Clinical Psychology Review 30 (2010) 37–50

Contents lists available at ScienceDirect

Clinical Psychology Review

Psychological treatment of panic disorder with or without agoraphobia: A meta-analysis ☆

Julio Sánchez-Meca *, Ana I. Rosa-Alcázar, Fulgencio Marín-Martínez, Antonia Gómez-Conesa

University of Murcia, Spain

ARTICLE INFO

Article history:
Received 11 September 2008
Received in revised form 24 August 2009
Accepted 25 August 2009

Keywords:
Panic disorder
Agoraphobia
Psychological treatment
Outcome evaluation
Meta-analysis

ABSTRACT

Although the efficacy of psychological treatment for panic disorder (PD) with or without agoraphobia has been the subject of a great deal of research, the specific contribution of techniques such as exposure, cognitive therapy, relaxation training and breathing retraining has not yet been clearly established. This paper presents a meta-analysis applying random- and mixed-effects models to a total of 65 comparisons between a treated and a control group, obtained from 42 studies published between 1980 and 2006. The results showed that, after controlling for the methodological quality of the studies and the type of control group, the combination of exposure, relaxation training, and breathing retraining gives the most consistent evidence for treating PD. Other factors that improve the effectiveness of treatments are the inclusion of homework during the intervention and a follow-up program after it has finished. Furthermore, the treatment is more effective when the patients have no comorbid disorders and the shorter the time they have been suffering from the illness. Publication bias and several methodological factors were discarded as a threat against the validity of our results. Finally the implications of the results for clinical practice and for future research are discussed.

Contents

☆ This research was supported by a grant of the Ministerio de Ciencia y Tecnología of the Spanish Government and by European Union FEDER funds (Project No. BSO2001-0491).
* Corresponding author. Dept. Basic Psychology and Methodology, Faculty of Psychology, Espinardo Campus, University of Murcia, 30100-Murcia, Spain. Tel.: +34 868884114.
E-mail address: jsmeca@um.es (J. Sánchez-Meca).
URL: http://www.um.es/metaanalysis (J. Sánchez-Meca).

0272-7358/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cpr.2009.08.011

1. Introduction

Initially called agoraphobia with panic attacks (American Psychiatric Association, 1980), and later renamed panic disorder (PD) with or without agoraphobia (American Psychiatric Association, 1987, 1994, 2004), PD is one of the most researched anxiety disorders due to its high rate of lifetime prevalence (about 5.1% of adults in USA; Bienvenu, 2006). PD is characterized by its resistance to spontaneous remission, its comorbidity with other disorders (e.g., depression, alcohol or substance disorders), and the decrease in quality of life. Additionally, PD can have serious social and economic consequences, since a large percentage of individuals with PD suffer social isolation and many of them have to give up work (Klerman et al., 1991; Mitte, 2005; Tsao, Mystowski, Zucker, & Craske, 2005).

In order to be diagnosed with PD a patient must have suffered recurrent and unexpected panic attacks over a minimum period of a month, followed by persistent concern about having additional attacks. Panic attacks are commonly accompanied by uncontrollable fear, worry about the implications of the attacks (e.g., losing control, having a heart attack), or a significant change in behavior relating to these symptoms. Furthermore, the attacks are not due to the direct effects of substance abuse or to a medical condition, and they cannot be explained by the presence of another mental illness. On the other hand, panic attacks often come together with agoraphobia, that is, an uncontrollable fear of having a panic attack in a setting from which it may be difficult to escape or receive help. About one in three people with PD develops agoraphobia, but agoraphobia without a history of panic attacks is very uncommon, with a lifetime prevalence of about 0.17% (Bienvenu, 2006).

1.1. The treatment of panic disorder

Since the recognition of PD as a separate diagnostic entity in the *Diagnostic and Statistical Manual of Mental Disorders*, DSM-III-R (American Psychiatric Association, 1980), much research has been devoted to examining the efficacy of different psychological and pharmacological interventions in ameliorating panic symptoms. Particular attention has been paid to cognitive–behavioral and pharmacological type interventions, alone or in combination (Barlow, Gorman, Shear, & Woods, 2000). Prior to 1980, the study of the etiology and treatment of PD was focused on biological theories, which enable the development of pharmacological treatments. Since 1980, the understanding of PD from the psychological perspective has advanced, as has the development of efficacious psychological treatments.

According to the criteria of the *Task Force on Promotion and Dissemination of Psychological Procedures* (1995), and in agreement with Barlow, Raffa, and Cohen (2002), the treatments for PD that have received empirical support are those based on the cognitive–behavioral model. Of particular notability are the panic control treatment developed by Barlow and his colleagues (Barlow & Craske, 1989; Craske & Barlow, 2006) and cognitive therapy by Clark's research group (Clark, 1997; Clark & Salkovskis, 1989).

In the treatment model developed by Barlow's group the exposure of the patient to interoceptive sensations plays a central role. Interoceptive exposure consists of inducing the feared sensations through exercises such as visualization of anxiety scenes, overbreathing and spinning. The treatment includes an educational component which teaches the patient about panic and the factors that influence its origin and recurrence. Cognitive therapy procedures are also included, with the objective of modifying erroneous beliefs about panic and anxiety, as well as cognitions that overestimate the threat and the danger that the attacks represent. The program includes progressive muscle relaxation training, which involves systematically constricting and relaxing various muscle groups paying attention to the sensations as well as suggestions to induce relaxation and warmth. Finally, the program also includes homework exercises, which vary according to the phase of therapy.

The cognitive therapy developed by Clark's group includes both an educational and a cognitive component. As with Barlow's approach, the educational component aims to demystify panic attacks by explaining their causes and triggering mechanisms. The cognitive component helps to identify and challenge the patient's erroneous interpretations of their symptoms. The program includes breathing retraining to influence dysfunctional habitual breathing patterns through the direct or indirect control of respiratory muscles, in order to alleviate fearful sensations. The program also introduces behavioral procedures, such as the generation of feared sensations by carrying out small experiments (e.g., hyper-ventilation, attentional focus, etc.), which have a twofold effect on the patient. Firstly, these exercises show him/her the possible causes of the sensations. Secondly, they help to give up the safety behaviors, disproving any catastrophic thoughts about the consequences of the symptoms. Finally, the program incorporates a series of homework exercises, in addition to a daily record of attacks, negative thoughts and rational interpretations of fearful symptoms.

In practice the most obvious difference between the approaches by Barlow and Clark is that in the former the emphasis is on exposure to interoceptive sensations, while the latter is more focused on the cognitive component.

Other psychological treatments for PD have been examined, but have not provided such clear benefits in terms of a statistically significant reduction of panic and agoraphobia symptoms. These include 'Eye Movement Desensitization and Reprocessing' (EMDR; Feske & Goldstein, 1997; Goldstein, de Beurs, Chambless, & Wilson, 2000), emotion regulation therapy (Shear, Houck, Greeno, & Masters, 2001), and Gestalt therapy (Chambless, Goldstein, Gallagher, & Bright, 1986).

PD with or without agoraphobia has been the focus of various meta-analytic studies to examine the differential efficacy of psychological and/or pharmacological interventions (Bakker et al., 1998; Chambless & Gillis, 1993; Clum, Clum, & Surls, 1993; Cox, Endler, Lee, & Swinson, 1992; Gould, Otto, & Pollack, 1995; Mattick, Andrews, Hadzi-Pavlovic, & Christensen, 1990; Mitte, 2005; Oei et al., 1999; Trull, Nietzel, & Main, 1988; van Balkom et al., 1997; van Balkom, Nauta, & Bakker, 1995; Westen & Morrison, 2001; Wilkinson, Balestrieri, Ruggeri, & Bellantuono, 1991). The results of these studies clearly prove the efficacy of cognitive therapy, *in vivo* exposure, and both techniques combined. *In vivo* exposure is a cognitive–behavioral technique consisting of gradually exposing the patient to feared situations. There is evidence that the main component in treating PD is *in vivo* exposure, with an effect size ranging between $d = 0.78$ and $d = 1.34$ in terms of the standardized mean difference. Furthermore, the effects increase over the course of time (between $d = 1.09$ and $d = 1.53$), although the follow-up periods were short, not exceeding 12 months on average. Meta-analyses that have addressed the differential efficacy of psychological and pharmacological treatments have shown good results for both cognitive–behavioral and pharmacological interventions, alone or in combination (cf. e.g., Cox et al., 1992; Mitte, 2005; van Balkom et al., 1997; Wilkinson et al., 1991).

Other relevant treatment characteristics have been empirically examined. One of these is the application format of the therapy,

distinguishing between individual and group treatment. Sharp, Power, and Swanson (2004) showed that both individual and group cognitive–behavioral therapies (CBT) were clearly superior to a non-active control group, but did not differ significantly from each other. Another important therapeutic format refers to the extent of therapist assistance. In an attempt to provide cost-effective treatment, Carlbring, Ekselius, and Anderson (2003) obtained good results for Internet-based CBT with minimal contact via e-mail. On the other hand, Klein, Richards, and Austin (2006) found better results for CBT via the Internet than CBT manual in reducing clinical-rated agoraphobia and improving physical health rating.

1.2. Objectives of the study

The purpose of our study was to analyze the efficacy of psychological interventions in the treatment of PD with or without agoraphobia. We were also interested in identifying treatment, subject, methodological, and extrinsic characteristics that can influence the effect magnitude obtained in the studies. We included several new elements in our meta-analysis with regard to the previous ones. First, we made it a pre-requisite that the studies included a control group. This is a characteristic that our meta-analysis shares only with those of Clum, Clum et al., 1993 and Gould et al. (1995). Studies with a control group are less vulnerable to threats to internal validity than those that do not include one and, at the same time, this requisite allows the use of the standardized mean difference as the effect size index. Second, we updated the time period of the search, to the year 2006 inclusively. The most recent meta-analysis (Mitte, 2005) only includes up to the year 2004. Third, in our meta-analysis random- and mixed-effects models were applied, which are more appropriate than the fixed-effects models usually applied in the previous meta-analyses (with the exception of Mitte, 2005).[1] Fourth, we included as a methodological moderator variable the standardized mean difference, *d*, obtained in the pretest when comparing the means of the treated and control groups. Thus, we were able to analyze the possible influence of the *d* index in the pretest on the effect size in the posttest. Finally, we propose a predictive model for the differential efficacy of the different techniques of psychological intervention, controlling for the methodological quality of the studies, which can be used in clinical practice and future research in this field.

2. Method

2.1. Selection criteria of the studies

In order to be included in our meta-analysis, the studies had to fulfill several selection criteria. First, the paper had to be an empirical study in which a psychological treatment was applied to a sample of subjects diagnosed with PD with or without agoraphobia via a diagnostic criterion recognized by the scientific community (any version of the DSM, the International Classification of Diseases and Related Health Problems, ICD, or any other conventional and standardized classification). Second, studies that combined psychological treatment with psychoactive drugs were excluded, as our interest was focused on the benefits of different components of the psychological treatments. Third, studies whose samples were composed of patients with agoraphobia without panic attacks were also excluded, as our interest was in patients with PD. Fourth, the study had to include a control group, either non-active (waiting list) or active (psychological placebo, pharmacological placebo or both). Fifth, the study had to report statistical data from the groups involved (means, standard deviations, *t*-tests, ANOVAs, etc.) in the posttest and, optionally, in the pretest and in any follow-up. Sixth, the sample size of every group in the posttest could not be less than five subjects. Seventh, the study had to be carried out or published between 1980 and 2006. Finally, due to limitations in the languages spoken by the authors of this research, the study had to be written in English, Spanish or French.

2.2. Search procedures

To select the studies that could fulfill our selection criteria the following databases were consulted in June 2007: PsycINFO, Medline, the Cochrane Library, and the Spanish databases CSIC and Psicodoc. In addition, books, monographs and scientific journals were consulted, as well as the references of 12 meta-analyses published on the efficacy of psychological interventions for PD (Bakker et al., 1998; Chambless & Gillis, 1993; Clum, Clum et al., 1993; Cox et al., 1992; Gould et al., 1995; Mattick et al., 1990; Mitte, 2005; Oei et al., 1999; Trull et al., 1988; van Balkom et al., 1995, 1997; Westen & Morrison, 2001). Also, in an attempt to recover unpublished papers, letters were sent to researchers in the field.

In order to locate electronically the studies that fulfilled the selection criteria, abstracts were read of all the studies that included the following combination of key words in the title of the article: ("treatment*" or "therap*" or "behavio*" or "cognitive" or "program*" or "intervent*") and ("panic" or "agoraphob*"). From all of the electronic databases, this combination of key words produced 2500 references. Forty-two articles that fulfilled our selection criteria were selected definitively. As some of the studies included more than one treatment group, a total of 65 independent comparisons between a treatment and a control group were included in the meta-analysis. The total sample size involved 2560 individuals in the pretest assessments (1712 in the 65 treatment groups and 848 in the 42 control groups), which reduced to 2357 subjects in the posttest (1559 in the treatment groups and 798 subjects in the control groups; median sample size = 17 subjects). Although an attempt was made to locate unpublished works, all those selected were published papers. The studies selected came from four continents. North America was the most represented, with 37 comparisons (53.6%; 35 from the USA and 2 from Canada). Europe had the next largest representation, with 23 comparisons (33.3%; Great Britain being the country most represented, with 13 comparisons). This was followed by Australia (6 comparisons, 8.7%) and South America (3 comparisons from Brazil, 4.3%).[2] All of the studies were written in English.

The degree of overlap of the studies included in our meta-analysis with respect to the meta-analyses previously published on the efficacy of psychological treatments for PD ranged between 0% (Chambless & Gillis, 1993; van Balkom et al., 1995), and 45% (Mitte,

[1] The consequences of assuming a random-effects model instead of a fixed-effects one concern the interpretation of the results and also the actual results obtained. A meta-analyst that applies a fixed-effects model is assuming that his/her results can only be generalized to an identical population of studies to that of the individual studies included in the meta-analysis. In a random-effects model however, the results can be generalized to a wider population of studies. On the other hand, in a fixed-effects model the error attributed to the effect size estimates is smaller than in a random-effects model, which is why in the first model the confidence intervals are narrower and the statistical tests more liberal than in the second one. The principal consequence of assuming a fixed-effects model when the meta-analytic data come from a random-effects model is that we may attribute more precision to the effect size estimates than is appropriate and that, in addition, we may find statistically significant relationships between variables that are actually spurious (*cf.* Field, 2003; Hedges & Vevea, 1998; Marín-Martínez & Sánchez-Meca, 1998; National Research Council, 1992; Sánchez-Meca & Marín-Martínez, 1998, 2008).

[2] The total number of comparisons computed here rises to 69 instead of 65 because 4 comparisons were carried out with samples of subjects from Australia and the United Kingdom (Casey, Newcombe, & Oei, 2005; Kenardy et al., 2003, studies a, b and c).

2005). Together with Mitte's (2005) meta-analysis, the others that showed greater overlap with our study were those of van Balkom et al. (1997, with an overlap of 26%), Gould et al. (1995, with 19%), Bakker et al. (1998, with 12%, although this meta-analysis is actually a continuation of that of van Balkom et al. (1997)), and Westen and Morrison (2001, with 11.9%). These data demonstrate the minor overlap between our meta-analysis and previous studies, and guarantee the originality of our results.

2.3. Coding of studies

In order to check the characteristics of the studies that can be correlated to the effect magnitude, treatment, subject, methodological, and extrinsic variables of the studies were coded.

The *treatment characteristics* coded were: (a) the type of intervention received by the treatment group, distinguishing between relaxation training techniques (e.g., progressive relaxation), breathing retraining techniques (e.g., deep breathing, breathing retraining), exposure (*in vivo*, in imagination, mixed), systematic desensitization (e.g., eye movement desensitization reprocessing, EMDR),[3] cognitive therapy, anxiety management training, and other techniques; (b) the duration of the treatment (in weeks); (c) the intensity of the treatment (number of weekly hours of treatment); (d) the magnitude of the treatment (total number of hours received by each subject); (e) the number of sessions of treatment; (f) the homogeneity of the treatment (whether all patients received the treatment in the same conditions); (g) the inclusion of homework; (h) the inclusion of a follow-up program; (i) the mode of training (group, individual or mixed); (j) the therapeutic format (direct, written or mixed), (k) the mode of application of the intervention (therapist-assisted versus self-help treatment), and (l) the experience of the therapists.

The *subject characteristics* coded for the samples of each study were: (a) the mean age of the sample (in years); (b) the gender of the sample (percentage of males); (c) the type of disorder suffered by the subjects (PD with agoraphobia, PD or mixed); (d) the illness duration (mean in years of the group); (e) the presence of comorbidity in the sample (percentage of subjects), and (f) whether individuals had received any previous treatment.

The *methodological characteristics* were coded as follows: (a) design type (quasi-experimental versus experimental); (b) type of control group (non-active versus active); (c) control of consumption of psychoactive drugs by the subjects during the treatment; (d) percentage of patients in the sample that took psychoactive drugs;[4] (e) attrition in the posttest for the treatment and control groups, as well as differential attrition between the two groups; (f) follow-up measures (in months); (g) quality of the study (on a scale of 0 — minimal quality — to 9 points — maximum quality),[5] and (h) the *d* index in the pretest.

Finally, the *extrinsic characteristics* coded were: (a) the year of publication of the study; (b) the discipline of the first author (psychologist, psychiatrist, other), and (c) the country and continent in which the study was carried out.

With the aim of assuring the highest possible objectivity, a codebook was produced, which contained the details of the criteria used in coding all of the characteristics of the studies. In order to check the reliability of the coding process a random sample of the meta-analyzed studies was selected (20% of the total) and two teams of researchers independently coded the subset of selected studies. Each coding team was composed of two researchers who coded each study independently and inconsistencies were then resolved by consensus. The coding reliability reached kappa coefficients which were highly satisfactory on the whole: in all cases they were greater than 0.60 (mean kappa coefficient = 0.78). The effect size calculations were also subjected to reliability analysis. They produced, on average, an intra-class correlation coefficient of 0.85 between the estimates obtained by the two coding teams. Both the codebook and the manual for calculating the effect size can be obtained from the corresponding author.

2.4. Computation of effect size

The standardized mean difference, *d*, was used as the effect size index. *d* is defined as the difference between the means of the treatment and the control groups, both in the posttest, divided by a pooled estimate of the within-study standard deviation, *S*, and corrected by the factor $c(m)$ for small samples (Hedges and Olkin, 1985): $d = c(m)(\overline{y}_E - \overline{y}_C)/S$. Positive values of *d* indicated a favorable result to the treatment. When the study did not report the means or standard deviations of the groups, the procedures developed by Glass, McGaw, and Smith (1981) were used to calculate the *d* index from the results of *t*-tests, ANOVAs, etc.

In each study a *d* index was calculated for each of the outcome measures of panic, agoraphobia, general anxiety, depression, fear of bodily sensations, global adjustment, and other measures and, in turn, for two types of report: self-reports and clinician assessments. Therefore, in the same study up to 14 *d* indices could be calculated (from 7 outcome measures × 2 types of report). In addition to these, a mean *d* index was also calculated for each of the two types of report and for each of the seven outcome measures (pooling self-reports and clinical assessments). Finally, the mean *d* index of all the types of reports and measures was also obtained. Therefore this made a total of 24 *d* indices calculated in the posttest for each study.

The standardized mean difference between the two groups in the pretest was also calculated in order to use this index as a moderator variable, which would enable us to verify if it is related to the *d* index in the posttest. Therefore, in each study that reported data from the pretest, a total of up to 24 *d* indices was calculated in the pretest.

A review of the measurement instruments applied in the studies of this meta-analysis revealed that those most frequently used to assess panic measures were the *Panic Attack Symptoms Questionnaire* (PASQ; Clum, Broyles, Borden, & Watkins, 1990) and the *Panic Appraisal Inventory* (PAI; Telch et al., 1993). The most frequently used instruments in assessing agoraphobic symptoms were the Agoraphobia subscale of the *Fear Questionnaire* (Marks & Mathews, 1979), the *Agoraphobic Cognitions Questionnaire* (ACQ; Chambless, Caputo, Bright, & Gallagher, 1984), and the *Mobility Inventory for Agoraphobia* (Chambless et al., 1985). To assess general anxiety the following were used: the *State-Trait Anxiety Inventory* (STAI; Spielberger, Gorsuch, & Lushene, 1970), the *Hamilton Anxiety Scale* (HAM-A; Hamilton, 1959), the *Anxiety Sensivity Index* (ASI; Reiss, Peterson, Gursky, & McNally, 1986), and the *Beck Anxiety Inventory* (BAI; Beck & Steer, 1987). Depression symptoms were mainly assessed with the *Beck Depression Inventory* (BDI; Beck et al., 1961). The *Bodily Sensations Questionnaire* (BSQ; Chambless et al., 1984) was used to assess fear of bodily

[3] Following its originator (Shapiro, 1995), we have classified EMDR as a type of systematic desensitization. In any case, only three studies in our meta-analysis applied the EMDR technique (Feske & Goldstein, 1997, studies a and b; Goldstein, de Beurs, Chambless, & Wilson, 2000).

[4] Although we focused on studies that applied treatments that were composed exclusively of psychological components, many of the patients included in these studies were taking psychoactive drugs at controlled and/or reduced doses. The moderator variables (c) and (d) enabled us to code whether the researchers had controlled the patients' consumption of psychotropic substances under psychiatric prescription as well as the percentage of patients that were consuming psychoactive drugs during the treatment at reduced and/or controlled doses.

[5] The scale of quality takes into consideration random assignment, sample size, the use of pretest measures, attrition, the use of blind evaluators, the reporting in the posttest of all the variables recorded in the pretest, the homogeneity of the treatment, the reporting of follow-up measures, and the use of normed and standardized assessment instruments. The intra-class correlation between two independent coding teams was 0.72. The scale can be requested from the corresponding author.

sensations. Other measurement instruments aimed to assess the global adjustment of the individual, such as the *Global Assessment of Severity* (GAS) and the *Social Adjustment Scale* (SAS; Weissman & Bothwell, 1976). Finally, in the category 'other measures' we included instruments that could not be classified in the previous categories because they assessed psychological constructs such as quality of life, general satisfaction, or satisfaction in specific areas (social, marital, and work). The most frequent instruments in this category were the *Symptom Checklist-90* (SCL-90; Derogatis, Lipman, & Covi, 1973), the *Clinical Global Impression Scale* (CGI-S; Guy, 1976), and the *Quality of Life Inventory* (QOLI; Frisch, Cornell, Villanueva, & Retzlaff, 1992).

2.5. Statistical analysis

To avoid problems of statistical dependence, separate meta-analyses were carried out for each *d* index according to the outcome measure and type of report. In each one a random-effects model was applied according to which each *d* index was weighted by its inverse variance (the sum of the within-study variance and an estimate of the between-studies variance). The process of analysis consisted of calculating the mean effect size with its 95% confidence interval, the heterogeneity test, *Q*, and the I^2 index to assess the degree of heterogeneity of the effect sizes around the mean effect (Cooper, Hedges, & Valentine, 2009; Hedges & Olkin, 1985; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). Once confirmed that the effect sizes were heterogeneous, mixed-effects models were applied to test the influence of the moderator variables. These consisted of ANOVAs and simple and multiple regression analyses by weighted least squares. Finally, since our meta-analysis did not include any unpublished papers, a test for publication bias was carried out (Rothstein, Sutton, & Borenstein, 2005). All the statistical analyses were carried out using the SPSS macros created by David B. Wilson.[6]

3. Results

3.1. Distribution of effect sizes

A list of the 65 studies with the main moderator variables and effect sizes is presented in Appendix A. For each combination of outcome measures (panic, agoraphobia, anxiety, depression, bodily sensations, global adjustment and others) and type of instrument (self-reports, clinicians and combination), we carried out a meta-analysis to obtain an estimate of the mean effect size together with its confidence interval, the heterogeneity *Q* statistic and the I^2 index. Additionally, separate meta-analyses were applied for self-reports, for clinician assessments, and for the total combination of outcome measures and report types. Table 1 shows the results. As we can see in this table, the global result for the 65 comparisons gave a statistically significant mean effect size, $d_+ = 0.784$ (95% CI: 0.663, 0.905), and according to the classification proposed by Cohen (1988) this can be considered to be of high magnitude in favor of psychological treatments of PD.

As was to be expected, the best results were obtained with the panic measures, $d_+ = 1.015$ (95% CI: 0.855, 1.175), followed by those of global adjustment, $d_+ = 0.895$ (95% CI: 0.665, 1.126), bodily sensations, $d_+ = 0.874$ (95% CI: 0.656, 1.092), agoraphobia, $d_+ = 0.856$ (95% CI: 0.679, 1.033), general anxiety, $d_+ = 0.840$ (95% CI: 0.686, 0.994), depression, $d_+ = 0.645$ (95% CI: 0.500, 0.791), and other measures, $d_+ = 0.627$ (95% CI: 0.446, 0.808). Self-reports gave slightly lower effect sizes than clinician assessments in the panic measures (self-reports: $d_+ = 1.037$; clinicians: $d_+ = 1.182$), and more prominent ones in the measures of agoraphobia (self-reports: $d_+ = 0.784$; clinicians: $d_+ = 1.961$) and general anxiety (self-reports: $d_+ = 0.773$; clinicians: $d_+ = 1.128$). On the other hand, the effect sizes for self-reports were slightly higher than those of clinicians for measures of depression (self-reports: $d_+ = 0.689$; clinicians: $d_+ = 0.545$), global adjustment (self-reports: $d_+ = 0.919$; clinicians: $d_+ = 0.840$), and other measures (self-reports: $d_+ = 0.644$; clinicians: $d_+ = 0.586$).

Table 1
Summary results for the effect size as a function of the outcome measure and type of measurement instrument.

Outcome/report type	*k*	d_+ (95% C.I.)	*Q*	I^2
Panic:				
Self-reports	50	1.037 (0.848; 1.227)	180.78**	72.9
Clinician	25	1.182 (0.924; 1.441)	86.44**	72.2
Combined	61	1.015 (0.855; 1.175)	202.77**	70.4
Agoraphobia:				
Self-reports	40	0.784 (0.621; 0.948)	93.81**	58.4
Clinician	7	1.961 (1.608; 2.315)	8.48	29.2
Combined	42	0.856 (0.679; 1.033)	118.03**	65.3
General anxiety:				
Self-reports	41	0.773 (0.609; 0.936)	98.90**	59.6
Clinician	18	1.128 (0.949; 1.307)	20.44	16.8
Combined	44	0.840 (0.686; 0.994)	99.59**	56.8
Depression:				
Self-reports	35	0.689 (0.539; 0.840)	50.19*	32.3
Clinician	12	0.545 (0.270; 0.820)	24.23*	54.6
Combined	42	0.645 (0.500; 0.791)	72.08**	43.1
Bodily sensations (self-reports only)	18	0.874 (0.656; 1.092)	33.04**	48.5
Global adjustment:				
Self-reports	20	0.919 (0.648; 1.189)	75.33**	74.8
Clinician	10	0.840 (0.481; 1.200)	31.05**	71.0
Combined	25	0.895 (0.665; 1.126)	87.03**	72.4
Other outcomes:				
Self-reports	21	0.644 (0.450; 0.838)	38.99**	48.7
Clinician	3	0.586 (−0.051; 1.223)	5.25	61.9
Combined	24	0.627 (0.446; 0.808)	44.09**	47.8
Global results:				
Self-reports	59	0.811 (0.686; 0.936)	117.27**	50.5
Clinician	34	1.080 (0.864; 1.296)	115.50**	71.4
Total (self-reports + clinician)	65	0.784 (0.663; 0.905)	136.91**	53.3

k: number of studies. d_+: weighted mean effect size. 95% C.I.: 95 per cent confidence interval around the mean effect size. *Q*: heterogeneity *Q* statistic. I^2: I^2 heterogeneity index (%).
* $p < .05$.
** $p < .01$.

Of all the meta-analyses reported in Table 1, there was a high heterogeneity obtained between the individual effect sizes of the studies. This is reflected by the fact that statistically significant *Q* values ($p < .05$) and I^2 indices of at least 50% were obtained in most of the outcome measures (see Table 1). The existence of high heterogeneity between the effect sizes led us to examine the influence of characteristics of the studies that may explain part of this heterogeneity. Before presenting the results of these analyses however, a study of publication bias is outlined in the next section.

3.2. Publication bias

Since all the studies included in the meta-analysis were published papers, we tested whether publication bias against null results could be a source of bias in the effect size estimates obtained in our meta-analysis. To do this, we carried out two complementary analyses on the 65 global effect sizes resulting from combining all the outcome

[6] These macros can be obtained from the following web address: http://mason.gmu.edu/~dwilsonb/ma.html.

measures and all the report types. First, we applied the Egger test,[7] obtaining a non-statistically significant result for the intercept of the regression model [Intercept = 0.178 ; $t(63) = 0.242$, $p = .810$]. Second, we calculated the fail-safe N index, N_{fs} (Becker, 2005), obtaining an N_{fs} value of 304, which means that to cancel the mean effect size obtained in our meta-analysis ($d_+ = 0.784$) there had to be 304 non-published studies with null effects not included in the meta-analysis.[8] Therefore, in the light of these analyses it seems reasonable to discard publication bias as a serious threat to the validity of our meta-analytic results.

3.3. Type of treatment

Of the 65 comparisons between a treatment group and a control group, only two incorporated an external component to the behavioral or cognitive–behavioral approaches. In particular, Shear et al. (2001, study a) applied a treatment based on emotion regulation therapy, and Chambless et al. (1986) combined Gestalt therapy with exposure, cognitive therapy, and breathing retraining techniques. The most frequent psychological techniques were exposure (55 comparisons out of the total 65, or 84.6%), followed by cognitive therapy (46 comparisons, 70.8%), and relaxation training and/or breathing retraining techniques (30 comparisons, 46.2%). Of the various types of exposure, the most frequent were those based on *in vivo* exposure, both interoceptive (29 comparisons, 44.6%) and non interoceptive (28 comparisons, 43.1%). The use of imaginal exposure was very infrequent (interoceptive: 6 comparisons; non interoceptive: 6 comparisons).

Analysis of the moderator variables was carried out only on the effect sizes obtained with panic measures (self-reports and clinician assessments combined), since this is the outcome measure most frequently used in the studies whilst also being the most relevant from a clinical basis. Of the 65 comparisons, 4 did not give data on panic measures (Barlow et al., 1984; Michelson & Mavissakalian, 1985; Sharp, Powell, & Swanson, 2004, studies a and b), so the following analyses are based on 61 comparisons.

As the most frequent techniques were exposure, cognitive therapy (CT), and relaxation training and/or breathing retraining techniques (RB), and given that the majority of the studies applied a combination of treatments using these and other techniques, we carried out an ad hoc categorization. The treatments were classified as follows: RB alone, exposure alone, CT alone, or any combination of these (RB + exposure, RB + CT, exposure + CT, and RB + exposure + CT). Furthermore, we added the categories EMDR and "other techniques"[9] to include those that would not fit in any of the previous ones. Table 2 shows the mixed-effects ANOVA of the type of treatment on the effect sizes for panic measures, which showed statistically significant differences between the nine treatment combinations tested, with 24.6% of variance accounted for [$Q_B(8) = 33.311$, $p < .001$; $\omega^2 = 0.246$]. Fig. 1 shows a forest plot with the mean effect sizes for the different treatment combinations in panic measures. According to these results, the most efficacious combined treatment was RB + exposure ($d_+ = 1.837$), followed by exposure alone ($d_+ = 1.528$), exposure + CT ($d_+ = 1.285$), RB alone ($d_+ = 0.862$),[10] and RB + exposure + CT ($d_+ = 0.833$). With the exception of RB alone, the results were clearly lower when the treatment did not include exposure. Non-statistically significant confidence intervals were even obtained (RB + CT: $d_+ = 0.697$, although with only one study; EMDR alone: $d_+ = 0.613$, with 3 studies, CT alone: $d_+ = 0.338$, with 3 studies; other techniques: $d_+ = -0.020$, with 2 studies).

Table 2
Results of comparing different treatment combinations on the effect size for panic and agoraphobia measures.

Moderator variable/ outcome measure	k	d_+	95% C. I.		ANOVA results
			d_l	d_u	
Treatment combination/panic measures:					
Relaxation/breathing (RB)	3	0.862	0.203	1.522	$Q_B(8) = 33.311$, $p < .001$
Exposure	4	1.528	0.928	2.128	$Q_E(52) = 61.939$, $p = .163$
Cognitive therapy (CT)	3	0.338	−0.253	0.930	$\omega^2 = 0.246$
EMDR	3	0.613	−0.023	1.249	
RB + exposure	4	1.837	1.265	2.409	
RB + CT	1	0.697	-0.278	1.672	
Exposure + CT	19	1.285	1.042	1.528	
RB + Exposure + CT	22	0.833	0.611	1.055	
Other techniques	2	−0.020	−0.732	0.692	
Type of exposure/panic measures:					
In vivo exposure	39	1.246	1.053	1.438	$Q_B(2) = 6.890$, $p = .032$
Exposure in imagination	6	0.646	0.128	1.165	$Q_E(47) = 57.503$, $p = .140$
Mixed exposure	5	0.751	0.246	1.256	$\omega^2 = 0.068$
Treatment combination/agoraphobia measures:					
Relaxation/breathing (RB)	1	0.150	−0.867	1.167	$Q_B(8) = 10.848$, $p = .210$
Exposure	1	0.702	−0.407	1.811	$Q_E(33) = 36.910$ $p = .293$
Cognitive therapy (CT)	2	−0.130	−0.970	0.709	$\omega^2 = 0.039$
EMDR	1	0.721	−0.524	1.966	
RB + exposure	2	1.034	0.205	1.864	
RB + CT	2	1.600	0.725	2.474	
Exposure + CT	16	0.907	0.620	1.194	
RB + Exposure + CT	15	0.899	0.611	1.187	
Other techniques	2	0.562	−0.323	1.448	
Type of exposure/agoraphobia measures:					
In vivo exposure	28	0.894	0.677	1.110	$Q_B(2) = 0.467$, $p = .792$
Exposure in imagination	2	0.610	−0.196	1.416	$Q_E(31) = 35.544$, $p = .263$
Mixed exposure	4	0.919	0.400	1.439	$\omega^2 = 0.0$

k: number of studies. d_+: mean effect size. 95% C.I.: 95% confidence interval. d_l and d_u: lower and upper confidence limits. Q_B: between-categories Q statistic. p: probability level. Q_E: within-categories Q statistic. ω^2: proportion of variance accounted for.

Out of the 61 comparisons with panic measures, 50 applied one type of exposure (*in vivo*, in imagination or mixed). As we can see in Table 2, the comparison of these three categories referring to the type of exposure gave statistically significant results, with 6.8% of variance accounted for. *In vivo* exposure achieved the best results ($d_+ = 1.246$), followed by mixed exposure ($d_+ = 0.751$).

Although our analyses focused on panic measures, in Table 2 we have also included comparisons between the different combined treatments with regard to agoraphobia measures. In this case, we did not obtain statistically significant differences between them

[7] The "Egger test" is an unweighted regression consisting of taking the precision of each study as the independent variable (precision being defined as the inverse of the standard error of each effect size) and the effect size divided by its standard error as the dependent variable. A non-statistically significant result of the t-test for the hypothesis of an intercept equal to zero enables us to discard publication bias as a threat to the validity of our overall effect size (Sterne & Egger, 2005). Non-statistically significant results for the intercept were obtained when the Egger test was applied to the other outcome measures ($p > .05$), with the exception of panic measures ($p = .029$).

[8] The N_{fs} index for each outcome measure was also clearly high, enabling us to discard publication bias as a threat to the validity of our results (panic measures: $N_{fs} = 511$ studies; agoraphobia measures: $N_{fs} = 240$; general anxiety: $N_{fs} = 241$; depression: $N_{fs} = 125$; bodily sensations: $N_{fs} = 108$; global adjustment: $N_{fs} = 158$; other measures: $N_{fs} = 67$).

[9] In the category 'other techniques' only two studies were included: one that applied a combination of exposure and systematic desensitization (Mavissakalian & Michelson, 1986) and another that applied emotion regulation therapy (Shear et al., 2001, study a).

[10] Only three studies applied relaxation and/or breathing training techniques alone: two of them applied relaxation training (Barlow, Craske, Cerny, & Klosko, 1989, study a; Taylor, Kenigsberg, & Robinson, 1982) and the other one applied both relaxation and breathing training (Beck et al., 1994, study b).

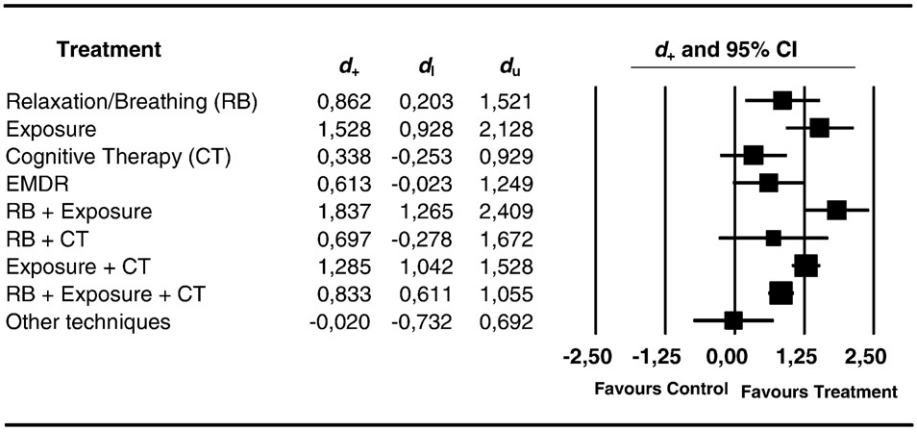

Fig. 1. Forest plot of the mean effect sizes for the different treatment combinations on panic measures. d_+: mean effect size. d_l and d_u: lower and upper confidence limits.

($p = .210$). The highest effect sizes were obtained when any of the three most frequent techniques (RB, exposure, and CT) were combined with each other: RB + CT ($d_+ = 1.600$), RB + exposure ($d_+ = 1.034$), exposure + CT ($d_+ = 0.907$), and RB + exposure + CT ($d_+ = 0.899$). When the techniques were applied separately they did not obtain statistically significant results: exposure alone ($d_+ = 0.702$), RB alone ($d_+ = 0.150$), and CT alone ($d_+ = -0.130$). The categories EMDR and 'other techniques' did not reach statistical significance either (EMDR: $d_+ = 0.721$, with only one study; other techniques: $d_+ = 0.562$). The comparison between the different types of exposure did not achieve a statistically significant result either ($p = .792$), although imaginal exposure showed a lower effect size ($d_+ = 0.610$) in comparison with *in vivo* exposure ($d_+ = 0.894$) and combined exposure ($d_+ = 0.919$).

3.4. Other treatment characteristics

Table 3 shows the results of the ANOVAs for other qualitative moderator variables related to the characteristics of the treatment on panic measures. The only moderator variables that had a statistically significant association with effect size are detailed as follows. First, the interventions that included homework ($d_+ = 1.183$) obtained better results than those that did not ($d_+ = 0.569$) ($p = .002$). Second, a moderator variable that had a marginally significant association with effect size was the inclusion of a follow-up program ($p = .076$), with better results obtained for the interventions that did include one ($d_+ = 1.208$) with respect to those that did not ($d_+ = 0.907$). Another treatment characteristic that showed a marginally significant relationship with effect size was the type of intervention ($p = .083$), with

Table 3
Results of analyzing the influence of qualitative moderator variables related with the treatment implementation on the effect sizes for panic measures.

Moderator variable	k	d_+	95% C. I.		ANOVA results
			d_l	d_u	
Homogeneity of the treatment:					$Q_B(1) = 0.394$, $p = .530$
Homogeneous treatment	57	1.030	0.862	1.197	$Q_W(59) = 70.701$, $p = .141$
Nonhomogeneous treatment	4	0.823	0.201	1.446	$\omega^2 = 0.0$
Homework?					$Q_B(1) = 9.720$, $p = .002$
Yes	44	1.183	1.002	1.364	$Q_W(53) = 64.514$, $p = .133$
No	11	0.569	0.228	0.910	$\omega^2 = 0.113$
Follow-up program?					$Q_B(1) = 3.145$, $p = .076$
Yes	22	1.208	0.941	1.475	$Q_W(59) = 69.002$, $p = .175$
No	39	0.907	0.707	1.106	$\omega^2 = 0.027$
Approach of the intervention:					$Q_B(1) = 1.615$ $p = .204$
Self-help treatment	8	0.745	0.299	1.191	$Q_W(59) = 70.841$, $p = .139$
Therapist-assisted treatment	53	1.054	0.883	1.225	$\omega^2 = 0.006$
Type of intervention:					$Q_B(2) = 4.981$, $p = .083$
Group intervention	13	1.008	0.668	1.348	$Q_W(56) = 65.466$, $p = .181$
Individual intervention	41	1.104	0.905	1.302	$\omega^2 = 0.037$
Mixed intervention	5	0.413	−0.161	0.987	
Therapeutic format:					$Q_B(2) = 1.411$, $p = .494$
Direct (oral or by e-mail)	48	1.021	0.839	1.202	$Q_W(58) = 69.510$, $p = .143$
Written	5	0.735	0.175	1.295	$\omega^2 = 0.0$
Mixed	8	1.172	0.712	1.631	
Therapist experience:					$Q_B(3) = 4.456$, $p = .216$
High	28	1.250	1.003	1.497	$Q_W(47) = 56.582$, $p = .160$
Medium	12	0.908	0.537	1.280	$\omega^2 = 0.013$
Low	10	0.804	0.373	1.236	
Mixed	1	0.757	−0.538	2.052	

k: number of studies. d_+: mean effect size. 95% C.I.: 95% confidence interval. d_l and d_u: lower and upper confidence limits. Q_B: between-categories Q statistic. p: probability level. Q_W: within-categories Q statistic. ω^2: proportion of variance accounted for.

better results for individual ($d_+ = 1.104$) and group ($d_+ = 1.008$) interventions in comparison with mixed interventions ($d_+ = 0.413$). Finally, Table 5 shows the mixed-effects regression analyses applied with each continuous moderator variable on the effect sizes for panic measures. In the case of the treatment characteristics, none of the moderator variables analyzed showed a statistically significant association ($p > .05$): duration (mean = 9.5 weeks, $SD = 3.7$), intensity (mean = 2.2 h per week, $SD = 5.5$), magnitude of treatment (mean = 12.4 total hours per subject, $SD = 11.1$) and number of sessions (mean = 10.2 sessions, $SD = 3.6$).

3.5. Subject characteristics

Table 4 shows the ANOVAs applied with the subject moderator variables on panic measures. Firstly, the interventions obtained larger effect sizes when the subject samples were composed of patients with PD with agoraphobia ($d_+ = 1.376$) than without agoraphobia ($d_+ = 0.430$) ($p = .004$). Since the majority of the subject samples mixed patients with and without agoraphobia, we coded as a continuous moderator variable the percentage of individuals with agoraphobia in the sample and applied a regression model on the effect sizes. As we can see in Table 5, from the 46 comparisons that reported the percentage of patients with agoraphobia, a positive and statistically significant association with effect size was reached, with 20% of variance accounted for. This means that the higher the percentage of patients with agoraphobia the larger the effect size. Previous studies have shown the opposite result (e.g., Rosenberg & Hougaard, 2005; Williams & Falbo, 1996). Thus, in order to analyze this relationship in depth, we examined the relationship between the percentages of individuals with agoraphobia in the sample with two methodological variables of the studies: the design quality and the type of control group. We did not find a statistical relationship between the percentage of patients with agoraphobia and the design quality of the study ($r = 0.082$, $p = .581$). The studies with a non-active control group however, had a higher percentage of patients with agoraphobia (mean: 80.5%) than the studies with active control groups (mean: 40.6%), the difference between these two mean percentages being statistically significant [$t(44) = 3.540$, $p = .001$]. As we will mention later (see Table 4), the studies with non-active control groups obtained larger effect sizes than those with active control groups. Therefore, the positive statistical relationship found between the percentage of individuals with agoraphobia and the effect size was spurious as it was due to the relationship between the percentage of patients with agoraphobia and the type of control group. In fact, when we repeated the regression analysis for the percentage of individuals with agoraphobia on the effect size separately for the studies with non-active and active control groups, the statistical relationship disappeared [non-active control groups: $Z = 1.752$, $p = .079$, $R^2 = 0.079$, $k = 33$ studies; Active control groups: $Z = 1.128$, $p = .259$, $R^2 = 0.107$, $k = 13$ studies]. Thus, contrary to previous research, we found a null relationship, and not a negative one, between the presence of agoraphobia in the patients and the effect size. There remained the possibility however, that the null relationship between the percentage of individuals with agoraphobia and the effect size might be due to the fact that the effect size was based on panic measures and not on agoraphobia outcomes. Thus, we applied a regression model to test whether the percentage of individuals with agoraphobia was associated with the effect size for agoraphobia measures. Once again, we did not find a statistical

Table 5
Simple weighted regression analyses of each continuous moderator variable on the *d* index for panic measures.

Moderator variable	*k*	*B*	*Z*	Q_E	R^2
(A) Treatment characteristics:					
Duration (in weeks)	59	0.011	0.468	68.226	0.003
Intensity (*n.* hours a week)	48	−0.007	−0.435	57.609	0.003
Magnitude (total n. of hours)	50	−0.007	−0.906	60.392	0.013
Number of sessions	54	−0.038	−1.548	63.091	0.037
(B) Subject characteristics:					
Mean age (in years)	59	0.000	−0.001	69.701	0.0
Gender (% male)	61	0.010	1.424	69.827	0.028
History (years with the problem)	52	−0.037	−1.872[a]	62.644	0.053
Psychodrugs (% that did not take)	52	−0.005	−1.358	57.713	0.031
Agoraphobia in the sample (%)	46	0.009	3.534[b]	49.832	0.200
(C) Methodological characteristics:[c]					
Treatment attrition (%)	61	0.000	−0.007	71.492	0.0
Control attrition (%)	61	−0.002	−0.344	71.455	0.002
Global attrition (%)	61	−0.002	−0.245	71.418	0.001
Differential attrition (%)[d]	61	0.003	0.381	71.211	0.002
Design quality (scale of 0–9 points)	61	0.294	3.194[b]	67.137	0.132
Consumption of psycho drugs (%)[e]	52	0.005	1.358	57.713	0.031
d index in the pretest	51	−0.073	−0.218	61.060	0.001
(D) Extrinsic characteristics:					
Publication year	61	0.027	1.848[f]	70.008	0.046

k: number of studies. *B*: unstandardized regression coefficient. *Z*: statistical test for the regression coefficient. Q_E: statistical test to assess the model misspecification. R^2: proportion of variance accounted for

[a] $p = .061$.
[b] $p < .01$.
[c] All the moderator variables on attrition are referred to the posttest.
[d] The differential attrition was calculated as the difference between the attrition for the treatment group and the attrition for the control group. Thus, a positive value indicates that the treatment group exhibited a larger attrition than the control group.
[e] This moderator variable represent the percentage of patients in the sample that toke psychoactive drugs at controlled and/or reduced doses.
[f] $p = .065$.

Table 4
Results of analyzing the influence of different subject, methodological and extrinsic characteristics on the effect sizes for panic measures.

Moderator variable	*k*	d_+	95% C. I.		ANOVA results
			d_l	d_u	
(A) Subject characteristics:					
Type of disorder:					$Q_B(2) = 11.150$, $p = .004$
PD with agoraphobia	13	1.376	1.041	1.710	$Q_W(58) = 66.438$, $p = .209$
PD without agoraphobia	7	0.430	−0.014	0.874	$\omega^2 = 0.112$
Mixed	41	1.001	0.813	1.188	
Comorbidity in the sample:					$Q_B(1) = 5.432$, $p = .020$
> 50%	24	0.828	0.586	1.070	$Q_W(41) = 52.281$, $p = .111$
< 50%	19	1.272	0.987	1.556	$\omega^2 = 0.070$
Previous treatment?					$Q_B(1) = 1.649$, $p = .199$
Yes	25	1.118	0.844	1.391	$Q_W(32) = 42.813$, $p = .096$
No	9	0.778	0.337	1.219	$\omega^2 = 0.007$
(B) Method characteristics:					
Design type:					$Q_B(1) = 5.033$, $p = .025$
Quasi-experimental	9	0.582	0.174	0.990	$Q_W(59) = 69.873$, $p = .157$
Experimental	52	1.088	0.918	1.257	$\omega^2 = 0.050$
Type of control group:					$Q_B(1) = 14.147$, $p < .001$
Non-active (waiting list)	45	1.187	1.014	1.360	$Q_W(59) = 71.890$, $p = .121$
Active (placebo)	16	0.579	0.314	0.844	$\omega^2 = 0.149$
Control of psychoactive drugs:					$Q_B(1) = 0.235$, $p = .628$
Yes	51	1.027	0.848	1.207	$Q_W(56) = 67.224$, $p = .145$
No	7	1.143	0.710	1.576	$\omega^2 = 0.0$
(C) Extrinsic characteristics:					
Researcher's affiliation:					$Q_B(1) = 5.594$, $p = .018$
Psychology	50	1.101	0.929	1.273	$Q_W(57) = 71.409$, $p = .095$
Psychiatry	9	0.598	0.218	0.978	$\omega^2 = 0.055$

k: number of studies. d_+: mean effect size. 95% C.I.: 95% confidence interval. d_l and d_u: lower and upper confidence limits. Q_B: between-categories *Q* statistic. *p*: probability level. Q_W: within-categories *Q* statistic. ω^2: proportion of variance accounted for.

relationship between them [$Z = 0.077$, $p = .938$, $R^2 = 0.0$, $k = 29$ studies].

A characteristic of the subject samples that was expected to have a statistical association with the effect size was the presence of comorbidity in the patients. Indeed, the samples in which over 50% of the subjects had comorbid disorders produced worse results ($d_+ = 0.828$) than those in which the percentage of subjects with comorbidity was less than 50% ($d_+ = 1.272$) ($p = .020$). Finally, Table 5 shows the results of the regression analyses for different continuous subject variables on the effect sizes. As Table 5 shows, the mean illness duration (mean = 8 years, $SD = 4.4$) had a negative and marginally significant association with the effect size ($p = .065$), which implies that the efficacy of the interventions is reduced as the time that patients suffer the disorder increases. Neither the mean age (mean = 36.4 years, $SD = 2.5$) nor the percentage of males in the samples (mean = 26.8%, $SD = 12.3$%) had a statistical association with effect size.

3.6. Methodological characteristics

Another cluster of moderator variables that may be related to the effect size of the interventions refers to the methodological aspects of the studies. Table 4 shows the results of the ANOVAs carried out with qualitative methodological variables on panic measures. Of the 61 comparisons analyzed, 52 applied an experimental design (random assignment), while only 9 comparisons had a quasi-experimental design (non-random assignment). The analysis of design type showed statistically significant differences between both designs ($p = .025$), the mean effect size obtained being clearly higher with experimental designs ($d_+ = 1.088$) than with quasi-experimental ones ($d_+ = 0.582$). We therefore observed a higher effect size with the better quality designs. As expected, the type of control group also affected the effect size ($p < .001$), which were higher when the control group was non-active — waiting list — ($d_+ = 1.187$) than when it was an active control group — psychological and/or pharmacological placebo — ($d_+ = 0.579$).[11] In general, the effect size decreases by approximately half when an active control group is used, as opposed to a non-active one. These data allow us to estimate the magnitude of the non-specific effects of the psychological intervention of PD, $d_{\text{Non-specific}}$, calculating the difference between both mean effect sizes: $d_{\text{Non-specific}} = d_{\text{Non-active}} - d_{\text{Active}} = 1.187 - 0.579 = 0.608$. Following Cohen (1988), we can say that an effect size of 0.608 can be considered as exhibiting a practical significance of medium magnitude and is, therefore, clearly relevant in practice.

Many of the patients in the samples take psychoactive drugs under psychiatric prescription at reduced and/or controlled doses and this consumption can affect the effect size estimates. To analyze the possible influence of this variable we carried out two complementary analyses. Firstly, out of the 58 studies with panic measures that reported this information, only 7 did not control the consumption of psychotropic drugs during the psychological treatment, but we did not find statistically significant differences between these studies ($d_+ = 1.143$) and those that did control the consumption of these drugs ($d_+ = 1.027$). Secondly, 52 studies with panic measures reported the percentage of patients that had taken psychoactive drugs during the treatment at reduced and/or controlled doses, with a mean percentage of 28.4% ($SD = 24.4$%). In line with the previous result, a regression analysis for this percentage of patients applied on the effect size reflected the absence of a statistical relationship between the two (see Table 5). Therefore, our results seem to guarantee that the studies controlled the consumption of psychoactive drugs reasonably well and that the reduced consumption of some patients in the samples did not influence the effect estimates of the psychological treatments implemented in the studies.

The results of the regression analyses carried out on other continuous methodological variables are also shown in Table 5. Design quality (mean = 6.2, $SD = 0.9$) showed a positive and statistically significant association with effect size ($p < .01$), such that the higher the quality the greater the effect. An important aspect to take into account is the attrition suffered by the treatment and control groups in the posttest. The treatment groups showed a mean attrition of 9.28% (median: 5.26%), whereas for the control groups the mean attrition was 6.39% (median: 0%). The difference between the two mean percentages was statistically significant [$t(64) = 2.264$, $p = .027$].[12] A mean differential attrition of 2.89% (median: 0%) between the treatment and the control groups could bias the effect sizes obtained in the posttest. However, as Table 5 shows, neither the attrition of the treatment groups, nor the attrition of the control groups, the global attrition of the studies, or the differential attrition between the treatment and the control groups showed a statistical relationship with effect size ($p > .05$). Therefore, we can discard differential attrition as a confounding factor in the effect estimates.

3.6.1. Controlling for the d index in the pretest

The d values in the posttest may provide a biased estimate of the treatment effect, if the treatment and the control groups are not equated in the pretest with regard to the dependent variable(s). This can happen even when studies apply an experimental design (*i.e.*, random assignment of subjects to the groups), if the sample sizes are small ($n < 30$ subjects per group) and/or there is differential attrition. A way of testing whether the studies examined equated the dependent variables in the treated and control groups in the pretest consists of calculating the d index. This is defined as the difference between the means of the two groups in the pretest divided by a pooled estimate of the standard deviation. For all the outcome measures the mean d index in the pretest was close to zero and the corresponding confidence intervals reflected an absence of statistically significant differences between the means of the treated and control groups. Therefore we can confirm that the studies examined equated the treatment and control groups in the pretest with regard to the dependent variables.

Nevertheless, there still remained the possibility that there was a relationship between the d index in the pretest and the d index in the posttest of the studies. In order to examine this possibility, we carried out simple weighted regression analyses for the different outcome

[11] Out of the 16 studies that used an active control group, 10 applied pill placebo (e.g., Bakker et al., 2002; Bakker, van Dyck, Spinhoven, & van Balkom, 1999; Sharp et al., 1996, studies a and b), 3 applied some kind of psychological placebo, such as relaxation training (Carlbring, Ekselius, & Anderson, 2003) and information (Klein, Richards, & Austin, 2006, studies a and b), and the remaining 3 studies applied both pill placebo and psychological placebo, the latter being empathic listening (Beck et al., 1994, studies a and b) and clinical management (Loerch et al., 1999). An ANOVA taking the three types of active control group (pill placebo, psychological placebo and both) revealed marginally statistically significant differences between them [$Q_B(2) = 5.388$, $p = .068$, $\omega^2 = 0.178$], pill placebo being the type of active control group that showed the smallest effect size ($d_+ = 0.449$) in comparison with psychological placebo ($d_+ = 1.008$) and both pill placebo and psychological placebo together ($d_+ = 0.684$).

[12] This same analysis was carried out separately for the studies that used non-active and active control groups. With the 48 studies that used non-active control groups the mean percentages of attrition for the treatment and control groups were 9.53% and 6.33% respectively, and the difference was also statistically significant [$t(47) = 2.286$, $p = .027$]. However, with the 17 studies that used active control groups the mean percentages of attrition for the treatment and control groups were more similar, 8.59% and 6.56% respectively, and the difference was not statistically significant [$t(16) = 0.689$, $p = .501$].

measures, taking the *d* index in the pretest as a predictor variable and the *d* index in the posttest (the effect size) as a dependent variable. A non-statistically significant relationship was found for the measures of panic ($Z = 1.440$, $p > .05$; $R^2 = 0.032$), general anxiety ($Z = 1.401$, $p > .05$; $R^2 = 0.046$), and bodily sensations ($Z = 0.764$, $p > .05$; $R^2 = 0.037$). However, a positive and statistically significant relationship was found for the measures of agoraphobia ($Z = 3.138$, $p < .01$; $R^2 = 0.184$), depression ($Z = 3.161$, $p < .01$; $R^2 = 0.213$), global adjustment ($Z = 2.499$, $p < .05$; $R^2 = 0.248$), other measures ($Z = 4.812$, $p < .01$; $R^2 = 0.525$), and the combination of all of these ($Z = 2.333$, $p < .05$; $R^2 = 0.088$). The positive sign of the regression coefficient for these outcome measures means that the better the subjects of the treated group in the pretest with respect to those of the control group, the more favorable the *d* index in the posttest for the treated group. Therefore, depending on the outcome measure the studies show a null or a positive relationship between the *d* index in the pretest and the effect size in the posttest. In respect to panic measures, which are the most relevant in the treatment of patients with PD, the results seem to reveal an absence of relationships between the pretest and posttest data.

3.7. Extrinsic characteristics

Two extrinsic variables were coded: the affiliation of the first researcher in the study and the publication year. As Table 4 shows, the researcher's affiliation showed a statistical relationship with effect size ($p = .018$), the mean effect size being higher for psychologists ($d_+ = 1.101$) than for psychiatrists ($d_+ = 0.598$). Finally, we obtained a positive and marginally significant relationship between the publication year and the effect size ($p = .065$). The positive sign of the regression coefficient indicates that the most recent studies seem to obtain larger effect sizes than the older studies.

3.8. A predictive model of treatment efficacy

As the methodological quality of the studies and the type of control group influenced the effect size, the direct comparison between the mean effect sizes obtained with the different intervention techniques may be affected by biases in their estimates. In order to equate this comparison in terms of the design quality of the studies, we applied a multiple weighted regression model, in which we introduced seven predictor variables referring to the treatments and two predictors referring to the methodological quality of the studies. A cluster formed by seven treatment predictors allowed us to identify, by means of dummy coding (1, present; 0, absent) if the treatment included a technique of relaxation training (Re), breathing retraining (Br), exposure (Exp), cognitive therapy (CT), EMDR, anxiety management training (AMT), or "other techniques".[13] A second cluster was formed by two methodological predictors: the design quality (on a scale of 0 to 9 points) and the type of control group (0, non-active; 1, active). The dependent variable in the model was the *d* index in the posttest for panic measures. Two separate multiple regression analyses, one for each cluster of variables, revealed a statistically significant association with effect size. The percentage of variance accounted for was 31.5% for the cluster of treatment variables and 29.5% for that of methodological predictors. In addition, the nine predictors, taken together, explained 46.8% of the total variance in effect size. The most relevant result however, was that both the treatment and method clusters showed a statistically significant association with effect size once the influence of the other cluster was controlled.

The predictive equation of the model was: $d' = -0.095 - 0.378 \times \mathrm{Re} - 0.016 \times \mathrm{Br} + 0.513 \times \mathrm{Exp} - 0.501 \times \mathrm{CT} - 0.602 \times \mathrm{EMDR} - 0.066 \times \mathrm{AMT} - 0.550 \times \mathrm{Other} + 0.218 \times \mathrm{Quality} - 0.477 \times \mathrm{Control}$. The fact that only the exposure techniques have a positive and statistically significant regression coefficient, as opposed to the negative ones of the other six treatment categories, indicates that exposure is the critical component for the treatment of PD ($p = .026$). With this equation it is possible to obtain predictions of the effect size for certain combinations of treatment and methodological variables. Thus, fixing the maximum score of methodological quality (Quality = 9) and assuming an active control group (Control = 1), the highest predictions of efficacy are obtained with the technique of exposure, either on its own ($d' = 1.90$) or in combination with breathing retraining ($d' = 1.89$) or anxiety management training ($d' = 1.84$). Lower predictions of efficacy are obtained by breathing retraining ($d' = 1.37$), anxiety management ($d' = 1.32$), relaxation training ($d' = 1.01$), CT ($d' = 0.89$), the category 'other techniques' ($d' = 0.84$), and EMDR ($d' = 0.79$).

3.9. Follow-up measures

Although 42 of the 65 comparisons included follow-up data for the treated group, only 8 of these comparisons also reported follow-up data for the control group (Barlow et al., 2000, studies a and b; Clum, Clum et al., 1993, studies a and b; Gould & Clum, 1995; Mavissakalian & Michelson, 1986; Shear et al., 2001, studies a and b). The follow-up periods of the 8 comparisons for which we were able to calculate the effect size, varied between 2 and 24 months (median: 6 months) and the treatments applied were: RB + Exposure + CT (5 comparisons), Exposure + CT (one comparison), Exposure + systematic desensitization (one comparison), and emotion regulation therapy (one comparison). The mean effect size obtained for panic measures was $d_+ = 0.349$ (95% confidence interval: 0.051; 0.646), reflecting an ostensibly lower level of efficacy than that obtained in the posttest. The reduced number of comparisons included in this analysis limits its possibilities for generalization.

4. Discussion

In this paper we have presented the results of a meta-analytic review on the efficacy of psychological treatments for PD with or without agoraphobia. With this purpose, 42 studies that fulfilled our selection criteria were selected giving a total of 65 comparisons between a treatment group and a control group. The results proved that the psychological treatment of PD has a clinically relevant efficacy, for panic measures as well as for those of agoraphobia, general anxiety, depression, bodily sensations, global adjustment, and other related measures. The high heterogeneity found among the effect sizes of the studies indicates that there were multiple factors causing this variability. It is worth noting however that testing multiple moderator variables produces an increase in the Type I error rate of the meta-analytic statistical tests applied. This problem is even more serious when the number of studies in the meta-analysis is small. Therefore, the results of these statistical tests should be interpreted very cautiously, only as tentative relationships with effect size, and they must be examined more thoroughly in future research.

Of the different techniques of intervention examined, the treatment that obtained the most consistent results in favor of its efficacy, once methodological variables were controlled, was the combination of exposure (both interoceptive and *in vivo*) with relaxation training and/or breathing retraining techniques. Relaxation training and breathing retraining techniques proved to be more efficacious in reducing panic behaviors than CT, and CT alone does not offer such positive results as when exposure, RB and anxiety

[13] Please see Appendix A for details of the treatment components included in each study.

management techniques are applied. The application of techniques other than those cited here does not seem to contribute an additional clinically relevant improvement. Therefore, our results showed that the most efficacious treatment for PD with or without agoraphobia is one that combines exposure (both interoceptive and non interoceptive) with relaxation training, breathing retraining or anxiety management training. Of the different variants of exposure, *in vivo* exposure is the one that provides the greatest benefits. As regards the measures of agoraphobia, once again exposure was the most relevant technique, although the existence of differences in efficacy between the different intervention techniques did not prove to be as marked as with the panic measures. These results concur, in general, with some of the previously published meta-analyses (Clum, Clum et al., 1993; Mattick et al., 1990; van Balkom et al., 1995).

It is worth highlighting other factors that seemed to be related to effect size. Coinciding with the meta-analyses by Bakker et al. (1998) and Cox et al. (1992), the interventions were more efficacious when they included homework and a follow-up program. With regard to patient characteristics, we found that psychological treatments achieved better results the shorter the time that the patients had been suffering the disorder and in the absence of comorbidity with other disorders. We found, however, the absence of a statistical relationship between the percentage of patients with agoraphobia and effect size both in panic and in agoraphobia measures.

Our analyses, as well as those of Mitte (2005), showed how methodological characteristics of the studies influence the effect estimates, the greater effect sizes being obtained with higher design quality, with random assignment of the patients to the groups and when the control group was non-active. The comparison between the mean effects obtained when the control group was active or non-active revealed the existence of non-specific effects of the therapy that we estimated as $d_{\text{Non-specific}} = 0.608$, an effect that can be considered of medium magnitude and, therefore, clinically relevant (Cohen, 1988). Our analyses enabled us to discard publication bias as a threat to the validity of our results. We can also confirm that there are no biases in the effect estimates due to differential attrition between the treatment and control groups or the consumption of psychoactive drugs at controlled and/or reduced doses. As an additional methodological control, we also calculated the *d* index in the pretest for the different outcome measures. Although we found that the studies equated the groups reasonably well in the outcome measures in the pretest, we observed that in some outcome measures a positive and statistically significant relationship was found between the *d* index in the pretest and the effect size in the posttest (for measures of agoraphobia, depression, global adjustment, other measures, and the overall outcome). For panic measures however, which were the main outcome measure in our meta-analysis, we found a null relationship between the *d* indices in the pretest and the posttest.

The inclusion of methodological factors together with the treatment techniques in a multiple regression model enabled us to propose a predictive model of the effect size expected for a certain combination of techniques after controlling for the influence of methodological factors. This predictive equation can be used in the future by other researchers and clinicians to make predictions of efficacy for different intervention techniques. In addition, the multiple regression showed that the differential efficacy of the treatments was still maintained, once the influence of the method variables was controlled.

4.1. Implications for clinical practice

Although we can affirm that exposure is the treatment of choice for reducing panic behaviors, the inclusion of relaxation training and breathing retraining techniques is highly recommended since these techniques contribute to improving the effects of exposure. This combination of techniques for the treatment of PD, in addition to being included in the treatment guides for therapists, has also been proposed in self-help manuals for the patients (e.g., Beck & Zebb, 1994). However, it is still necessary to obtain more evidence in order to be able to recommend these programs as a standard alternative that could substitute the currently existing ones. A possible mechanism of change operated by exposure may be related to decreased anxiety sensitivity. In particular, exposure to fearful situations and/or sensations produces a modification of the fear structures leading to the habituation or extinction of dysfunctional beliefs.

The benefits of the treatment are also improved by the inclusion of homework and a follow-up program after the treatment has finished. These variables contribute to extending the treatment out of the therapeutic context, so that the patient learns how to generalize the results of the therapy to other more natural contexts where panic behaviors and agoraphobic avoidance often appear.

It is highly important for clinical practice to take into consideration a number of limitations that these programs of intervention have. It is worth highlighting the high cost of therapy, since the application of the treatments implies the therapist's presence throughout the whole process. Furthermore, its application requires a highly qualified professional. All of this, together with the long waiting lists in the public health service and the extended duration of these treatments, complicates the viability of its routine application. Moreover, the absence of a relationship between the duration of the therapy and the effect size should lead to the exploration of strategies to shorten the real time of the duration of therapy, as well as the use of new communication technologies to achieve a greater efficiency in its application by reducing the time of real contact between the patient and the therapist (Carlbring et al., 2003).

4.2. Implications for future research

Some methodological deficiencies in the empirical studies should be addressed in future research in this field. Firstly, very few studies reported follow-up data for the control groups. Secondly, and ethical problems aside, an important recommendation for future studies on the efficacy of the treatment of PD is to include follow-up data in order to obtain long-term comparisons between the treated and control groups that would enable us to evaluate the duration of the benefits of therapy. Thirdly, only 17 of the 65 comparisons of our meta-analysis incorporated an active control group. Therefore, another recommendation would be to incorporate psychological and/or pharmacological placebos more often in order to control the non-specific effects of therapy, which in this case seem to be considerable according to our estimates. For example, patients in an attention placebo control group receive treatment that mimics the amount of time and attention received by the treatment group but is thought not to have a specific effect upon them.

As Rosenberg and Hougaard (2005) stated, patients with PD that also suffer severe agoraphobia and those that suffer other comorbid pathologies (depression, personality disorders, and generalized anxiety) predict a lesser treatment efficacy. Furthermore, we know that *in vivo* exposure alone improves subjects with agoraphobia, while subjects with PD need other intervention techniques, such as relaxation training, breathing retraining or cognitive therapy. Therefore, we recommend that the studies report the percentage of patients with agoraphobia in the sample, as well as the distribution of other comorbid disorders. The incorporation of these data into the studies would allow us to verify whether the benefits of the intervention techniques vary according to these characteristics of the patients.

Appendix A. Some of the main characteristics and *d* indices of each study included in the meta-analysis

Author(s) and year	Treatment combination	Duration (in weeks)	Mean age	% male	% of agoraphobia	Design type	Attrition (%)	Design quality	N^a	d_{Panic}	$d_{Agoraph}$
Arntz & van den Hout (1996)a	E + CT	12.0	33.9	61.1	–	2	0.00	5.5	36	1.341	0.724
Arntz & van den Hout (1996)b	RB + E	12.0	33.9	61.1	–	2	2.70	5.0	36	0.581	0.645
Bakker et al. (1999)	CT	12.0	34.4	29.8	92.5	1	16.40	6.5	67	−0.086	0.152
Bakker et al. (2002)	E + CT	12.0	34.5	32.0	–	1	19.10	5.5	67	0.602	2.228
Barlow et al. (1984)	RB + CT[b]	14.0	38.0	65.0	0.0	1	0.00	6.5	11	–	1.531
Barlow et al. (1989)a	RB	15.0	37.0	16.0	–	1	19.35	6.5	25	1.282	–
Barlow et al. (1989)b	E + CT	15.0	36.0	20.0	–	1	6.25	6.5	30	1.023	1.236
Barlow et al. (1989)c	RB + E + CT	15.0	33.9	16.3	–	1	13.89	6.5	31	0.927	0.098
Barlow et al. (2000)a	RB + E + CT	12.0	37.5	36.8	0.0	1	30.00	7.0	101	0.483	0.958
Barlow et al. (2000)b	RB + E + CT	12.0	37.8	41.9	0.0	1	32.00	7.0	87	0.692	–
Beck et al. (1994)a	CT	10.0	37.5	25.7	0.0	1	11.36	6.5	39	0.594	–
Beck et al. (1994)b	RB	10.0	37.5	25.7	0.0	1	2.30	7.0	41	0.582	0.150
Black et al. (1993)	RB + CT	8.0	37.5	24.0	82.0	1	32.00	6.5	50	0.697	1.670
Carlbring et al. (2001)	RB + E + CT	9.5	34.0	29.3	–	1	12.50	5.0	17	0.352	1.003
Carlbring et al. (2003)	RB + E + CT	20.0	38.0	31.8	90.1	1	22.50	4.5	41	0.833	1.300
Carter et al. (2003)	RB + E + CT	11.0	41.5	0.0	100.0	1	21.50	5.5	25	2.000	0.541
Casey et al. (2005)	RB + E + CT	–	37.5	30.0	80.0	1	0.00	8.0	60	1.466	–
Clark et al. (1994)a	E + CT	12.0	34.6	22.0	81.0	1	13.50	6.0	32	2.651	0.264
Clark et al. (1994)b	RB + E	12.0	34.6	22.0	81.0	1	13.50	6.0	32	1.232	1.559
Clark et al. (1999)a	E + CT	12.0	34.0	38.0	85.0	1	3.45	8.0	28	1.896	0.806
Clark et al. (1999)b	E + CT	2.0	34.0	38.0	85.0	1	0.00	8.5	28	1.837	2.185
Clum, Clum et al. (1993a)	RB + E + CT	6.0	35.5	16.7	–	2	0.50	4.5	18	0.508	–
Clum, Watkins et al. (1993)	E	6.0	35.5	16.7	–	2	0.60	4.5	14	−0.023	–
Craske et al. (2005)	RB + E + CT	10.0	35.1	49.0	29.7	1	13.90	6.0	37	1.939	−0.129
Chambless et al. (1986)	RB + E + CT[c]	2.0	35.2	14.3	100.0	2	3.00	4.5	35	0.757	–
Febbraro et al. (1999)a	RB + E + CT	8.0	44.4	25.4	1.0	1	0.00	6.0	31	0.281	1.458
Febbraro et al. (1999)b	RB + E + CT	8.0	44.4	25.4	1.0	1	0.00	6.0	29	0.410	–
Feske & Goldstein (1997)a	EMDR	3.0	35.2	22.5	95.3	1	3.60	6.0	27	0.797	–
Feske & Goldstein (1997)b	EMDR	3.0	35.2	22.5	95.3	1	3.20	6.0	30	0.431	–
Goldstein et al. (2000)	EMDR	4.0	38.2	19.6	100.0	1	15.60	6.0	27	0.623	0.721
Gould & Clum (1995)	RB + E + CT	4.0	30.2	16.0	84.0	1	16.70	6.0	25	0.403	0.429
Gould et al. (1993)a	RB + E + CT	4.0	35.7	35.0	94.0	1	8.30	6.0	22	1.702	1.187
Gould et al. (1993)b	RB + E + CT	4.0	35.7	35.0	94.0	1	4.80	6.0	20	1.073	–
Ito et al. (2001)a	E	10.0	37.0	36.0	100.0	1	5.10	7.0	37	3.468	0.702
Ito et al. (2001)b	RB + E	10.0	37.0	36.0	100.0	1	0.00	7.0	38	3.263	–
Ito et al. (2001)c	RB + E	10.0	37.0	36.0	100.0	1	7.70	7.0	36	3.023	–
Kenardy et al. (2003)a	E + CT	6.0	36.8	24.5	76.1	1	12.10	6.5	80	1.227	–
Kenardy et al. (2003)b	E + CT	6.0	36.8	24.5	76.1	1	14.60	6.5	82	1.441	–
Kenardy et al. (2003)c	E + CT	12.0	36.8	24.5	76.1	1	8.80	6.5	83	1.835	−0.096
Klein et al. (2006)a	E + CT	6.0	–	20.0	81.8	1	16.20	7.0	31	1.351	–
Klein et al. (2006)b	E + CT	6.0	–	20.0	81.8	1	22.20	6.5	28	0.919	–
Klosko et al. (1990)	RB + E + CT	15.0	37.0	26.0	–	1	11.76	5.5	30	1.040	1.231
Lidren et al. (1994)a	E + CT[d]	8.0	37.5	31.0	83.3	1	0.00	6.5	24	1.473	–
Lidren et al. (1994)b	E + CT[d]	8.0	32.0	17.0	83.3	1	0.00	6.5	24	1.164	1.054
Loerch et al. (1999)	E + CT	8.0	35.1	25.5	100.0	1	12.00	6.0	28	0.993	0.142
Mavissakalian & Michelson (1986)	Other[e]	12.0	36.5	16.0	0.0	1	16.22	6.5	34	−0.098	0.652
Michelson & Mavissakalian (1985)	E	12.0	36.5	16.0	100.0	1	11.40	5.5	34	–	–
Öst and Thulin (2004)a	E	14.0	36.1	31.5	100.0	1	10.60	7.5	47	1.434	–
Öst and Thulin (2004)b	E + CT	14.0	36.1	31.5	100.0	1	8.30	7.5	48	1.779	1.900
Rosenberg & Hougaard (2005)	RB + E + CT	–	33.1	36.0	81.1	2	9.00	5.0	93	0.721	0.302
Ross et al. (2005)	RB + E + CT	8.0	39.0	0.0	–	1	47.90	5.0	25	0.815	–
Sharp et al. (1996)a	E + CT	12.0	33.2	26.7	–	1	42.00	6.0	58	0.905	0.699
Sharp et al. (1996)b	E + CT	12.0	38.8	18.2	–	1	16.20	6.5	61	0.665	0.060
Sharp et al. (2004)a	E + CT	12.0	40.2	–	–	1	0.00	8.0	39	–	0.368
Sharp et al. (2004)b	E + CT	12.0	36.7	–	–	1	0.00	8.0	50	–	0.485
Shear et al. (2001)a	Other[f]	12.0	36.7	24.5	0.0	2	30.00	5.5	53	0.044	0.475
Shear et al. (2001)b	RB + E + CT[d]	12.0	34.6	39.0	0.0	2	38.90	5.5	59	0.428	1.142
Smits et al. (2004)	RB + E + CT	8.0	33.9	23.9	100.0	1	0.00	7.0	130	0.674	0.852
Swinson et al. (1995)	E + CT	10.0	40.5	11.9	100.0	1	7.50	6.0	42	0.938	1.147
Taylor et al. (1982)	RB	6.0	34.9	30.6	0.0	2	10.00	5.0	16	0.819	–
Telch et al. (1993)	RB + E + CT	8.0	34.6	26.9	–	1	0.00	7.5	67	0.727	1.054
Telch et al. (1995)	RB + E + CT[d]	8.0	34.8	31.7	–	1	10.26	6.0	140	0.879	2.100
Williams & Falbo (1996)a	CT	8.0	38.0	12.5	91.7	1	0.00	6.5	23	0.681	−0.547
Williams & Falbo (1996)b	E	8.0	38.0	12.5	91.7	1	0.00	6.5	21	1.193	–
Williams & Falbo (1996)c	E + CT	8.0	38.0	12.5	91.7	1	0.00	6.5	22	1.131	1.350

In the variable 'Treatment combination', RB: this category implies that the study applied relaxation training, breathing training or both techniques, E: Exposure, CT: Cognitive Therapy, Other: Other techniques, EMDR: Eye Movement Desensitization and Reprocessing. In the variable 'design type', 1 indicates experimental design (random assignment), whereas 2 indicates quasi-experimental design (non-random assignment). d_{Panic}: standardized mean difference in the posttest between treatment and control groups for panic measures. $d_{Agoraph}$: standardized mean difference in the posttest between treatment and control groups for agoraphobia measures.

[a] *N*: total sample size in the posttest.

[b] The treatment also included biofeedback and coping skills training in the management of anxiety.

[c] The treatment also included Gestalt therapy.

[d] The treatment also included anxiety management training.

[e] The treatment consisted of exposure and systematic desensitization.

[f] The treatment consisted of emotion regulation therapy.

References[14]

American Psychiatric Association (APA). (1980). *Diagnostic and statistical manual of mental disorders*, 3rd ed. Washington, DC: Author.

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

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders*, 3rd ed. Washington, DC: Author.

American Psychiatric Association (APA). (2004). *Diagnostic and statistical manual of mental disorders*, 4th ed. Washington, DC: Author.

*Arntz, A., & van den Hout, M. (1996). Psychological treatments of panic disorder without agoraphobia: Cognitive therapy versus applied relaxation. *Behaviour Research and Therapy*, *34*, 113–121.

*Bakker, A., Spinhoven, P., van der Boes, A. J. W., van Balkom, A. J. L. M., & van Dyck, R. (2002). Locus of control orientation in panic disorder and the differential effects of treatment. *Psychotherapy and Psychosomatics*, *71*, 85–89.

Bakker, A., van Balkom, A. J. L. M., Spinhoven, P., Blaauw, B. M., & van Dyck, R. (1998). Follow-up on the treatment of panic disorder with or without agoraphobia: A quantitative review. *Journal of Nervous and Mental Disease*, *186*, 414–419.

*Bakker, A., van Dyck, R., Spinhoven, P., & van Balkom, A. J. L. M. (1999). Paroxetine, clomipramine, and cognitive therapy in the treatment of panic disorder. *Journal of Clinical Psychiatry*, *60*, 831–838.

*Barlow, D. H., Cohen, A. S., Waddell, M. T., Vermilyea, B. B., Klosko, J. S., Blanchard, E. B., et al. (1984). Panic and generalized anxiety disorders: Nature and treatment. *Behavior Therapy*, *15*, 431–449.

Barlow, D. H., & Craske, M. G. (1989). *Mastery of your anxiety and panic.* Albany, NY: Greywind.

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

*Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2000). Cognitive–behavioral therapy, imipramine, or their combination for panic disorder: A randomized controlled trial. *Journal of the American Medical Association*, *283*, 2529–2536.

Barlow, D. H., Raffa, S. D., & Cohen, E. M. (2002). Psychological treatment for panic disorder, phobias and generalized anxiety disorder. In P. E. Nathan, & J. M. Gorman (Eds.), *A guide of treatment that work*, 2nd ed. London: Oxford University Press.

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

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

*Beck, J. G., Stanley, M. A., Baldwin, L. E., Deagle, E. A., III, & Averill, P. M. (1994). Comparison of cognitive therapy and relaxation training for panic disorder. *Journal of Consulting and Clinical Psychology*, *62*, 818–826.

Beck, J. G., & Zebb, B. J. (1994). Behavioral assessment and treatment of panic disorder: Current status, future directions. *Behavior Therapy*, *25*, 581–611.

Becker, B. J. (2005). Failsafe *N* or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111–125). Chichester, UK: Wiley.

Bienvenu, O. J. (2006). Lifetime prevalence of panic disorder is about 5% in the USA. *Evidence-based mental health*, *9*, 114.

*Black, D. W., Wesner, R., Bowers, W., & Gabel, J. (1993). A comparison of fluvoxamine, cognitive therapy, and placebo in the treatment of panic disorder. *Archives of General Psychiatry*, *50*, 44–50.

*Carlbring, P., Ekselius, L., & Anderson, G. (2003). Treatment of panic disorder via the Internet: A randomized trial of CBT vs. relaxation. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 129–140.

*Carlbring, P., Westling, B. E., Ljungstrand, P., Ekselius, L., & Anderson, G. (2001). Treatment of panic disorder via the internet: A randomized trial of a self-help program. *Behavior Therapy*, *32*, 751–764.

*Carter, M. M., Sbrocco, T., Gore, K. L., Marin, N. W., & Lewis, E. L. (2003). Cognitive–behavioral group therapy versus a wait-list control in the treatment of African American women with panic disorder. *Cognitive Therapy and Research*, *27*, 505–518.

*Casey, L. M., Newcombe, P. A., & Oei, T. P. S. (2005). Cognitive mediation of panic severity: The role of catastrophic misinterpretation of bodily sensations and panic self-efficacy. *Cognitive Therapy and Research*, *29*, 187–200.

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of far in agoraphobics: The Bodily Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. *Journal of Consulting and Clinical Psychology*, *32*, 1090–1097.

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

*Chambless, D. L., Goldstein, A. J., Gallagher, R., & Bright, P. (1986). Integrating behavior therapy and psychotherapy in the treatment of agoraphobia. *Psychotherapy*, *23*, 150–159.

Chambless, D. L., & Gillis, M. M. (1993). Cognitive therapy of anxiety disorders. *Journal of Consulting and Clinical Psychology*, *61*, 248–260.

Clark, D. M. (1997). Panic disorder and social phobia. In D. M. Clark, & C. G. Fairburn (Eds.), *Science and practice of cognitive behaviour therapy* (pp. 121–153). New York: Oxford University Press.

Clark, D. M., & Salkovskis, P. M. (1989). *Cognitive therapy for panic and hypochondriasis.* Oxford: Pergamon Press.

*Clark, D. M., Salkovskis, P. M., Hackmann, A., Middleton, H., Anastasiades, P., & Gelder, M. (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry*, *164*, 759–769.

*Clark, D. M., Salkovskis, P. M., Hackmann, A., Wells, A., Ludgate, J., & Gelder, M. (1999). Brief cognitive therapy for panic disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *67*, 583–589.

Clum, G. A., Broyles, S. E., Borden, J. W., & Watkins, P. L. (1990). Validity and reliability of the Panic Attacks Symptoms and Cognitions Questionnaires. *Journal of Psychopathology and Behavioral Assessment*, *12*, 233–245.

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.

*Clum, G. A., Watkins, P. L., Borden, J. W., Broyles, S. E., & Hayes, J. (1993). A comparison of guided imaginal coping and imaginal exposure in the treatment of panic disorder. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, *11*, 179–193.

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

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*, 2nd ed. New York: Russell Sage Foundation.

Cox, B. J., Endler, N. S., Lee, P. S., & Swinson, R. P. (1992). A meta-analysis of treatments for panic disorder with agoraphobia: Imipramine, alprazolam and *in vivo* exposure. *Journal of Behaviour Therapy and Experimental Psychiatry*, *23*, 175–182.

Craske, M. G., & Barlow, D. H. (2006). *Mastery of your anxiety and panic: Therapist guide*, 4th ed. New York, NY: Oxford University Press.

*Craske, M. G., Lang, A. J., Aikins, D., & Mystkowski, J. L. (2005). Cognitive behavioral therapy for nocturnal panic. *Behavior Therapy*, *36*, 43–54.

Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale: Preliminary report. *Psychopharmacology Bulletin*, *9*, 13–25.

*Febbraro, G. A. R., Clum, G. A., Roodman, A. A., & Wright, J. H. (1999). The limits of bibliotherapy: A study of the differential effectiveness of self-administered interventions in individuals with panic attacks. *Behavior Therapy*, *30*, 209–222.

*Feske, U., & Goldstein, A. J. (1997). Eye movement desensitization and reprocessing treatment for panic disorder: A controlled outcome and partial dismantling study. *Journal of Consulting and Clinical Psychology*, *65*, 1026–1035.

Field, A. P. (2003). The problems of using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, *2*, 77–96.

Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the Quality of Life Inventory: A measure of life satisfaction for use in the treatment planning and outcome assessment. *Psychological Assessment*, *4*, 92–101.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

*Goldstein, A. J., de Beurs, E., Chambless, D. L., & Wilson, K. A. (2000). EMDR for panic disorder with agoraphobia: Comparison with waiting list and credible attention-placebo control conditions. *Journal of Consulting and Clinical Psychology*, *68*, 947–956.

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.

*Gould, R. A., & Clum, G. A. (1995). Self-help plus minimal therapist contact in the treatment of panic disorder: A replication and extension. *Behavior Therapy*, *26*, 533–546.

*Gould, R. A., Clum, G. A., & Shapiro, D. (1993). The use of bibliotherapy in the treatment of panic: A preliminary investigation. *Behavior Therapy*, *24*, 241–252.

Guy, W. (1976). ECDEU *Assessment Manual for Psychopharmacology* (218–222). US Dept. Health, Education, and Welfare publication (ADM) 76–338. Rockville, MD: National Institute of Mental Health.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, *32*, 50–55.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychological Methods*, *11*, 193–206.

*Ito, L. M., de Araujo, L. A., Tess, V. L. C., de Barros-Neto, T. P., Asbahr, F. R., & Marks, I. (2001). Self-exposure therapy for panic disorder with agoraphobia: Randomised controlled study of external versus interoceptive self-exposure. *British Journal of Psychiatry*, *178*, 331–336.

*Kenardy, J. A., Dow, M. G. T., Johnston, D. W., Newman, M. G., Thomson, A., & Taylor, C. B. (2003). A comparison of delivery methods of cognitive–behavioral therapy for panic disorder: An international multicenter trial. *Journal of Consulting and Clinical Psychology*, *71*, 1068–1075.

*Klein, B., Richards, J. C., & Austin, D. W. (2006). Efficacy of Internet therapy for panic disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *37*, 213–238.

Klerman, G., Weissman, M. M., Oullette, R., Johnson, J., & Greenwald, S. (1991). Panic attacks in the community: Social morbidity and health care utilization. *Journal of the American Medical Association*, *265*, 742–746.

*Klosko, J. S., Barlow, D. H., Tassinari, R., & Cerny, J. A. (1990). A comparison of alprazolam and behavior therapy in treatment of panic disorder. *Journal of Consulting and Clinical Psychology*, *58*, 77–84.

*Lidren, D. M., Watkins, P. L., Gould, R. A., Clum, G. A., Asterino, M., & Tulloch, H. L. (1994). A comparison of bibliotherapy and group therapy in the treatment of panic disorder. *Journal of Consulting and Clinical Psychology*, *62*, 865–869.

*Loerch, B., Graf-Morgenstern, M., Hautzinger, M., Schlegel, S., Hain, C., Sandmann, J., et al. (1999). Randomised placebo-controlled trial of moclobemide, cognitive–behavioural therapy and their combination in panic disorder with agoraphobia. *British Journal of Psychiatry*, *174*, 205–212.

[14] References marked with an asterisk represent studies included in the meta-analysis.

Marín-Martínez, F., & Sánchez-Meca, J. (1998). Testing dichotomous moderators in meta-analysis. *Journal of Experimental Education*, *67*, 69–81.
Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, *17*, 263–267.
Mattick, R. P., Andrews, G., Hadzi-Pavlovic, D., & Christensen, H. (1990). Treatment of panic and agoraphobia: An integrative review. *Journal of Nervous and Mental Disease*, *178*, 567–576.
*Mavissakalian, M., & Michelson, L. (1986). Agoraphobia: Relative and combined effectiveness of therapist-assisted *in vivo* exposure and imipramine. *Journal of Clinical Psychiatry*, *47*, 117–122.
*Michelson, L., & Mavissakalian, M. (1985). Psychophysiological outcome of behavioral and pharmacological treatments of agoraphobia. *Journal of Consulting and Clinical Psychology*, *53*, 229–236.
Mitte, K. (2005). A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *Journal of Affective Disorders*, *88*, 27–45.
National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy of Sciences Press.
Oei, T. P. S., Llamas, M., & Devilly, G. J. (1999). The efficacy and cognitive processes of cognitive behaviour therapy in the treatment of panic disorder with agoraphobia. *Behavioural and Cognitive Psychotherapy*, *27*, 63–88.
*Öst, L. -G., Thulin, U., & Ramnerö, J. (2004). Cognitive behaviour therapy vs. exposure *in vivo* in the treatment of panic disorder with agoraphobia. *Behaviour Research and Therapy*, *42*, 1105–1127.
Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency, and the prediction of fearfulness. *Behaviour Research and Therapy*, *24*, 1–8.
*Rosenberg, N. K., & Hougaard, E. (2005). Cognitive–behavioural group treatment of panic disorder and agoraphobia in a psychiatric setting: A naturalistic study of effectiveness. *Nordic Journal of Psychiatry*, *59*, 198–204.
*Ross, C. J. M., Davis, T. M. A., & MacDonald, G. F. (2005). Cognitive–behavioral treatment combined with asthma education for adults with asthma and coexisting panic disorder. *Clinical Nursing Research*, *14*, 131–157.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments* Chichester, UK: Wiley.
Sánchez-Meca, J., & Marín-Martínez, F. (1998). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical and Statistical Psychology*, *51*, 311–326.
Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*, 31–48.
Shapiro, F. (1995). *Eye movement desensitization and reprocessing: Basic principles, protocols, and procedures.* New York: Guilford Press.
*Sharp, D. M., Power, K. G., Simpson, R. J., Swanson, V., Moodie, E., Anstee, J. A., et al. (1996). Fluvoxamine, placebo, and cognitive behaviour therapy used alone and in combination in the treatment of panic disorder and agoraphobia. *Journal of Anxiety Disorders*, *10*, 219–242.
*Sharp, D. M., Power, K. G., & Swanson, V. (2004). A comparison of the efficacy and acceptability of group versus individual cognitive behaviour therapy in the treatment of panic disorder and agoraphobia in primary care. *Clinical Psychology and Psychotherapy*, *11*, 73–82.
*Shear, M. K., Houck, P., Greeno, C., & Masters, S. (2001). Emotion-focused psychotherapy for patients with panic disorder. *American Journal of Psychiatry*, *158*, 1993–1998.
*Smits, J. A. J., Powers, M. B., Cho, Y., & Telch, M. J. (2004). Mechanism of change in cognitive–behavioral treatment of panic disorder: Evidence for the fear of fear meditational hypothesis. *Journal of Consulting and Clinical Psychology*, *72*, 646–652.
Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologist Press.
Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–100). Chichester, UK: Wiley.
*Swinson, R. P., Fergus, K. D., Cox, B. J., & Wickwire, K. (1995). Efficacy of telephone-administered behavioral therapy for panic disorder with agoraphobia. *Behaviour Research and Therapy*, *33*, 465–469.
Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training and dissemination of empirically-validated psychosocial treatments: Report and recommendations. *The Clinical Psychologist*, *48*, 3–23.
*Taylor, C. B., Kenigsberg, M. T., & Robinson, J. M. (1982). A controlled comparison of relaxation and diazepam in panic disorder. *Journal of Clinical Psychiatry*, *43*, 423–425.
*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.
*Telch, M. J., Schmidt, N. B., Jaimez, T. L., Jacquin, K. M., & Harrington, P. J. (1995). Impact of cognitive–behavioral treatment on quality of life in panic disorder patients. *Journal of Consulting and Clinical Psychology*, *63*, 823–830.
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.
Tsao, J. C. I., Mystowski, J. L., Zucker, B. G., & Craske, M. G. (2005). Impact of cognitive–behavioral therapy for panic disorder on comorbidity: A controlled investigation. *Behaviour Research and Therapy*, *43*, 959–970.
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., Nauta, M. C. E., & Bakker, A. (1995). Meta-analysis on the treatment of panic disorder with agoraphobia: Review and re-examination. *Clinical Psychology and Psychotherapy*, *2*, 1–14.
Weissman, K. E., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry*, *33*, 1111–1115.
Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *69*, 875–899.
Wilkinson, G., Balestrieri, M., Ruggeri, M., & Bellantuono, C. (1991). Meta-analysis of double-blind placebo controlled trials of antidepressants and benzodiazepines for patients with panic disorder. *Psychological Medicine*, *21*, 991–998.
*Williams, S. L., & Falbo, J. (1996). Cognitive and performance-based treatments for panic attacks in people with varying degrees of agoraphobic disability. *Behaviour Research and Therapy*, *34*, 253–264.