Relative effects of cognitive and behavioral therapies on generalized anxiety disorder, social anxiety disorder and panic disorder: A meta-analysis

Pim Cuijpers (PhD.) (Professor)\textsuperscript{a,b,*}, Claudio Gentili\textsuperscript{c}, Rosa M. Banos\textsuperscript{d,e}, Javier García-Campayo\textsuperscript{f}, Cristina Botella\textsuperscript{e,g}, Ioana A. Cristea\textsuperscript{c,h}

\textsuperscript{a} Department of Clinical Psychology, VU University Amsterdam, The Netherlands
\textsuperscript{b} EMGO Institute for Health and Care Research, The Netherlands
\textsuperscript{c} Department of General Psychology, University of Padova, Padova, Italy
\textsuperscript{d} Department of Personality, Assessment and Psychological Treatments, Valencia, Spain
\textsuperscript{e} Ciber Fisiopatología Obesidad y Nutrición (CB06/03), Instituto Salud Carlos III, Madrid, Spain
\textsuperscript{f} Department of Psychiatry, Miguel Servet Hospital & University of Zaragoza, Red Investigación en Atención Primaria (REDIAPP), Instituto Aragonés de Ciencias de la Salud, Zaragoza, Spain
\textsuperscript{g} Department of Psychología Básica, Clínica y Psicobiología, Castellón, Spain
\textsuperscript{h} Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Cluj-Napoca, Romania

A R T I C L E   I N F O

Article history:
Received 8 August 2016
Accepted 7 September 2016

Keywords:
Social anxiety disorder
Generalized anxiety disorder
Panic disorder
Cognitive therapy
Behavior therapy
Meta-analysis

A B S T R A C T

Although cognitive and behavioral therapies are effective in the treatment of anxiety disorders, it is not clear what the relative effects of these treatments are. We conducted a meta-analysis of trials comparing cognitive and behavioral therapies with a control condition, in patients with social anxiety disorder (SAD), generalized anxiety disorder (GAD) and panic disorder. We included 42 studies in which generic measures of anxiety were used (BAI, HAMA, STAI-State and Trait). Only the effects of treatment for panic disorder as measured on the BAI (13.33 points; 95% CI: 10.58–16.07) were significantly (p<0.001) larger than the effect sizes on GAD (6.06 points; 95% CI: 3.96–8.16) and SAD (5.92 points; 95% CI: 4.64–7.20). The effects remained significant after adjusting for baseline severity and other major characteristics of the trials. The results should be considered with caution because of the small number of studies in many subgroups and the high risk of bias in most studies.

© 2016 Published by Elsevier Ltd.

Contents

1. Introduction .................................................................................................................. 80

2. Methods ..................................................................................................................... 80
2.1. Identification and selection of studies ................................................................. 80
2.2. Quality assessment and data extraction ............................................................... 81
2.3. Meta-analyses .......................................................................................................... 81

3. Results ....................................................................................................................... 81
3.1. Selection and inclusion of studies ......................................................................... 81
3.2. Characteristics of included studies ....................................................................... 81
3.3. Quality assessment ................................................................................................. 83
3.4. Baseline differences among patients in treatments on GAD, SAD and panic disorders .................................................................................................................. 83
3.5. Differential outcomes on anxiety of trials across GAD, SAD and panic disorders .................................................................................................................. 84

* Corresponding author at: Clinical Psychology, Department of Clinical, Neuro and Developmental Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.

E-mail address: p.cuijpers@vu.nl (P. Cuijpers).

http://dx.doi.org/10.1016/j.janxdis.2016.09.003
0887-6185/© 2016 Published by Elsevier Ltd.
1. Introduction

It is well-established that cognitive and behavioral therapies are effective in the treatment of anxiety disorders, including social anxiety disorder (Acarturk, Cuijpers, van Straten, & de Graaf, 2009; Esksildsen, Hougaard, & Rosenberg, 2010), generalized anxiety disorder (Cuijpers, Sijbrandsj et al., 2014; Hunot, Churchill, Silva de Lima, & Teixeira, 2007) and panic disorder (Sánchez-Meca, Rosa-Alcázar, Martín-Martínez, & Gómez-Conesa, 2010). Although some other types of treatment have been developed for the treatment of anxiety disorders, like psychodynamic (Leichsenring et al., 2009; Milrod et al., 2007) and interpersonal psychotherapies (Dagöö et al., 2014; Lipsitz et al., 2008; Vos, Huibers, Diels, & Arntz, 2012), cognitive and behavioral therapies have been examined in dozens of randomized trials and have been consistently shown to be effective in the treatment of anxiety disorders with large effect sizes across disorders.

It is not clear, however, what are the relative effects of the treatment of one anxiety disorder compared to another. Most outcome instruments are specifically designed to measure the effects of each disorder separately. For example many studies examining the effects of treatments of social anxiety disorder use the Liebowitz Social Anxiety Scale (Baker, Heinrichs, Kim, & Hofmann, 2002; Liebowitz, 1987) as outcome measure, but many studies on generalized anxiety disorder use the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990), while studies on panic disorder use the frequency of panic attacks as main outcome measure. This makes it impossible to compare the relative effects of cognitive and behavioral treatments in different anxiety disorders.

The relative effects of treatments for different anxiety disorders are, however, important for several reasons. Firstly, how well a disorder can be treated is important from a public health perspective (GBD 2013 DALYS and HALE Collaborators et al., 2015). If a disorder can be treated effectively it is less important to develop new treatments that could potentially be better than existing ones since the disease burden of this disorder can be ameliorated with existing treatments. If a disorder cannot be treated effectively, it is more important to develop new or improved therapies. Understanding the relative effectiveness of a treatment is also important for clinicians and patients when deciding whether and how to treat the disorder. From a scientific perspective, differences in effectiveness may be helpful in understanding the underlying processes that lead to these disorders and in explaining how treatments work.

The relative effects of cognitive and behavioral treatments between various anxiety disorders can be assessed in meta-analyses, which delineate the relative effects by giving estimates in terms of effect sizes (standardized mean difference). However, these effect sizes are still statistical concepts, indicating the difference between a treatment and control group in terms of standard deviations, and do not say anything about the clinical effect of a treatment (Cuijpers, Turner, Koole, van Dijke, & Smit, 2014). For example, an effect size of d = 0.1 on mortality would be most clinicians and patients be considered a highly clinically important effect, while an effect of d = 0.1 on social skills would not be considered be relevant by most. This implies that effect sizes cannot be directly compared across disorders, because from a clinical perspective an effect size in one disorder may have different implications than that same effect size in another disorder.

However, in the field of anxiety there are several outcome instruments that measure general levels of anxiety and that are not related to specific anxiety disorders, such as the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988)), the Hamilton Rating Scale for Anxiety (HAMA; Hamilton (1959)), and the State-Trait Anxiety Inventory (STAI-S and STAI-Trait; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). These instruments may not fully capture the specific effects of treatments on a specific disorder, because they measure anxiety in general, and not the specific symptoms of particular anxiety disorders. However, they can give an indication about the relative effects of treatments across disorders because they give a generic assessment of anxiety, not attached to a specific anxiety disorder. Effect sizes based on these measures can be considered comparable across disorders, because they use exactly the same instrument, in contrast to disorder-specific outcomes whose effect sizes cannot be compared directly. These outcome instruments are used in a considerable number of trials on cognitive and behavioral therapies for anxiety disorders.

We decided to conduct a meta-analysis of trials including instruments that measure general anxiety symptoms in order to make a comparison between the outcomes of cognitive and behavioral therapies in three of the most common anxiety disorders: panic disorder, social anxiety disorder, and generalized anxiety disorder. We focused on these three disorders because these are among the most important and common anxiety disorders in adults according to the DSM-5 (American Psychiatric Association, 2013).

2. Methods

2.1. Identification and selection of studies

We searched four major bibliographical databases (PubMed, PsycInfo, Embase and the Cochrane Database of randomized trials) by combining terms (both MeSH terms and text words) indicative of social anxiety disorder (such as social phobia, social anxiety, public-speaking anxiety), generalized anxiety disorder (such as worry and generalized anxiety), and panic (such as panic, panic disorder), with filters for randomized controlled trials. We also checked the references of earlier meta-analyses of psychological treatments for the included disorders. The exact search string for PubMed is given in Appendix A. The deadline for the searches was August 14, 2015.

We included (a) randomized trials (b) in which the effects of a cognitive or behavioral treatment (c) on anxiety measured with the BAI, HAMA, STAI-Trait and/or STAI-State (d) was directly compared with a control group (waiting list, care-as-usual, placebo or other) (e) in adults (f) with a panic disorder (with or without agoraphobia), generalized anxiety disorder (GAD), or social anxiety disorder (SAD). Only studies in which subjects met diagnostic criteria for the disorder according to a structured diagnostic interview (such as the SCID, CIDI, or MINI) were included. We chose the BAI, HAMA and STAI for inclusion because these are by far the most used generic measures of anxiety in outcome studies. Cognitive and behavioral therapies were defined as therapies aimed at cognitive restructuring or at changing current anxiety behav-
ior. Studies on EMDR, interpersonal and psychodynamic therapy were excluded, because they are not considered to be cognitive behavior therapy, and don't offer the cognitive and behavioral strategies that are typically offered in CBT. We also excluded studies in which (applied) relaxation was examined as a stand-alone treatment, because relaxation is effective in GAD, but less so for example in panic disorder (Siever & Chambless, 2007) and may therefore affect pooled outcomes across disorders. Comorbid or specific disorders were not used as an exclusion criterion. Studies on inpatients, adolescents and children (below 18 years of age) were excluded. We also excluded maintenance studies, aimed at people who had already recovered or partly recovered after an earlier treatment, and studies that did not report sufficient data to calculate standardized effect sizes. Studies in English, German, and Dutch were considered for inclusion.

2.2. Quality assessment and data extraction

We assessed the validity of included studies using four criteria of the ‘Risk of bias’ assessment tool, developed by the Cochrane Collaboration (Higgins & Green, 2011). This tool assesses possible sources of bias in randomized trials, including the adequate generation of allocation sequence; the concealment of allocation to conditions; the prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat analyses were conducted, meaning that all randomized patients were included in the analyses). The assessment of the validity of the included studies was conducted by two independent researchers, and disagreements were solved through discussion.

We also coded participant characteristics (disorder; recruitment method; target group); characteristics of the psychotherapies (treatment format; number of sessions); and general characteristics of the studies (country where the study was conducted; year of publication).

2.3. Meta-analyses

For each comparison between a psychotherapy and a control condition, the effect size indicating the difference between the two groups at post-test was calculated (Hedges’s g). Effect sizes were calculated by subtracting (at post-test) the average score of the psychotherapy group, from the average score of the control group, and then dividing this result by the pooled standard deviation. Because some studies had relatively small sample sizes we corrected the effect size for small sample bias (Hedges & Olkin, 1985). If means and standard deviations were not reported, we used the procedures of the Comprehensive Meta-analysis software (see below) to calculate the effect size using dichotomous outcomes; and if these were not available either, we used other statistics (such a t-value or p-value) to calculate the effect size. We also calculated (unstandardized) mean differences that indicate the difference between treatment and control groups in terms of points on the specific scale used (the BAI, HAMA, STAI-State, STAI-Trait).

We also calculated effect sizes indicating the improvement from baseline to post-test for the treatment groups in the studies. Because baseline and post-test values are not independent from each other, we assumed a conservative correlation between baseline and post-test score of r = 0.70.

Apart from the outcomes of the studies on the BAI, HAMA, STAI-State and STAI-Trait, we also collected effect sizes based on disorder specific measures from the included trials, such as the PSWQ for generalized anxiety disorder, and the LSAS for social anxiety. By collecting these disorder-specific outcomes we could assess whether the effect sizes based on the BAI, HAMA and STAI (measuring anxiety in general) differed from disorder-specific outcomes. We did not include measures on cognitions or other indirect outcomes or generic measures of anxiety (other than the BAI, HAMA and STAI).

To calculate pooled mean effect sizes, we used the computer program Comprehensive Meta-Analysis (version 3.3070: CMA). Because we expected considerable heterogeneity among the studies, we employed a random effects pooling model in all analyses (Borenstein, Hedges, Higgins, & Rothstein, 2009).

In order to assess baseline difference among patients with GAD, SAD and panic disorder, we pooled the mean on the BAI, HAMA, STAI-State and STAI-Trait at baseline using the means, standard deviations and N of the treatment groups according to the procedures implemented in CMA. Numbers-needed-to-treated (NNT) were calculated using the formulae provided by Kraemer and Kupfer (2006).

As a test of homogeneity of effect sizes, we calculated the I² statistic, which is an indicator of heterogeneity in percentages. A value of 0% indicates no observed heterogeneity, and larger values indicate increasing heterogeneity, with 25% as low, 50% as moderate, and 75% as high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). We calculated 95% confidence intervals around I² (Ioannidis, Patsopoulos, & Evangelou, 2007), using the non-central chi-squared-based approach within the heterogi module (Orsini, Bottai, Higgins, & Buchan, 2006) for Stata.

We conducted subgroup analyses according to the mixed effects model, in which studies within subgroups are pooled with the random effects model, while tests for significant differences between subgroups are conducted with the fixed effects model. For continuous variables, we used meta-regression analyses to test whether there was a significant relationship between the continuous variable and effect size, as indicated by a Z-value and an associated p-value. Multivariate metaregression analyses, with the effect size as the dependent variable, were conducted in CMA.

We tested for publication bias by inspecting the funnel plot on primary outcome measures and by Duval and Tweedie’s trim and fill procedure (Duval & Tweedie, 2000), which yields an estimate of the effect size after the publication bias has been taken into account (as implemented in CMA).

3. Results

3.1. Selection and inclusion of studies

After examining a total of 10,368 abstracts (6196 after removal of duplicates), we retrieved 1072 full-text papers for further consideration. We excluded 1031 of the retrieved papers. The PRISMA flowchart describing the inclusion process, including the reasons for exclusion, is presented in Fig. 1. A total of 42 studies met inclusion criteria for this meta-analysis (two of the studies were described in one paper; Mohlman et al. (2003)), 20 studies on GAD, 12 studies on panic disorder, and 10 on SAD. Selected characteristics of the included studies are reported in Table 1.

3.2. Characteristics of included studies

In the 42 studies a total of 2477 patients participated (1504 in the treatment groups and 973 in the control groups). Thirty-three studies were aimed at adults in general, six were aimed at older adults and three at other target groups (undergraduate students, unemployed homeless adults and African-American women). Fifteen studies recruited patients exclusively from clinical populations, 22 recruited (also) from the community, and five used other recruitment methods. The 42 studies included a total of 56 comparisons between a CBT condition and a control condition. In the 56 CBT conditions, the treatment was delivered in individual format in 33 studies, in 12 studies it was delivered in group format, in six studies
Table 1
Selected characteristics of randomized controlled trials examining the effects of cognitive and behavioral treatments of generalized anxiety disorders, panic disorder and social anxiety disorder on the BAL, HAMA, STAI-Trait and STAI-State.

<table>
<thead>
<tr>
<th>Study</th>
<th>Disorder</th>
<th>Recr</th>
<th>Target group</th>
<th>Conditions</th>
<th>N</th>
<th>Format</th>
<th>N sess-ions</th>
<th>Study Qual</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abramowitz et al. (2009)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT</td>
<td>11</td>
<td>Gsh</td>
<td>8</td>
<td>-- + *</td>
<td>US</td>
</tr>
<tr>
<td>Akillas and Efran (1995)</td>
<td>SAD</td>
<td>Comm</td>
<td>Undergraduate students</td>
<td>Waiting list</td>
<td>10</td>
<td>Gsh</td>
<td>3</td>
<td>-- *</td>
<td>US</td>
</tr>
<tr>
<td>Andersson et al. (2006)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT (internet + group)</td>
<td>32</td>
<td>Gsh/grp</td>
<td>11</td>
<td>+ *</td>
<td>SW</td>
</tr>
<tr>
<td>Andersson et al. (2012a)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>Internet-based CBT</td>
<td>102</td>
<td>Gsh</td>
<td>9</td>
<td>-- *</td>
<td>SW</td>
</tr>
<tr>
<td>Andersson et al. (2012b)</td>
<td>GAD</td>
<td>Comm</td>
<td>Adults</td>
<td>Internet-based CBT</td>
<td>27</td>
<td>Gsh</td>
<td>8</td>
<td>+ + + +</td>
<td>SW</td>
</tr>
<tr>
<td>Bakhrshani et al. (2007)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT</td>
<td>7</td>
<td>Ind</td>
<td>8</td>
<td>-- *</td>
<td>IR</td>
</tr>
<tr>
<td>Bakker et al. (1999)</td>
<td>Panic</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT</td>
<td>35</td>
<td>Ind</td>
<td>12</td>
<td>-- *</td>
<td>NL</td>
</tr>
<tr>
<td>Barlow et al. (1989)</td>
<td>Panic</td>
<td>Clin</td>
<td>Adults</td>
<td>RT + CBT Waiting list</td>
<td>15</td>
<td>Ind</td>
<td>15</td>
<td>-- +</td>
<td>US</td>
</tr>
<tr>
<td>Barlow et al. (1992)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>Applied relaxation</td>
<td>13</td>
<td>Ind</td>
<td>15</td>
<td>-- +</td>
<td>US</td>
</tr>
<tr>
<td>Beck et al. (1994)</td>
<td>Panic</td>
<td>Comm</td>
<td>Adults</td>
<td>Minimal contact control</td>
<td>17</td>
<td>Grp</td>
<td>10</td>
<td>-- --</td>
<td>US</td>
</tr>
<tr>
<td>Beidel et al. (2014)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>Exposure Exposure + soc skills tr</td>
<td>46</td>
<td>Ind</td>
<td>24</td>
<td>-- + +</td>
<td>US</td>
</tr>
<tr>
<td>Butler et al. (1991)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT Behavior therapy</td>
<td>19</td>
<td>Ind</td>
<td>12</td>
<td>-- +</td>
<td>UK</td>
</tr>
<tr>
<td>Carlbring, Westling, Ljungstrand, Ekselius, and Andersson (2001)</td>
<td>Panic</td>
<td>Comm</td>
<td>Adults</td>
<td>Internet CBT Waiting list</td>
<td>21</td>
<td>Gsh</td>
<td>6</td>
<td>* * sr</td>
<td>SW</td>
</tr>
<tr>
<td>Carlbring et al. (2006)</td>
<td>Panic</td>
<td>Other</td>
<td>Adults</td>
<td>Internet CBT Waiting list</td>
<td>30</td>
<td>Gsh</td>
<td>10</td>
<td>* * sr</td>
<td>SW</td>
</tr>
<tr>
<td>Carter et al. (2003)</td>
<td>Panic</td>
<td>Other</td>
<td>Afro Am women</td>
<td>CBT Waiting list</td>
<td>17</td>
<td>Grp</td>
<td>11</td>
<td>-- *</td>
<td>US</td>
</tr>
<tr>
<td>Clark et al. (1999)</td>
<td>Panic</td>
<td>Clin</td>
<td>Adults</td>
<td>Full CBT</td>
<td>15</td>
<td>Ind</td>
<td>14</td>
<td>-- +</td>
<td>UK</td>
</tr>
<tr>
<td>Clark et al. (2006)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CT Exposure + relaxation Waiting list</td>
<td>21</td>
<td>Ind</td>
<td>14</td>
<td>* * *</td>
<td>UK</td>
</tr>
<tr>
<td>Dugas et al. (2003)</td>
<td>GAD</td>
<td>Comm</td>
<td>Adults</td>
<td>Group CBT Waiting list</td>
<td>25</td>
<td>Grp</td>
<td>14</td>
<td>-- + +</td>
<td>CAN</td>
</tr>
<tr>
<td>Dugas et al. (2010)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT Waiting list</td>
<td>27</td>
<td>Ind</td>
<td>12</td>
<td>* * *</td>
<td>CAN</td>
</tr>
<tr>
<td>Himmel et al. (2014)</td>
<td>SAD</td>
<td>Other</td>
<td>Unemployed homeless adults</td>
<td>Work-related CBT CAU</td>
<td>29</td>
<td>Grp</td>
<td>8</td>
<td>+ +</td>
<td>US</td>
</tr>
<tr>
<td>Hoyer et al. (2009)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>Worry exposure</td>
<td>29</td>
<td>Ind</td>
<td>15</td>
<td>+ + +</td>
<td>GER</td>
</tr>
<tr>
<td>Ito et al. (2001)</td>
<td>Panic</td>
<td>Comm</td>
<td>Adults</td>
<td>External cues Interoceptive Combined Waiting list</td>
<td>21</td>
<td>Ind</td>
<td>7</td>
<td>--</td>
<td>BR</td>
</tr>
<tr>
<td>Ladouceur et al. (2000)</td>
<td>GAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT Waiting list</td>
<td>14</td>
<td>Ind</td>
<td>16</td>
<td>-- +</td>
<td>CAN</td>
</tr>
<tr>
<td>Linden et al. (2005)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>Contact + Waiting list CBT (+ placebo) Placebo</td>
<td>12</td>
<td>CBT</td>
<td>36</td>
<td>22</td>
<td>-- *</td>
</tr>
<tr>
<td>Loerch et al. (1999)</td>
<td>Panic</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT Waiting list</td>
<td>36</td>
<td>Ind</td>
<td>14</td>
<td>9</td>
<td>-- +</td>
</tr>
<tr>
<td>Mohlman et al. (2003)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>CBT Waiting list</td>
<td>11</td>
<td>Ind</td>
<td>13</td>
<td>-- +</td>
<td>US</td>
</tr>
<tr>
<td>Mohlman et al. (2003)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>Enhanced CBT Waiting list</td>
<td>8</td>
<td>Ind</td>
<td>13</td>
<td>-- +</td>
<td>US</td>
</tr>
<tr>
<td>Power et al. (1989)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT Placebo</td>
<td>10</td>
<td>Ind</td>
<td>4</td>
<td>--</td>
<td>UK</td>
</tr>
<tr>
<td>Power et al. (1990)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT Placebo</td>
<td>11</td>
<td>Ind</td>
<td>7</td>
<td>--</td>
<td>UK</td>
</tr>
<tr>
<td>Salaberry and Echeburúa (1998)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>Self exposure in vivo CBT + Self exposure in vivo Waiting list</td>
<td>24</td>
<td>Grp</td>
<td>8</td>
<td>-- *</td>
<td>SP</td>
</tr>
</tbody>
</table>
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Disorder</th>
<th>Recr</th>
<th>Target group</th>
<th>Conditions</th>
<th>N</th>
<th>Format</th>
<th>N sess-ions</th>
<th>Study Quala</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharp et al. (1997)</td>
<td>Panic</td>
<td>Clin</td>
<td>Adults</td>
<td>CBT (+ placebo)</td>
<td>36</td>
<td>Ind</td>
<td>nr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharp et al. (2004)</td>
<td>Panic</td>
<td>Other</td>
<td>Adults</td>
<td>CBT group</td>
<td>38</td>
<td>Grp</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stangier et al. (2003)</td>
<td>SAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT Individual</td>
<td>22</td>
<td>Ind</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2003a)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>CBT</td>
<td>29</td>
<td>Grp</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2003b)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>CBT</td>
<td>6</td>
<td>Ind</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanley et al. (2014)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>CBT lay therapists</td>
<td>76</td>
<td>Ind (tel)</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swinson et al. (1995)</td>
<td>Panic</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT</td>
<td>20</td>
<td>Ind (tel)</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van der Heiden et al. (2012)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>Metacogn therapists</td>
<td>54</td>
<td>Ind</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wetherell et al. (2003)</td>
<td>GAD</td>
<td>Comm</td>
<td>Older adults</td>
<td>CBT</td>
<td>18</td>
<td>Grp</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White et al. (1992)</td>
<td>GAD</td>
<td>Clin</td>
<td>Adults</td>
<td>CT</td>
<td>31</td>
<td>Grp</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinbarg et al. (2007)</td>
<td>GAD</td>
<td>Comm</td>
<td>Adults</td>
<td>CBT</td>
<td>8</td>
<td>Ind</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BR: Brazil; CAN: Canada; CAU: Care as Usual; CBT: Cognitive Behavioral Therapy; Clin: participants recruited in a clinical setting; Comm: participants recruited in a community setting; CT: Cognitive Therapy; GAD: Generalized Anxiety Disorder; GER: Germany; Grp: group format; Gsh: guided self-help format; Ind: individual format; Ind (tel): individual telephone format; Intol-Uncert ther: Intolerance-of-Uncertainty Therapy; IR: Iran; MDD: Major Depressive Disorder; NL: The Netherlands; Panic: Panic Disorder; Recr: Recruitment; RT: Relaxation Therapy; SAD: Social Anxiety Disorder; Soc. skills tr.: Social Skills Training; SP: Spain; SW: Sweden; UK: United Kingdom; US: United States of America.

a In this column a positive (+) or negative (−) sign is given for four quality criteria of the study, respectively: allocation sequence; concealment of allocation to conditions; blinding of assessors; intention-to-treat analyses; and selective outcome reporting. Sr in the third criterion indicates that only self-report measures were used (and no assessor was used).

**Disorder**

<table>
<thead>
<tr>
<th>GAD</th>
<th>SAD</th>
<th>Panic</th>
<th>Total</th>
</tr>
</thead>
</table>

**References identified by literature search:**

- Pubmed: 757
- Cochrane: 1831
- PsycINFO: 558
- Embase: 596

**Total:** 3742

**After removal of duplicates:** 2267

**Full-text retrieved:** 196

**Excluded:**

- Duplicate papers: 36
- No diagnosis: 44
- No relevant comparison: 30
- No CBT: 22
- No BAI, HAMA, STAI: 29
- Other reason: 15

**Total excluded:** 176

**Included in meta-analysis:** 20

**Fig. 1.** Flowchart for the inclusion of studies.

a guided self-help format was utilized, and 5 studies used another format (telephone or mixed). The number of treatment sessions ranged from 3 to 24, with the majority (35 out of 56) having 12 sessions or less. In 31 studies a waiting list control group was used, 6 used a pill placebo control group, three had a care-as-usual control group, and two a minimal contact control. Fifteen studies were conducted in the US, 22 in Europe, four in Canada and one in Brazil.

### 3.3. Quality Assessment

The quality of the studies varied. Ten studies reported an adequate sequence generation, while the other 32 did not. Eight of the 42 studies reported allocation to conditions by an independent (third) party. Thirty-three studies reported blinding of outcome assessors or used only self-report outcomes and 22 studies conducted intention-to-treat analyses (a post-treatment score was analyzed for every randomized patient even if the last observation prior to attrition had to be carried forward or that score was estimated from earlier response trajectories). Eleven studies met three or four quality criteria, another 8 studies met two criteria, and the remaining 23 studies met one or none of the criteria.

### 3.4. Baseline differences among patients in treatments on GAD, SAD and panic disorders

We first examined whether the baseline scores on the BAI, HAMA, STAI-Trait and STAI-State differed among patients with GAD, SAD and panic disorder. The pooled means are reported in Table 2. As can be seen, the baseline scores of the three disorders differed significantly for all outcome measures, except for the BAI (although there was a trend of p < 0.1 suggesting a possible significant difference). However, the number of studies was very small in some subgroups. Heterogeneity was very high in most analyses...
with larger samples of studies, as is typically the case when pooling absolute numbers (Higgins & Green, 2011). However, because of these baseline differences we decided to add baseline scores in the multivariate analyses examining whether the effect sizes of the therapies differed across disorders (see below).

### 3.5. Differential outcomes on anxiety of trials across GAD, SAD and panic disorders

The differential outcomes of the treatments for the three disorders on the BAI, HAMA, STAI-Trait and STAI-State are reported in Table 3. We only found significant differences across disorders for the BAI, but not for the HAMA, STAI-Trait and STAI-State. However, the number of comparisons in several of the subgroups was small, so the lack of significant relative effects may be related to low statistical power.

For the BAI, we found that the effects of treatments on panic disorder were considerably larger than for GAD and SAD (p < 0.001). The effect sizes for the BAI for all included studies, grouped according to disorder, are presented in Fig. 2. Pairwise comparison showed that the difference between GAD and panic (p < 0.001) and between SAD and panic (p < 0.1) were significant, but the difference between GAD and SAD was not significant (p > 0.1). The difference between the treatment and control groups at post-test was 13.33 points (95% CI: 10.58–16.07) on the BAI for panic disorder, 8.06 points (95% CI: 3.96–8.16) points for GAD, and 5.92 points (95% CI: 4.64–7.20) for SAD. The NNTs were 1.42, 2.54 and 2.54 respectively.

Because the BAI was the most used instrument across disorders, we examined potential publication bias in the studies using the BAI. For GAD we found that adjustment for publication bias resulted in a decrease of the effect size from g = 0.67 to g = 0.66 (95% CI: 0.48–0.84; number of missing studies: 1). For panic we found that the effect size decreased from g = 1.48 to g = 1.40 (95% CI: 1.07–1.72; number of missing studies: 1). For SAD we found no indication for publication bias (unadjusted and adjusted effect sizes were identical with no missing studies).

### 3.6. Improvement from baseline to post-test

Because the baseline scores for the BAI, HAMA, STAI-Trait and STAI-State differed at baseline for the disorders, we also calculated the effect sizes indicating the improvement from baseline to post-test within the treatment groups (Table 4). As can be seen from the Table, the results are very much in line with the findings for the effect sizes indicating the difference between the treatment and control group at post-test. There was a significant difference in the improvement from baseline to post-test between the three anxiety disorders on the BAI (p = 0.01), but not on the HAMA, the STAI-Trait and STAI-State. The improvement in patients with panic disorder was higher than in those with GAD or SAD, and there was no difference between SAD and GAD.

### 3.7. Generic and disorder-specific outcome instruments

In order to examine whether the effect sizes found for the generic anxiety measures (BAI, HAMA and STAI) differed from the disorder-specific outcomes, we calculated the pooled effect sizes for the disorder-specific outcomes (Table 3), as well as the effect size based on the disorder-specific instruments for each disorder that were reported in at least five studies (the PSWQ for GAD, SIAS for SAD, and number of panic attacks for panic). Four studies did not present data on disorder-specific outcome measures and were excluded from these analyses. As can be seen in Table 3, the outcomes for the disorder-specific instruments did not deviate considerably from the generic anxiety measures, and also indicated larger effects for panic disorder than for GAD and SAD. Because the generic and disorder-specific effect sizes came from the same studies, we could not directly compare them and test whether the differences between them were significant.

### 3.8. Multivariate metaregression analyses

We conducted a multivariate metaregression analysis with the effect size based on the BAI as dependent variable. As predictors we entered type of disorder (GAD, SAD or panic disorder), the baseline score on the BAI, and other main characteristics of the studies. The results are presented in Table 5. As can be seen, type of disorder remained a significant predictor of outcome, while adjusting for the other characteristics of the studies. We then conducted a manual back-step metaregression analysis. In this analysis, we dropped the least significant variable in each step, until only significant predictors were retained in the model (Table 5). The results of this parsimonious model indicated that type of disorder was still significantly associated with the effect size and was in fact the only significant predictor.
Table 3
Relative effects of cognitive and behavioral therapies for GAD, SAD and panic disorder on general measures of anxiety: Hedges’ g and Mean Difference. a

<table>
<thead>
<tr>
<th>Disorder-specific outcomes</th>
<th>Ncomp</th>
<th>g</th>
<th>95% CI</th>
<th>MD</th>
<th>95% CI</th>
<th>p</th>
<th>95% CI</th>
<th>p</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beck Anxiety Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>9</td>
<td>0.73</td>
<td>0.48–0.98</td>
<td>6.06</td>
<td>3.96–8.16</td>
<td>0</td>
<td>0–54</td>
<td>0.001</td>
<td>2.54</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>8</td>
<td>0.73</td>
<td>0.56–0.90</td>
<td>5.92</td>
<td>4.64–7.20</td>
<td>46</td>
<td>0–74</td>
<td>0.001</td>
<td>2.54</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>4</td>
<td>1.48</td>
<td>1.13–1.83</td>
<td>13.33</td>
<td>10.58–16.07</td>
<td>0</td>
<td>0–68</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>Hamilton</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>11</td>
<td>0.98</td>
<td>0.70–1.26</td>
<td>6.47</td>
<td>4.90–8.05</td>
<td>46</td>
<td>0–72</td>
<td>0.85</td>
<td>1.95</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>2</td>
<td>1.35</td>
<td>0.88–1.81</td>
<td>7.04</td>
<td>5.10–8.99</td>
<td>22</td>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>11</td>
<td>0.95</td>
<td>0.57–1.33</td>
<td>7.28</td>
<td>4.23–10.34</td>
<td>73</td>
<td>45–84</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>STAI-State</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>5</td>
<td>0.55</td>
<td>0.28–0.81</td>
<td>6.75</td>
<td>3.53–9.98</td>
<td>0</td>
<td>0–64</td>
<td>0.64</td>
<td>3.31</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>3</td>
<td>0.79</td>
<td>0.35–1.23</td>
<td>7.49</td>
<td>3.35–11.62</td>
<td>0</td>
<td>0–73</td>
<td>2.36</td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>2</td>
<td>0.62</td>
<td>–1.16–2.40</td>
<td>8.80</td>
<td>–13.35–31.16</td>
<td>91</td>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAI-Trait</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>16</td>
<td>0.50</td>
<td>0.37–0.64</td>
<td>5.46</td>
<td>4.04–6.88</td>
<td>0</td>
<td>0–45</td>
<td>0.86</td>
<td>3.62</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>2</td>
<td>0.65</td>
<td>–0.55–1.84</td>
<td>7.11</td>
<td>–5.57–19.79</td>
<td>77</td>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>5</td>
<td>0.30</td>
<td>–0.50–1.09</td>
<td>3.41</td>
<td>–2.65–9.47</td>
<td>85</td>
<td>59–92</td>
<td>5.95</td>
<td></td>
</tr>
<tr>
<td>All disorder-specific outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>22</td>
<td>0.64</td>
<td>0.49–0.78</td>
<td>24</td>
<td>0–54</td>
<td>0.006</td>
<td>2.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>13</td>
<td>1.01</td>
<td>0.70–1.31</td>
<td>66</td>
<td>28–80</td>
<td>1.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panic disorder</td>
<td>16</td>
<td>1.16</td>
<td>0.81–1.51</td>
<td>75</td>
<td>55–83</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorder-specific outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>13</td>
<td>0.73</td>
<td>0.47–0.99</td>
<td>64</td>
<td>23–79</td>
<td>2.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>7</td>
<td>0.80</td>
<td>0.21–1.38</td>
<td>84</td>
<td>64–90</td>
<td>2.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panic attacks(2) wks</td>
<td>5</td>
<td>2.10</td>
<td>0.69–3.50</td>
<td>93</td>
<td>87–95</td>
<td>1.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI: Confidence interval; MD: mean difference (not standardized); Ncomp: number of comparisons; NNT: Numbers-needed-to-treat.

a The p-values in this column indicate whether the difference between the effect sizes in the subgroups is significant.
b The 95% CI of P cannot be calculated when the number of studies is two or smaller.
c The mean difference can not be calculated when outcomes from different measurement instruments are pooled.

d The p-values in this column indicate whether the difference between the effect sizes in the subgroups is significant.

e The 95% CI of P cannot be calculated when the number of studies is two or smaller.
f The mean difference can not be calculated when outcomes from different measurement instruments are pooled.

Fig. 2. Forrest plot of effect sizes according to the BAI in randomized controlled trials in generalized anxiety disorder, social anxiety disorder and panic disorder.

4. Discussion

We wanted to examine whether the effects of cognitive and behavioral psychotherapies on generic anxiety measures differed across three of the most prevalent anxiety disorders: generalized anxiety disorder, social anxiety disorder and panic disorder. We examined generic anxiety measures utilized across anxiety disorders, because disorder-specific measures do not allow for direct comparisons of the effects across disorders. We found that the effects as measured with the BAI on panic disorder were significantly larger than those on GAD and SAD, and we found no significant difference between SAD and GAD. The difference between the effects on panic disorder on the one hand, and GAD and SAD on the other remained significant after adjusting for baseline severity and other major characteristics of the trials. These findings are in line with earlier meta-analyses using disorder-specific outcomes, indicating large effects for treatments of panic disorder (Sánchez-Meca et al., 2010) and somewhat smaller effects for GAD (Cuijpers, Sijbrandij et al., 2014) and SAD (Eskildsen et al., 2010).
Our findings suggest that treatment of panic disorder may result in better outcomes than treatment of GAD and SAD. However, this suggestion should be considered with caution because this difference was only found for the BAI and not for other generic outcomes. Also, the quality of the included studies was not optimal and risk of bias was considerable in most studies. And finally, the number of studies in panic disorders was small and the risk of a chance finding considerable. On the other hand the difference between panic disorder and the other disorders was large and remained significant after adjusting for major characteristics of the studies.

We also compared the effect sizes found for generic outcomes with those of disorder-specific outcome measures and found no major differences between these two categories. This suggests that there are no major differences between the two types of outcome instruments, although this finding has to be considered with caution because a direct comparison between the two categories was not possible.

All three anxiety disorders that were examined in this meta-analysis seem to result in high effect sizes, regardless of whether these were disorder-specific or generic anxiety outcomes. NNTs between 2 and 3 were found for most outcomes and for panic a NNT of less than 1.5 was found, which suggests that treatments of panic are even more effective than treatments of GAD and SAD. That is in line with earlier meta-analyses (Sánchez-Meca et al., 2010), and still good news for patients and clinicians. Furthermore, it suggests that at a population level, the disease burden of anxiety disorders can be reduced considerably with existing treatments and that is even more likely for panic disorders.

The reasons why panic disorder may be more treatable than the other included anxiety disorders is not clear. It may be possible that general levels of anxiety are more affected when the number of panic attacks is reduced after treatment than when more generic problems like worry or anxiety in social situations are reduced, but this remains a matter of speculation. It is also possible that the effects of treatments on panic disorder are larger because the treat-
ments are better. It was not possible to examine this in more detail, however, because the components in the studies differed considerably and it was not possible to make more specific categories of therapies in this analysis.

The reason that this difference was found for the BAI, but not for other generic measures of anxiety also remains unclear. The BAI was developed specifically to measure anxiety so that it was differentiated from depression and this was not the case for the other measures. It is also a self-report measure, unlike the HAMA, which was also used in a large number of the studies included. The STAI measures are also self-report, though it is worth noting that the number of studies using the STAI was small. It is also possible that the BAI is more sensitive to panic symptoms, with one study suggesting that this scale appears to actually measure panic attacks rather than anxiety in general (Cox, Cohen, Direnfeld, & Swinson, 1996). However, another factorial analysis concluded that as a measure of treatment-related changes, the BAI was more related to general anxiety than to panic symptoms (de Beurs, Wilson, Chambless, Goldstein, & Feske, 1997).

This study found that all three anxiety disorders can be treated effectively in many patients, and that panic disorder may be treated even more effectively than the other two disorders. However, this is only based on comparing the effects of treatments on generic anxiety measures. Future research is needed to assess the relative outcomes of treatments on other important outcomes that are not directly related to anxiety, like quality of life, functional limitations and functioning in daily life. Such measures may give a more comprehensive overview of the outcomes of treatments on the disease burden of anxiety disorders. More research is also needed to improve the effects of treatments, especially for GAD and SAD, since even if the effects of treatments are generally positive, there is still a considerable number of patients that do not benefit. Furthermore, this meta-analysis focused on short-term outcomes, while long-term outcomes are more important for patients, clinicians and the burden of disorders from a public health point of view.

This study has several important limitations that should be considered. Several of these have been mentioned already, including the small number of studies in several subgroups, the high risk of bias in most studies, differences between therapies that could not be categorized, and the fact that all outcomes may not be captured by measures that are not disorder-specific. To these we add that these analyses were only conducted for outcomes at post-test and long-term effects were not considered.

Despite these limitations, it is interesting to note that cognitive and behavioral therapies may have a different impact on anxiety and the disease burden in patients. Although more research is needed to verify whether this is indeed true, such differential effects of treatments across disorders may be helpful in designing optimal strategies for reducing the disease burden of anxiety disorders.

Acknowledgement

Ioana A. Cristea was supported for this work by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-III-RU-TE-2014-4-1316 awarded to Ioana A. Cristea.

Appendix A. Full search string for Pubmed.


References


1 References marked with an asterisk are included in the meta-analysis.


