior Therapy, 20,* 261–282.

Barlow, D. H., Esler, J. L., & Vitali, A. E. (1998). Psychosocial treatments for panic disorder, phobias, and generalized anxiety disorder. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (pp. 288–318). New York: Oxford University Press.

Beck, A. T., & Steer, R. A. (1987). *Beck Depression Inventory manual.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., & Steer, R. A. (1993). *Beck Anxiety Inventory - manual.* San Antonio, TX: The Psychological Corporation.

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: The Body Sensation Questionnaire and the Agoraphobic Cognitions Questionnaire. *Journal of Consulting and Clinical Psychology, 52,* 1090–1097.

Chambless, D. I., Caputo, G. C., Jasin, S. E., Gracely, E. J., & Williams, C. (1985). The Mobility Inventory for agoraphobia. *Behaviour Research and Therapy, 23,* 35–44.

Chambless, D. L., & Gillis, M. M. (1994). A review of psychological treatments for panic disorder. In B. E. Wolfe & J. D. Maser (Eds.), *Treatment of panic disorder. A consensus development conference* (pp. 149–173). Washington, DC: American Psychiatric Press.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Clum, G. A., Clum, G. A., & Surls, R. (1993). A meta analysis of treatments for panic disorder. *Journal of Consulting and Clinical Psychology, 61,* 317–326.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Derogatis, L. R. (1983). *SCL-90: Administration, scoring, and procedures manual for the revised version.* Baltimore: Clinical Psychometric Research.

DiNardo, P. A., & Barlow, D. H. (1988). *Anxiety Disorders Interview Schedule—Revised (ADIS–R).* Albany: State University of New York, Phobia and Anxiety Disorder Clinic.

Ehlers, A., & Margraf, J. (in press). *Beck Angst Inventar BAI* (Beck Anxiety Inventory BAI). Göttingen, Germany: Hogrefe.

Ehlers, A., Margraf, J., & Chambless, D. I. (1993). *Fragebogen zu koeperbezogenen Aengsten, Kognitionen und Vermeidung* [Questionnaire for body sensations, anxiety cognitions, and avoidance behavior]. Weinheim, Germany: Beltz.

Fiegenbaum, W. (1988). Long-term efficacy of ungraded versus graded massed exposure in agoraphobics. In I. Hand & H. Wittchen (Eds.), *Panic and phobias: Treatments and variables affecting course and outcome* (pp. 83–88). Berlin: Springer.

Franke, G. (1995). *Die Symptom-Checkliste von Derogatis—Deutsche Version* [Symptom Checklist SCL-90-R by Derogatis—German version]. Weinheim, Germany: Beltz.

Gould, R. A., Otto, M. W., & Pollack, M. H. (1995). A meta-analysis of treatment outcome for panic disorder. *Clinical Psychology Review, 15,* 819–844.

Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1995). *Beck-Depressions-Inventar (BDI).* (Beck Depression Inventory BDI) Bern, Germany: Huber.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352.

Kendall, P. C. (1999). Special section: Clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 283–284.

Kendall, P. C., & Chambless, D. L. (1998). Special section: Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology, 66,* 3–170.

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 285–299.

Kendall, P. C., & Southam-Gerow, M. (1996). Issues in the transportability of treatment. *Journal of Consulting and Clinical Psychology, 64,* 702–708.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM–III–R psychiatric disorders in the United States: Results from the national comorbidity survey. *Archives of General Psychiatry, 51,* 8–19.

Margraf, J., Schneider, S., & Ehlers, A. (1991). *Diagnostisches Interview bei psychischen Stoerungen DIPS* [Diagnostic Interview for Psychological Disorders]. Berlin: Springer.

Markowitz, J., Weissman, M., Quellete, R., Lish, J., & Klerman, G. (1989). Quality of life in panic disorder. *Archives of General Psychiatry, 46,* 984–992.

Rumland, M. (2000). *Effektivitaet psychologischer Therapien von Angst:*

*Meta-analytische Auswertungen auf Stoerungsebene* [Effectiveness of psychological therapies for anxiety disorder: Meta-analytic results for different diagnosis]. Unpublished doctoral thesis, Technische Universitaet Dresden, Dresden, Germany.

Shadish, W. R., Matt, G. E., Navarro, A. M., Siegle, G., Crits-Christoph, P., Hazelrigg, M. D., Jorm, A. F., Lyons, L. C., Nietzel, M. T., Prout, H. T., Robinson, L., Smith, M. L., Svartberg, M., & Weiss, B. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology, 65,* 355–365.

Telch, M. J., Lucas, J. A., Schmidt, N. B., Hanna, H. H., Jaimez, T. L., & Lucas, R. A. (1993). Group cognitive–behavioral treatment of panic disorder. *Behaviour Research and Therapy, 31,* 279–287.

Trull, T. J., Nietzel, M. T., & Main, A. (1988). The use of meta-analysis to assess the clinical significance of behavior therapy for agoraphobia. *Behavior Therapy, 19,* 527–538.

Turner, S. M., Beidel, D. C., Spaulding, S. A., & Brown, J. M. (1995). The practice of behavior therapy: A national survey of cost and methods. *Behavior Therapist, 18,* 1–4.

Tuschen, B., & Fiegenbaum, W. (1997). Techniques of exposure. In W. T. Roth (ed), *Treating anxiety disorders.* (pp. 31–55). San Francisco: Jossey-Bass.

Tuschen-Caffier, B., Pook, M., & Frank, M. (2001). Evaluation of manual-based behavioral therapy for bulimia nervosa in a service setting. *Behaviour Research and Therapy, 39,* 299–308.

Van Balkom, A. J. L. M., Bakker, A., Spinhoven, P., Blaauw, B. M. J. W., Smeenk, S., & Ruesink, B. (1997). A meta-analysis of the treatment of panic disorder with or without agoraphobia: A comparison of psychopharmacological, cognitive-behavioral, and combination treatments. *Journal of Nervous and Mental Disease, 185,* 510–516.

Van Balkom, A. J. L. M., van Oppen, P., Vermeulen, A. W. A., van Dyck, R., Nauta, M. C. E., & Vorst, H. C. M. (1994). A meta-analysis on the treatment of obsessive-compulsive disorder: A comparison of antidepressants, behavior, and cognitive therapy. *Clinical Psychology Review, 14,* 359–381.

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology, 66,* 231–239.

Weiss, B., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67,* 82–94.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 688–701.

Wetzel, C., Bents, H., & Florin, I. (1999). High-density exposure therapy for obsessive-compulsive inpatient: A 1-year follow-up. *Psychotherapy and Psychosomatics, 68,* 186–192.

Wilson, G. T. (1996). Manual-based treatments: The clinical application of research findings. *Behaviour Research and Therapy, 34,* 295–314.

Received April 24, 2000
Revision received October 30, 2000
Accepted November 9, 2000 ■