Efficacy of treatments for anxiety disorders: a meta-analysis

Borwin Bandelow\textsuperscript{a}, Markus Reitt\textsuperscript{a}, Christian Röver\textsuperscript{b}, Sophie Michaelis\textsuperscript{a}, Yvonne Görlich\textsuperscript{a} and Dirk Wedekind\textsuperscript{a}

To our knowledge, no previous meta-analysis has attempted to compare the efficacy of pharmacological, psychological and combined treatments for the three main anxiety disorders (panic disorder, generalized anxiety disorder and social phobia). Pre-post and treated versus control effect sizes (ES) were calculated for all evaluable randomized-controlled studies (\(n=234\)), involving 37 333 patients. Medications were associated with a significantly higher average pre-post ES [Cohen's \(d=2.02\) (1.90–2.15); 28 051 patients] than psychotherapies [1.22 (1.14–1.30); 6992 patients; \(P<0.0001\)]. ES were 2.25 for serotonin–noradrenaline reuptake inhibitors (\(n=23\) study arms), 2.15 for benzodiazepines (\(n=42\)), 2.09 for selective serotonin reuptake inhibitors (\(n=62\)) and 1.83 for tricyclic antidepressants (\(n=15\)). ES for psychotherapies were mindfulness therapies, 1.56 (\(n=4\)); relaxation, 1.36 (\(n=17\)); individual cognitive behavioural/exposure therapy (CBT), 1.30 (\(n=93\)); group CBT, 1.22 (\(n=18\)); psychodynamic therapy 1.17 (\(n=5\)); therapies without face-to-face contact (e.g. Internet therapies), 1.11 (\(n=34\)); exercise had an ES of 1.23 (\(n=3\)). For control groups, ES were 1.29 for placebo pills (\(n=111\)), 0.83 for psychological placebos (\(n=16\)) and 0.20 for waitlists (\(n=50\)). In direct comparisons with control groups, all investigated drugs, except for citalopram, opipramol and moclobemide, were significantly more effective than placebo. Individual CBT was more effective than waiting list, psychological placebo and pill placebo. When looking at the average pre-post ES, medications were more effective than psychotherapies. Pre-post ES for psychotherapies did not differ from pill placebos; this finding cannot be explained by heterogeneity, publication bias or allegiance effects. However, the decision on whether to choose psychotherapy, medications or a combination of the two should be left to the patient as drugs may have side effects, interactions and contraindications. *Int Clin Psychopharmacol* 00:000–000

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

International Clinical Psychopharmacology 2015, 00:000–000

Keywords: anxiety disorders, cognitive behavioural therapy, combined treatment, meta-analysis, pharmacological treatment

Introduction

According to treatment guidelines (Bandelow et al., 2008; National Institute for Health and Clinical Excellence (NICE), 2011; Baldwin et al., 2014), psychological therapies and psychopharmacological drugs have shown efficacy for the treatment of the three major anxiety disorders – panic disorder with or without agoraphobia (PDA), generalized anxiety disorder (GAD) and social anxiety disorder (SAD). Among psychotherapies, cognitive behavioural therapy (CBT) is the method studied most, but a number of trials have investigated relaxation, psychodynamic therapy (PDT\textsuperscript{h}), interpersonal therapy (IPT), eye movement desensitization reprocessing (EMDR), mindfulness meditation and therapies conducted through the Internet or computers.

Medications used for anxiety disorders include selective serotonin reuptake inhibitors (SSRIs), serotonin–noradrenaline reuptake inhibitors (SNRIs), pregabalin, tricyclic antidepressants and opipramol (TCAs), benzodiazepines, moclobemide, phenelzine, buspirone and hydroxyzine (Bandelow et al., 2008). The antipsychotic quetiapine has not been approved for GAD by the US Food and Drug Administration or the European Medicines Agency, but is licensed in a few countries for the treatment of this disorder.

The gold standard for meta-analyses is to compare the differences between an active treatment and a control condition as only randomized-controlled trials can control for unspecific effects such as placebo and expectation effects, tendency of regression to the mean, spontaneous remission or differences in measurement. However, these treated versus control effect sizes cannot be used for comparisons of psychological therapies and medications. In psychotherapy trials, the control condition is mostly a waiting list, which usually has a low pre-post effect size, whereas drugs are usually compared with a pill placebo, which has a larger effect (Rief et al., 2009). Moreover, patients are mainly interested in the question ‘How much will my anxiety improve with the treatment?’ which is reflected by the pre-post effect size. In
addition, more studies can be included in the pre–post analysis, namely, the studies that included two active treatments and no placebo condition.

Previous meta-analyses were mainly based on treated versus control effect sizes, did not include all three anxiety disorders, and/or were selective in including studies. Four meta-analyses only looked at the few available direct comparisons of CBT and drugs for PDA and found equal efficacy of CBT and medications (Bandelow et al., 2007; Cuijpers et al., 2013), superiority of the combination over monotherapies (Bandelow et al., 2007), superiority of CBT over medications (Roshanaci-Moghaddam et al., 2011), superiority of the combination to both monotherapies in the treatment phase and superiority to pharmacotherapy at follow-up (Furukawa et al., 2009). One meta-analysis for PDA found equal efficacy of CBT and drugs or superiority of CBT, respectively, depending on the calculation method used (Mitte, 2005), but this study compared the differences between treatments and control groups by using a correction factor and included drugs that were not available on the market. In an analysis of pre–post effect sizes for SAD, the largest effect sizes were found for medications, whereas treatment gains of CBT had longer durability after the termination of treatment (Fedoroff and Taylor, 2001).

To our knowledge, our meta-analysis represents the first evaluation of pre–post effect sizes for all treatments of all three anxiety disorders together. The rationale of combining all three disorders is that anxiety disorders have a high comorbidity among each other (Wittchen et al., 2011) and that there is no consistent evidence to suggest differential efficacy of treatments for different anxiety disorders. Moreover, previous meta-analyses showed large effect size heterogeneity among the studies despite the use of similar inclusion criteria (consecutive patients with DSM-defined anxiety disorders between 18 and 65 years of age) and similar outcome measures. Our approach was chosen to improve the statistical validity and to reduce the influence of heterogeneity by increasing the number of studies and to avoid multiple testing, which would have arisen if the three disorders were tested separately. Nevertheless, we also analysed the results for the three disorders separately and calculated the relative effect sizes of active treatments versus control conditions.

**Patients and methods**

**Selection of studies**

Randomized-controlled trials were selected with patients who fulfilled the criteria for PDA, GAD or SAD according to DSM-III or later versions. Treatments for specific phobias were not included as patients with these disorders rarely seek treatment. Journal articles were located using MEDLINE, ISI Web of Science and hand search [PRISMA statement (Moher et al., 2011); Fig. 1; Table, see Supplemental digital content, http://links.lww.com/ICP/A2]. After screening by title and abstract, full texts were assessed for eligibility. Study quality was assessed using the SIGN Statement (Scottish Intercollegiate Guidelines Network, 2012) (including randomization, blinding, standardized outcomes, dropouts and intent-to-treat analysis, results not shown because of space constraints). Reasons for exclusion were missing information, making it impossible to compute effect sizes, a sample size of any of the treatment arms at inclusion of less than 10, reports that were restricted to subsamples (e.g. only elderly patients) and studies with children and adolescents.

Drugs were included that had been shown to be effective in randomized-controlled studies and are licensed in at least some countries for the treatment of anxiety disorders (Bandelow et al., 2008): the SSRIs citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine and sertraline, the SNRIs duloxetine and venlafaxine, the calcium modulator pregabalin, the TCAs imipramine, clomipramine and the tricyclic anxiolytic opipramol, the benzodiazepines alprazolam, bromazepam, clobazam, clonazepam, delorazepam, diazepam and lorazepam, the antihistamine hydroxyzine, the irreversible monoamine oxidase inhibitor phenelzine, the reversible inhibitor of monoamine oxidase A (RIMA) moclobemide and the atypical antipsychotic quetiapine.

Psychological therapies were categorized as follows: ‘CBT/exposure’ included individual CBT or exposure techniques or a combination of both. ‘Relaxation’ comprised applied relaxation (Ost, 1988) and progressive muscle relaxation (Jacobson, 1938). ‘Non-face-to-face therapies’ include treatments conducted through the Internet, computers or self-help books, which were mostly based on CBT. The group ‘CBT + drug’ includes study arms with a combination of individual or group CBT/exposure with various drugs. Control conditions included pill placebo, waiting list and ‘psychological placebo’. The latter is used to differentiate between unspecific and specific psychotherapy effects and is defined as a treatment using sessions of the same length as a psychotherapy session, in which study staff establish a supportive, listening and nondirective relationship and/or provide psychoeducational instructions, but without applying any specific techniques such as exposure or cognitive restructuring.

**Meta-analytical procedure**

Two reviewers (B.B. and M.R.) independently extracted all data. Any discrepancies were discussed and resolved. To limit heterogeneity and to achieve maximum comparability, we preferably used the most commonly applied scales: the Hamilton Anxiety Scale (HAMA) (Hamilton, 1959) for PDA and GAD and the Liebowitz Social Anxiety Scale (LSAS) (Liebowitz, 1987) for SAD. We did not use a composite score of all kinds of outcome measures because many of the scales used in the trials were not designed to measure the core symptomatology.
of the anxiety disorder, but to cover other aspects (e.g., depression). The HAMA was available in 41% of the PDA studies and 78% of the GAD studies, whereas the LSAS was used in 68% of the SAD studies. If the HAMA or the LSAS were not available, other scales were used (for details, see Supplemental digital content, http://links.lww.com/ICP/A2). The selected scale was not necessarily the primary efficacy measure (PEM) of a study as there was a high variability in the choice of the PEM and the PEM was not provided in many studies.

Whenever available, intent-to-treat (ITT) data were used, which were usually based on the last observation carried forward method for missing values.

First, pre–post effect sizes were calculated. Prescores and postscores are not independent, and if their correlation is not accounted for, the significance of effect sizes may be overestimated. However, as this correlation was not provided in any of the studies selected for the current analysis, Cohen’s $d$ was calculated by subtracting the pretreatment mean from the posttreatment mean and dividing the difference by the pretreatment SD of the measure (Dunlap et al., 1996). If there was more than one treatment group in a study, a pooled baseline SD based on all treatment arms in the study was used. If the study failed to report the pretreatment SD, it was imputed from all the other available studies for the specific anxiety disorder that used the same scale (which was the case in 3.0% of the studies). Where a study only reported data from dichotomous outcomes (the proportion of respondents to treatment, e.g. defined by a 50% reduction on the HAMA), it was assumed that participants who ceased to engage in the study from whatever group had an unfavourable outcome. Odds ratios were transformed into

PRISMA statement.
Cohen’s $d$ (Borenstein et al., 2009). We calculated the pre–post effect sizes for all anxiety disorders together and then separately. We did not include open studies because these may have been influenced by expectation effects.

Second, the effect sizes for the comparisons drugs versus pill placebo and psychological treatments versus waiting list, psychological placebo or pill placebo were calculated using the difference between pretreatment and post-treatment scores for the active group minus the difference between pretreatment and post-treatment scores for the control group divided by the SD before therapy pooled from all treatments in the study. In studies using at least two arms with the same drug, the average effect size of all arms was calculated.

The random-effects model was used, in which studies are weighted on the basis of the inverse variances and an additional variance component reflecting the observed heterogeneity between studies. Analyses were carried out using Comprehensive Meta-Analysis Version 2.2.064 (Biostat, Englewood, New Jersey, USA). $T$-tests for independent samples (two-tailed) were used for significance testing of effect size differences.

$I^2$ was determined as an indicator for heterogeneity. Because moderate ($I^2 > 50\%$) to high ($>75\%$) heterogeneity was found for most comparisons, the random-effects model was used in all analyses. In general, including a random effect will lead to more conservative results than the fixed-effects model. Moreover, as fixed-effects meta-analyses are dominated by larger studies, the random model accounts better for the large differences in the average sample sizes between medication and psychotherapy studies.

To assess publication biases, ‘fail-safe $N$s’ (number of negative studies needed to reject the hypothesis that a treatment differs from a control group), $P$ values for Egger’s regression intercept (a method to quantify the bias captured by a funnel plot) for between-group effect sizes and effect sizes adjusted for publication bias using Duval and Tweedie’s (Duval and Tweedie, 2000) ’trim and fill’ method, which is used to correct the funnel plot by estimating the number of missing studies, and the effect sizes of these studies were determined.

Possible allegiance effects for all study arms were analysed by two independent raters and were assumed when a medication study was sponsored by the current manufacturer of the investigated drug, when authors disclosed financial support from the manufacturer or when one of the authors was a staff member of the manufacturer. In studies with more than two active treatments, allegiance effects were only assigned to the product associated with the medication. For psychological treatments, allegiance effects were assumed when authors had developed the treatment, contributed towards an aetiological model or published manuals for the treatment. Effect sizes were calculated separately for subgroups with or without allegiance effects.

Abbreviations are listed in the Supplemental digital content, http://links.lww.com/ICP/A2.

**Results**

A total of 232 papers including 234 studies with 37,333 patients were found to be eligible (Fig. 1, Table 1, and list of all included studies, Table, Supplemental digital content, http://links.lww.com/ICP/A2).

**Pre–post effect sizes**

In Fig. 2, the pre–post effect sizes are shown for all treatment groups. Table 2 shows the effect sizes for all treatments. Because of the problem of multiple testing, statistical tests were not performed for all possible comparisons. However, when the confidence intervals of two treatments do not overlap, the difference is significant. Numerically, among psychotherapies, mindfulness meditation yielded the highest effect size. Relaxation treatments were numerically more effective than individual behavioural treatments (CBT and exposure), which was more effective than group CBT. Exercise, non-face-to-face therapies, PDT, EMDR and IPT showed lower pre–post effect sizes. Among medication treatments, delorazepam, quetiapine and hydroxyzine had the highest effect sizes numerically, but these results were only based on a few studies.

When the pre–post effect sizes of all medications were pooled and compared with all psychotherapies, significantly higher efficacy was found for medications ($d = 2.02$; 95% confidence interval 1.90–2.15; 28,051 patients in 206 study arms) than for psychotherapies ($d = 1.22$, 1.14–1.30; 6922 patients in 184 study arms, $t = 84.2$; $d.f. = 37332$; $P < 0.0001$). Pill placebo was significantly more effective than psychological placebo and waiting list (Table 2). The average pre–post effect sizes for pill placebos were shown to have increased over the past decades, from 0.87 in the years 1983–1992 to 1.71 for the years 2003–2013 (Fig. 3).

**Active versus control effect sizes**

In direct comparisons, individual CBT/exposure was significantly more effective than waiting list, psychological placebo and pill placebo conditions (Table 3). Group CBT was superior to waiting list, but not to pill placebo. Relaxation and non-face-to-face therapies were superior to waiting lists. In one available study for each comparison, PDT, IPT and mindfulness therapy did not differ from a psychological placebo (Knijnik et al., 2004; Lipsitz et al., 2008; Hoge et al., 2013), whereas PDT (Leichsenring et al., 2013) and IPT (Stangier et al., 2011) were superior to waiting lists in single studies.

All drugs, with the exception of citalopram, moclobemide and opipramol, were significantly different from placebo (Table 4). Large between-group effect sizes were found for phenelzine, clomipramine and lorazepam.
Considerable heterogeneity was found (Tables 3 and 4). Therefore, the random-effects model was considered adequate and was used throughout. Tables 3 and 4 show indicators for publication bias (fail-safe $N$, $P$ scores for Egger’s regression intercept). After adjusting effect sizes for publication bias, none of the significant results became nonsignificant.

For psychological therapies, possible allegiance effects were found in 32.6% of 175 study arms (allegiance effects were undeterminable in 2.9%); the average pre–post effect sizes were 1.29 (1.16–1.43) for studies with allegiance effects and 1.19 (1.09–1.30) for those without (NS). For drug therapy arms, possible allegiance effects were found in 50.8% of 187 study arms (undeterminable in 1.3%). Effect sizes for studies with allegiance effects [2.00 (1.81–2.19)] did not differ from those without such effects [2.01 (1.82–2.20); NS].

To determine whether patients in psychotherapy and drug studies were comparable with respect to severity of illness, we looked at baseline scores of the two most commonly used scales. Baseline HAMA scores were significantly lower ($22.1 \pm 4.76$) in psychotherapy studies than in drug studies ($24.2 \pm 2.92$; $t = 2.97$, $P = 0.003$). Endpoint scores did not differ (psychotherapy $11.8 \pm 3.03$; drugs $12.06 \pm 2.54$; $t = 0.57$, NS). For SAD, baseline LSAS scores were $73.2 \pm 11.0$ in psychotherapy studies and $84.0 \pm 8.3$ in drug studies ($t = 3.93$, $P < 0.001$). Endpoint scores did not differ between psychotherapy ($49.7 \pm 10.7$) and drug studies ($51.3 \pm 10.2$).

Table 1 Number of studies and patients included

<table>
<thead>
<tr>
<th>Anxiety disorder</th>
<th>Psychotherapy</th>
<th>Medications</th>
<th>Combinations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$ studies</td>
<td>$N$ patients</td>
<td>$N$ studies</td>
<td>$N$ patients</td>
</tr>
<tr>
<td>Panic disorder/agoraphobia</td>
<td>40</td>
<td>2656</td>
<td>35</td>
<td>8524</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>22</td>
<td>1487</td>
<td>45</td>
<td>12 280</td>
</tr>
<tr>
<td>Social anxiety disorder</td>
<td>34</td>
<td>2799</td>
<td>30</td>
<td>7247</td>
</tr>
<tr>
<td>Total</td>
<td>94</td>
<td>6922</td>
<td>110</td>
<td>28 051</td>
</tr>
</tbody>
</table>

$N$ patients, post-treatment/ITT sample sizes.
Combinations = comparisons of psychological therapies and medication and their combination.

Fig. 2

Treatments for anxiety disorders (all anxiety disorders pooled). Pre–post effect sizes (Cohen’s $d$) and 95% confidence intervals. Black: drugs; white: psychological therapies; grey: control groups. Confidence intervals: see Table 2. $n$, number of studies. CBT, cognitive behavioural therapy; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.
and drug studies (50.1 ± 12.9, t = 0.12, NS). When only PDA and GAD studies with a minimum HAMA baseline score of 22 were analysed separately, medications still showed significantly higher effect sizes (d = 2.52; confidence interval 2.35–2.69; n = 95) than psychotherapies (d = 1.71, 1.34–2.08; n = 15; P < 0.0005).

Drug studies were significantly shorter (9.2 ± 4.4 weeks on average) than psychotherapy studies (12.4 ± 5.5 weeks; t = 4.6; df = 202; P < 0.0001). PDTTh studies had a mean duration of 21.0 ± 11.4 weeks. On average, psychological therapy trials had smaller sample sizes (mean patient number 29.7 ± 25.5 per study arm) than drug trials (109.7 ± 59.6). In 76.6% of the psychotherapy studies, patients were not excluded when they were on ongoing medication, and in an additional 8.5% of such studies, no information on the exclusion of medicated patients was provided. In 84.4% of the drug studies, in 71.4% of the combination studies and in 52.1% of the psychological treatment studies, ITT analyses were used for missing data.

**Discussion**

To our knowledge, this is the first meta-analysis comparing the absolute (pre–post) and relative (treated vs. control) effect sizes of all commonly used treatments for the three major anxiety disorders.

Some guidelines for the treatment of anxiety disorders, for example, that of the UK National Institute for Health and Clinical Excellence (NICE, 2011, 2013) evaluate psychotherapy as equivalent or even superior to drug therapy. However, the main result of our analysis is that most
psychopharmacological drugs used for anxiety disorders have markedly higher effect sizes than psychological therapies, and

the gains were achieved in a shorter time. Patients enrolled in psychotherapy studies were significantly less severely ill than patients recruited for drug studies. However, it is unlikely that the superiority of medications was because of the higher baseline severity in drug studies as a separate analysis of the studies with higher baseline severity still showed a significant difference in favour of medications.

Among the drugs that are currently most commonly used, licensed and recommended by guidelines, the SSRIs, the SNRIs and pregabalin, had the highest effect sizes numerically. High effect sizes were also found for delorazepam, hydroxyzine and quetiapine; however, the results with these two drugs were only based on a few studies. Quetiapine is only licensed in a few countries for the treatment of GAD.

The choice of a drug should not only be made solely on the basis of efficacy but also on possible side effects, contraindications and interactions. Benzodiazepines may cause dependency and are therefore not recommended for routine use. Tricyclic antidepressants have more adverse events than SSRIs according to direct comparisons (Bandelow et al., 2008). Pregabalin has been associated

with GAD.

![Fig. 3](image_url)

Pre–post effect sizes for pill placebo arms (n=111) by year of publication. Publication year and placebo effect sizes were significantly correlated (r=0.41; P<0.001). Regression analysis showed a function of effect size (d=0.058 x year − 111.1).

### Table 3 Psychological therapies versus control conditions (treated vs. control effect sizes)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>n</th>
<th>d</th>
<th>CI</th>
<th>P</th>
<th>P²</th>
<th>Fail-safe N</th>
<th>Egger P</th>
<th>Adjusted d</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBT (individual) vs. waiting list</td>
<td>25</td>
<td>1.23</td>
<td>1.02–1.45</td>
<td>&lt;0.0001</td>
<td>70.0</td>
<td>2701</td>
<td>0.44 (NS)</td>
<td>1.23</td>
<td>1.02–1.45</td>
</tr>
<tr>
<td>CBT (individual) vs. psychological placebo</td>
<td>12</td>
<td>0.75</td>
<td>0.34–1.16</td>
<td>&lt;0.0001</td>
<td>83.5</td>
<td>168</td>
<td>0.033</td>
<td>0.26</td>
<td>−0.19 to 0.72</td>
</tr>
<tr>
<td>CBT (individual) vs. pill placebo</td>
<td>9</td>
<td>0.37</td>
<td>0.20–0.94</td>
<td>0.003</td>
<td>64.8</td>
<td>49</td>
<td>0.023</td>
<td>0.24</td>
<td>−0.19 to 0.66</td>
</tr>
<tr>
<td>CBT (group) vs. waiting list</td>
<td>7</td>
<td>1.33</td>
<td>1.06–1.60</td>
<td>&lt;0.0001</td>
<td>0.0</td>
<td>157</td>
<td>0.47 (NS)</td>
<td>1.26</td>
<td>1.01–1.51</td>
</tr>
<tr>
<td>CBT (group) vs. pill placebo</td>
<td>5</td>
<td>0.57</td>
<td>−0.10 to 0.35</td>
<td>0.28 (NS)</td>
<td>−</td>
<td>−</td>
<td>0.27 (NS)</td>
<td>0.12</td>
<td>−0.10 to 0.35</td>
</tr>
<tr>
<td>Relaxation vs. waiting list</td>
<td>6</td>
<td>1.31</td>
<td>0.92–1.71</td>
<td>&lt;0.0001</td>
<td>41.3</td>
<td>114</td>
<td>0.054*</td>
<td>1.21</td>
<td>0.80–1.63</td>
</tr>
<tr>
<td>Non-face-to-face therapies vs. waiting list</td>
<td>20</td>
<td>0.91</td>
<td>0.73–1.09</td>
<td>&lt;0.0001</td>
<td>53.6</td>
<td>1108</td>
<td>0.43 (NS)</td>
<td>0.91</td>
<td>0.73–1.09</td>
</tr>
</tbody>
</table>

Adjusted d, Cohen’s d adjusted for publication bias (Duval and Tweedie’s trim and fill method); CBT, cognitive behavioural therapy; CI, confidence interval; d, Cohen’s d; Egger P, probability score for Egger’s regression intercept; P², heterogeneity; n, number of study arms.

*Trend towards statistical significance.

### Table 4 Medications versus pill placebo (treated vs. control effect sizes)

<table>
<thead>
<tr>
<th>Drugs</th>
<th>n</th>
<th>CI low</th>
<th>P</th>
<th>P²</th>
<th>Fail-safe N</th>
<th>Egger P</th>
<th>Adjusted d</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alprazolam</td>
<td>9</td>
<td>0.37–0.95</td>
<td>&lt;0.0001</td>
<td>64.4</td>
<td>123</td>
<td>0.24</td>
<td>0.49</td>
<td>0.18–0.81</td>
</tr>
<tr>
<td>Buspiroprone</td>
<td>5</td>
<td>0.05–0.64</td>
<td>0.02</td>
<td>52.7</td>
<td>12</td>
<td>0.29</td>
<td>0.34</td>
<td>0.05–0.64</td>
</tr>
<tr>
<td>Citalopram</td>
<td>2</td>
<td>−0.13 to 0.59</td>
<td>0.20 (NS)</td>
<td>75.8</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Clomipramine</td>
<td>5</td>
<td>0.43–1.31</td>
<td>&lt;0.0001</td>
<td>77.3</td>
<td>78</td>
<td>0.29</td>
<td>0.529</td>
<td>0.04–1.02</td>
</tr>
<tr>
<td>Clonazepam</td>
<td>4</td>
<td>0.37–0.90</td>
<td>&lt;0.0001</td>
<td>48.2</td>
<td>51</td>
<td>0.20</td>
<td>0.581</td>
<td>0.31–0.85</td>
</tr>
<tr>
<td>Diazepam</td>
<td>5</td>
<td>0.38–1.14</td>
<td>&lt;0.0001</td>
<td>69.3</td>
<td>83</td>
<td>0.31</td>
<td>0.611</td>
<td>0.18–1.04</td>
</tr>
<tr>
<td>Duloxetine</td>
<td>3</td>
<td>0.62–0.65</td>
<td>&lt;0.0001</td>
<td>0.0</td>
<td>60</td>
<td>0.003</td>
<td>0.53</td>
<td>0.42–0.66</td>
</tr>
<tr>
<td>Escitalopram</td>
<td>6</td>
<td>0.26–0.63</td>
<td>&lt;0.0001</td>
<td>77.5</td>
<td>148</td>
<td>0.64</td>
<td>0.45</td>
<td>0.26–0.63</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>4</td>
<td>0.07–1.05</td>
<td>0.025</td>
<td>80.1</td>
<td>42</td>
<td>0.001</td>
<td>0.56</td>
<td>0.07–1.05</td>
</tr>
<tr>
<td>Fluvoxamine</td>
<td>12</td>
<td>0.43–0.78</td>
<td>&lt;0.0001</td>
<td>58.3</td>
<td>342</td>
<td>0.18</td>
<td>0.56</td>
<td>0.37–0.75</td>
</tr>
<tr>
<td>Hydroxyzine</td>
<td>4</td>
<td>0.41–1.17</td>
<td>&lt;0.0001</td>
<td>77.8</td>
<td>72</td>
<td>0.32</td>
<td>0.49</td>
<td>0.07–0.91</td>
</tr>
<tr>
<td>Imipramine</td>
<td>4</td>
<td>0.46–1.01</td>
<td>&lt;0.0001</td>
<td>0.0</td>
<td>22</td>
<td>0.38</td>
<td>0.74</td>
<td>0.46–1.01</td>
</tr>
<tr>
<td>Lorazepam</td>
<td>3</td>
<td>0.13–1.61</td>
<td>0.022</td>
<td>92.6</td>
<td>49</td>
<td>0.25</td>
<td>0.87</td>
<td>0.13–1.61</td>
</tr>
<tr>
<td>Moclobemide</td>
<td>5</td>
<td>−0.23 to 0.56</td>
<td>0.4 (NS)</td>
<td>86.5</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Omprazolm</td>
<td>1</td>
<td>−0.07 to 0.48</td>
<td>0.14 (NS)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>19</td>
<td>0.41–0.70</td>
<td>&lt;0.0001</td>
<td>85.6</td>
<td>1603</td>
<td>0.06</td>
<td>0.42</td>
<td>0.26–0.57</td>
</tr>
<tr>
<td>Phenelzine</td>
<td>6</td>
<td>0.50–1.31</td>
<td>&lt;0.0001</td>
<td>49.5</td>
<td>79</td>
<td>0.90</td>
<td>0.96</td>
<td>0.60–1.31</td>
</tr>
<tr>
<td>Pregabalin</td>
<td>8</td>
<td>0.37–0.74</td>
<td>&lt;0.0001</td>
<td>78.0</td>
<td>314</td>
<td>0.92</td>
<td>0.55</td>
<td>0.37–0.74</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>3</td>
<td>0.47–0.66</td>
<td>&lt;0.0001</td>
<td>0.0</td>
<td>106</td>
<td>0.37</td>
<td>0.56</td>
<td>0.47–0.66</td>
</tr>
<tr>
<td>Sertaline</td>
<td>6</td>
<td>0.38–0.70</td>
<td>&lt;0.0001</td>
<td>60.0</td>
<td>169</td>
<td>0.23</td>
<td>0.54</td>
<td>0.38–0.70</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>19</td>
<td>0.31–0.69</td>
<td>&lt;0.0001</td>
<td>90.8</td>
<td>1305</td>
<td>0.55</td>
<td>0.33</td>
<td>0.14–0.52</td>
</tr>
</tbody>
</table>

Adjusted d, Cohen’s d adjusted for publication bias; CI, confidence interval; d, Cohen’s d; Egger P, probability score for Egger’s regression intercept; P², heterogeneity; n, number of study arms.
with withdrawal symptoms and abuse in individuals with multisubstance abuse; however, the relative abuse potential, compared with other medications, has not been established (Baldwin et al., 2013). Quetiapine, like other antidepressants, has been associated with a risk of metabolic abnormalities (Jin et al., 2004). Further, it has to be noted that some drugs analysed in our study are only licensed for one or two of the three anxiety disorders in many countries.

In direct comparisons, all drugs were superior to placebo, with the exception of citalopram, moclobemide and opipramol. The discrepancy between our findings and the superiority of these three drugs in the original studies can be explained by the fact that the PEMs used in these studies differed from the standardized outcome criteria used in our analysis.

For some drugs, for example, phenelzine, a discrepancy between high active-control difference effect sizes and relatively low pre–post effect sizes was found. This is probably because of the large heterogeneity of the studies and the fact that placebo effect sizes have increased markedly over the years; thus, drugs investigated in the 1980s and 1990s showed higher active-control differences.

In direct comparisons, CBT was significantly more effective than waiting list, psychological placebo and pill placebo control conditions. However, when looking at pre–post effect sizes, psychotherapies did not differ from pill placebos. This surprising finding cannot be explained by heterogeneity, publication bias or allegiance effects.

CBT should preferably be offered as individual treatment as group CBT was not superior to psychological placebo conditions in direct comparisons. For PDTth, despite its widespread use, the few existing studies showed relatively low pre–post effect sizes, despite the long average study length of 21 weeks.

Psychological therapies were compared with waiting lists in 70%, a psychological placebo in 14% and a pill placebo in 16% of the studies analysed. There has been some debate on whether it is possible at all to create a psychological control procedure that does not have any specific effects (Borkovec and Sibrava, 2005). However, as psychological placebo was significantly less effective than pill placebo in our study, it seems unlikely that a psychological placebo exerts considerable specific effects. As individual CBT was significantly superior to psychological placebo, this control procedure seems to be fair and feasible. The criterion ‘better than waiting list’ may be overly liberal as treatments that are markedly less effective than psychological or pill placebos would also fall into this category. Nevertheless, definitions of attention placebos differed widely among the studies, thus making comparisons difficult.

It is claimed that gains from CBT are maintained after termination of treatment, whereas patients receiving drugs experience a recurrence of anxiety symptoms after stopping medication. This would offer CBT an advantage over drug treatment. We intend to carry out another meta-analysis to investigate whether or not this is the case. Few controlled studies corroborate this opinion at present (Bandelow et al., 2007).

Large effect sizes were found for the few studies of combinations of CBT and various drugs. However, combination studies of CBT with currently recommended medications are still lacking.

Possible allegiance effects were found more frequently in drug than in psychotherapy studies. However, possible allegiance of the investigator did not alter effect sizes significantly.

**Limitations**

Large effect size heterogeneity was found among the studies, despite clear diagnoses, similar inclusion criteria and the use of standard scales in the majority of studies. Therefore, the results have to be interpreted with caution, but the strength of this study was the inclusion of a large number of studies by combining the three anxiety disorders to decrease the influence of heterogeneity.

Head-to-head RCTs of the different modalities are more reliable than meta-analyses. However, there are only a few such comparison trials. Studies comparing medications and psychotherapy mostly used drugs that are no longer first line in the treatment of anxiety disorders, for example benzodiazepines, monoamine oxidase inhibitors or tricyclic antidepressants.

Pre–post effect sizes may exaggerate the true treatment effect as they also comprise effects of expectancy, attention, regression to the mean and spontaneous remission. However, as these unspecific effects play a role in all evaluated treatments and the effect sizes of control groups differ largely, pre–post effects are better suited for a comparison of pharmacological and psychological treatments than treated versus control effects.

Psychological treatment studies had a longer average duration than drug studies. It is in the nature of psychotherapy that its effects do not occur immediately. Therefore, it would not have been feasible to use study length as a moderator variable.

Although ITT analysis was used in 84% of the drug studies, this was the case in only 52% of the psychotherapy studies. Moreover, blinding is difficult in psychotherapy studies, which may have affected effect sizes (Hrobjartsson et al., 2013). In most of the psychotherapy studies, patients were not excluded if they were on medication as long as it was not changed during the trial. Although this may not have affected the between-group effect sizes – as the condition was the same in the control groups – this may have led to an overestimation of the contribution of the psychological treatment towards the pre–post effect size.
Publication biases have not only been found in drug studies but also in psychotherapy trials (Cuijpers et al., 2010; Flint et al., 2014). Psychotherapy trials usually have markedly smaller sample sizes than drug trials and there is a risk that small studies with negative results are not published. We tested all treatments for publication bias and could not find any distortions that might have changed the main direction of our results. However, the tests applied for funnel plot asymmetry are not sufficient to exclude publication bias (Sterne et al., 2000) as it was shown that larger negative trials were often not published either, and more often so for industry-sponsored trials (Jones et al., 2013). We did not attempt to track unpublished trials with negative results because we included trials from the year 1983 onwards and trial registers for drug studies have only been available since 1997 (and their use is still not mandatory in Europe). Although the use of trial registers for psychotherapy studies has increased in the last few years, it is not widespread and not mandatory.

Conclusion

The decision on whether to choose psychotherapy, medications or a combination of the two should be left to the patient as drugs may have side effects, interactions and contraindications. In addition, costs, waiting periods and treatment duration have to be considered. Nevertheless, patients should be informed about the considerable differences in the pre–post effect sizes of treatments offered and their risk–benefit ratio.

Acknowledgements


Conflicts of interest

In the last 12 months and in the near future, Dr Bandelow has been/will be on the speakers/advisory board for Meiji-Seika, Lundbeck, Pfizer, and Servier. Dr Wedekind was on the speakers’ board of AstraZeneca, Essex Pharma, Lundbeck and Servier. For the remaining authors there are no conflicts of interest.

References


