Efficacy of treatments for anxiety disorders: a meta-analysis

Borwin Bandelow[a], Markus Reitt[a], Christian Röver[b], Sophie Michaelis[a], Yvonne Görlich[a] and Dirk Wedekind[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 ($n = 34$); eye movement desensitization reprocessing, 1.03 ($n = 3$); and interpersonal therapy 0.78 ($n = 4$). The ES was 2.12 ($n = 16$) for CBT/drug combinations. 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

International Clinical Psychopharmacology 2015, **00**:000–000

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

Departments of [a]Psychiatry and Psychotherapy and [b]Medical Statistics, University of Göttingen, Göttingen, Germany

Correspondence to Borwin Bandelow, Department of Psychiatry and Psychotherapy, University of Göttingen, von-Siebold-Str. 5, D-37075 Göttingen, Germany
Tel: +49 551 396607; fax: +49 551 396597; e-mail: bbandel@gwdg.de

Received 14 January 2015 **Accepted** 19 March 2015

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 (PDTh), 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

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.intclinpsychopharm.com*).

DOI: 10.1097/YIC.0000000000000078

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 (Roshanaei-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.

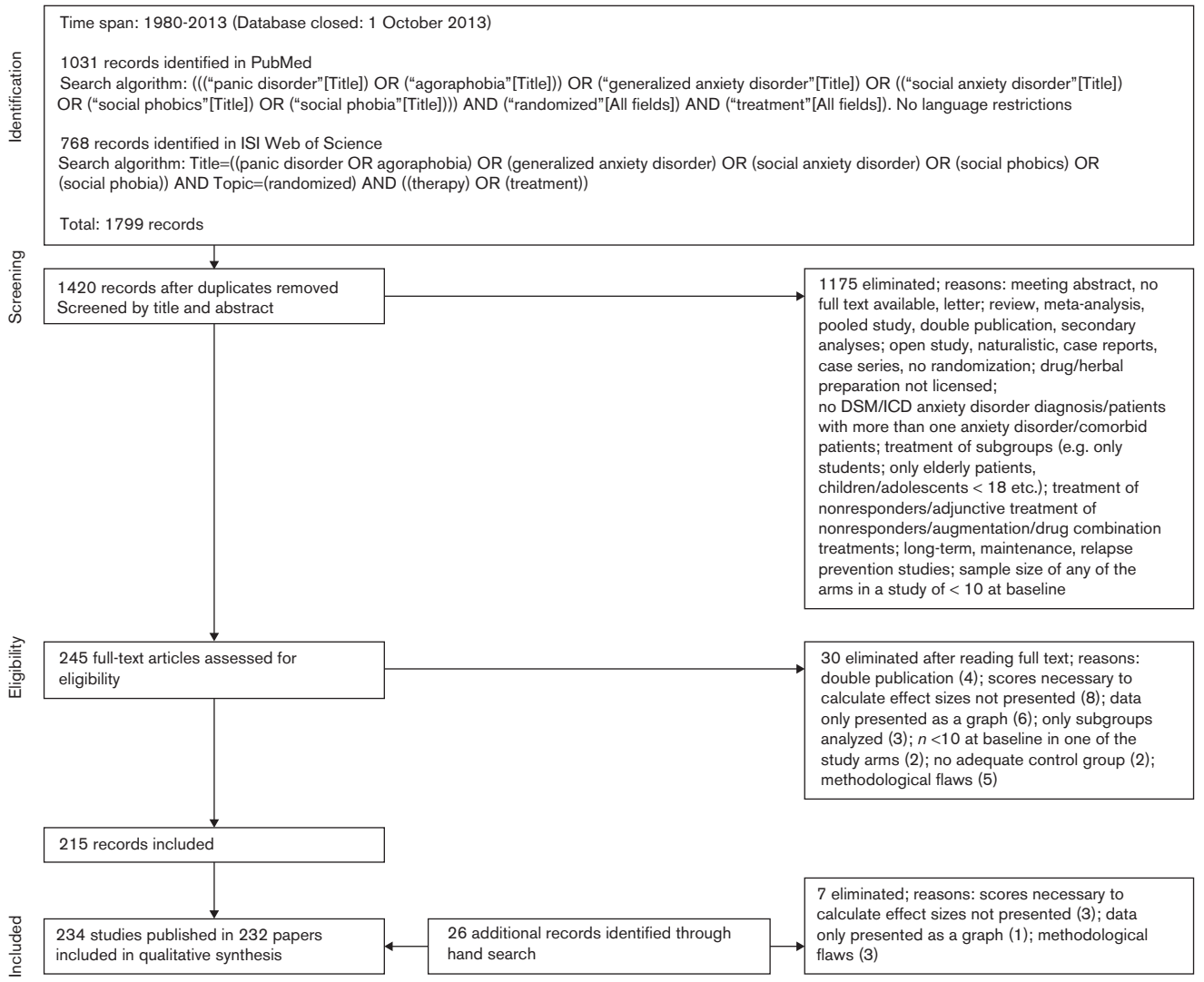
Drugs were included that had been shown to be effective in randomized-controlled studies and are licensed in at least some countries for the treatment of anxiety disorders (Bandelow *et al.*, 2008): the SSRIs citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine and sertraline, the SNRIs duloxetine 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

Fig. 1



PRISMA statement.

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 posttreatment mean from the pretreatment 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 responders 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

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 manufacturer. 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, PDTh, 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, $t = 84.2$; $d.f. = 37\,332$; $P < 0.0001$).

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, PDTh, IPT and mindfulness therapy did not differ from a psychological placebo (Knijnik *et al.*, 2004; Lipsitz *et al.*, 2008; Hoge *et al.*, 2013), whereas PDTh (Leichsenring *et al.*, 2013) and IPT (Stangier *et al.*, 2011) were superior to waiting lists in single studies.

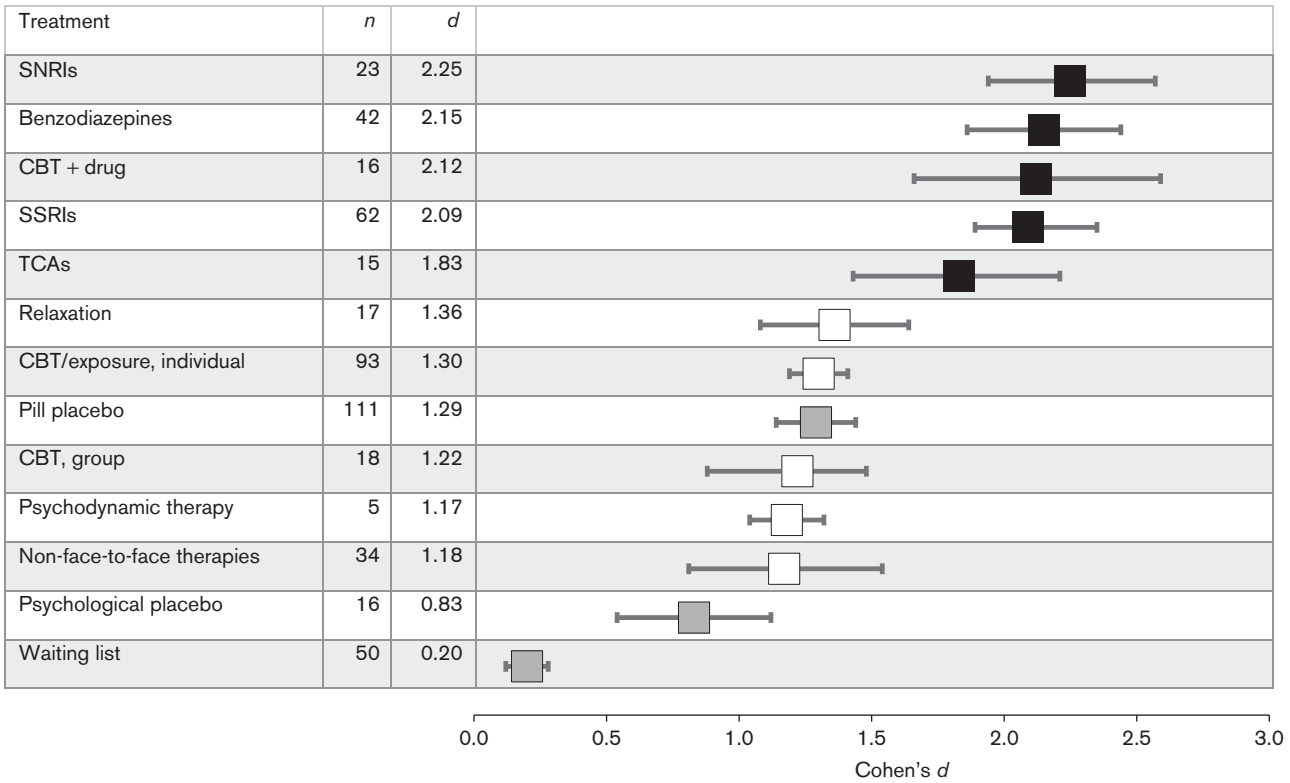
All drugs, with the exception of citalopram, moclobemide and opipramol, were significantly different from placebo (Table 4). Large between-group effect sizes were found for phenelzine, clomipramine and lorazepam.

Table 1 **Number of studies and patients included**

	Psychotherapy		Medications		Combinations		Total	
Anxiety disorder	*N* studies	*N* patients	*N* studies	*N* patients	*N* studies	*N* patients	*N* studies	*N* patients
Panic disorder/agoraphobia	40	2656	35	8524	15	1137	90	12 317
Generalized anxiety disorder	22	1467	45	12 280	4	203	71	13 950
Social anxiety disorder	34	2799	30	7247	9	1020	73	11 066
Total	94	6922	110	28 051	28	2360	234	37 333

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.

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 ± 4.76) in psychotherapy studies than in drug studies (24.2 ± 2.92; $t = 2.97$, $P = 0.003$). Endpoint scores did not differ (psychotherapy 11.8 ± 3.03; drugs 12.06 ± 2.54; $t = 0.57$, NS). For SAD, baseline LSAS scores were 73.2 ± 11.0 in psychotherapy studies and 84.0 ± 8.3 in drug studies ($t = 3.93$, $P < 0.001$). Endpoint scores did not differ between psychotherapy (49.7 ± 10.7)

Table 2 Pre–post effect sizes (Cohen's *d*)

	All anxiety disorders			Panic disorder			Generalized anxiety disorder			Social phobia		
Treatment	*n*	*d*	CI	*n*	*d*	CI	*n*	*d*	CI	*n*	*d*	CI
Psychological therapies												
CBT/exposure												
Individual	93	1.30	1.19–1.41	47	1.24	1.10–1.39	20	1.81	1.47–2.15	26	1.10	0.93–1.28
Group	18	1.22	0.95–1.49	4	1.81	1.50–2.12	1	1.63	0.97–2.28	13	1.01	0.72–1.29
EMDR	3	1.03	0.53–1.53	3	1.03	0.53–1.53	–	–	–	–	–	–
IPT	4	0.78	0.54–1.01	1	0.56	0.13–1.00	–	–	–	3	0.87	0.59–1.15
Mindfulness	4	1.56	1.20–1.92	–	–	–	1	1.16	0.73–1.60	3	1.75	1.40–2.09
Non-face-to-face therapies	33	1.11	0.98–1.23	9	1.21	1.01–1.40	6	1.15	0.95–1.36	18	1.04	0.92–1.16
Psychodynamic therapy	5	1.17	0.81–1.54	2	0.97	0.58–1.36	1	2.11	1.45–2.76	2	1.02	0.82–1.22
Relaxation	17	1.36	1.08–1.64	8	1.39	0.89–1.90	9	1.34	1.00–1.68	–	–	–
Medications												
Benzodiazepines	42	2.15	1.86–2.44	15	1.29	0.99–1.59	24	2.75	2.47–3.03	3	1.19	0.58–1.80
Alprazolam	16	1.79	1.35–2.23	9	1.10	0.71–1.49	6	2.98	2.25–3.71	1	1.23	0.36–2.11
Bromazepam	2	2.86	2.21–3.52	–	–	–	2	2.86	2.21–3.52	–	–	–
Clobazam	1	1.59	1.03–2.15	–	–	–	1	1.59	1.03–2.15	–	–	–
Clonazepam	5	1.78	1.09–2.47	3	2.61	1.17–4.04	–	–	–	2	1.17	0.28–2.06
Delorazepam	1	3.54	2.60–4.48	–	–	–	1	3.54	2.60–4.48	–	–	–
Diazepam	9	2.46	1.96–2.95	2	1.43	0.90–1.96	7	2.74	2.30–3.19	–	–	–
Lorazepam	8	2.44	1.83–3.05	1	0.95	0.41–1.49	7	2.87	2.24–3.14	–	–	–
Buspirone	7	1.35	0.75–1.95	2	0.82	0.13–1.50	5	1.56	0.78–2.33	–	–	–
Hydroxyzine	4	2.56	1.92–3.21	–	–		4	2.56	1.92–3.21	–	–	–
Moclobemide	9	1.47	0.94–2.00	3	1.51	0.19–2.84	–	–	–	6	1.51	0.87–2.16
Phenelzine	6	1.42	1.04–1.81	–	–	–	–	–	–	6	1.42	1.04–1.81
Pregabalin	8	2.30	1.71–2.89	–	–	–	6	2.66	2.14–3.18	2	1.22	1.04–1.40
Quetiapine	3	3.39	3.19–3.60	–	–	–	3	3.39	3.19–3.60	–	–	–
SNRIs	23	2.25	1.94–2.57	5	2.43	1.54–3.31	13	2.47	2.09–2.84	5	1.50	1.00–2.01
Duloxetine	3	1.95	1.69–2.20	–	–	–	3	1.95	1.69–2.20	–	–	–
Venlafaxine	20	2.32	1.94–2.70	5	2.43	1.54–3.31	10	2.70	2.17–3.22	5	1.50	1.00–2.01
SSRIs	62	2.09	1.89–2.35	25a	1.59	1.32–1.86	15	3.48	3.18–3.78	22	1.70	1.50–1.90
Citalopram	2	1.06	0.41–1.71	2	1.06	0.41–1.71	–	–	–	–	–	–
Escitalopram	8	2.75	2.09–3.41	1	0.89	0.63–1.15	5	3.34	2.78–3.90	2	2.26	1.83–2.69
Fluoxetine	7	1.69	1.16–2.22	4	2.27	1.10–3.43	–	–	–	3	1.30	0.64–1.96
Fluvoxamine	12	1.53	1.24–1.83	6	1.26	0.79–1.74	–	–	–	6	1.77	1.41–2.13
Paroxetine	23	2.42	2.03–2.82	8	2.16	1.56–2.77	7	3.46	3.23–3.69	8	1.62	1.32–1.92
Sertraline	9	2.23	1.56–2.90	3	1.53	0.34–2.72	3	3.59	2.20–4.98	3	1.66	1.29–2.03
TCAs	15	1.83	1.43–2.21	13	1.68	1.31–2.05	2	3.02	0.89–5.15	–	–	–
Clomipramine	7	1.85	1.10–2.60	7	1.85	1.10–2.60	–	–	–	–	–	–
Imipramine	7	1.82	1.36–2.28	6	1.55	1.24–1.86	1	4.17	3.15–5.19	–	–	–
Opipramol	1	1.99	1.65–2.33	–	–	–	1	1.99	1.65–2.33	–	–	–
CBT + drug combinations	16	2.12	1.66–2.59	10	1.55	1.17–1.93	2	6.04	3.71–8.37	4	2.15	1.35–2.95
Exercise	3	1.23	0.58–1.88	2	1.17	0.02–2.32	–	–	–	1	1.38	0.77–1.99
Control groups												
Pill placebo	111	1.29	1.14–1.44	36	1.02	0.78–1.25	39	1.85	1.61–2.09	37	0.94	0.77–1.12
Psychological Placebo	16	0.83	0.54–1.12	5	0.90	0.27–1.52	4	1.42	0.76–2.09	7	0.52	0.20–0.84
Waiting list	50	0.20	0.12–0.28	20	0.19	0.06–0.32	13	0.22	0.03–0.41	17	0.21	0.09–0.33

CBT, cognitive behavioural therapy; EMDR, eye movement desensitization reprocessing; IPT, interpersonal therapy; *n*, number of study arms; SNRIs, serotonin–noradrenaline reuptake inhibitors; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.
[a]One study used multiple SSRIs (van Apeldoorn *et al.*, 2008).

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; $t = 4.6$; $d.f. = 202$; $P < 0.0001$). PDTh 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 ± 95.6). 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.

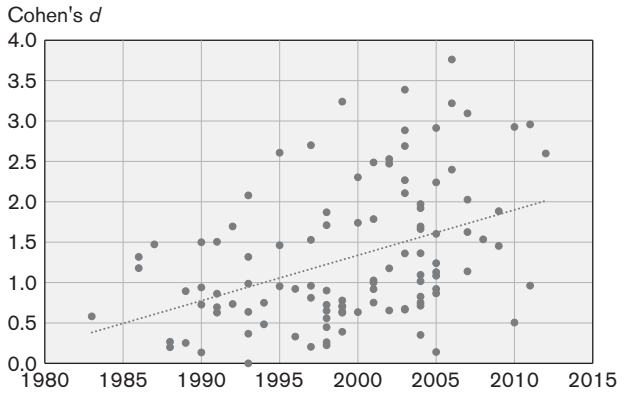
Discussion

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.

Fig. 3



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 ($d = 0.056 \times \text{year} - 111.1$).

Among the drugs that are currently most commonly used, licensed and recommended by guidelines, that is 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, contra-indications 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

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

Treatment	*n*	*d*	CI	*P*	I^2	Fail-safe *N*	Egger *P*	Adjusted *d*	CI
CBT (individual) vs. waiting list	25	1.23	1.02–1.45	< 0.0001	70.0	2701	0.44 (NS)	1.23	1.02–1.45
CBT (individual) vs. psychological placebo	12	0.75	0.34–1.16	< 0.0001	83.5	168	0.033	0.26	−0.19 to 0.72
CBT (individual) vs. pill placebo	9	0.57	0.20–0.94	0.003	64.8	49	0.023	0.24	−0.19 to 0.66
CBT (group) vs. waiting list	7	1.33	1.06–1.60	< 0.0001	0.0	157	0.47 (NS)	1.26	1.01–1.51
CBT (group) vs. pill placebo	5	0.12	−0.10 to 0.35	0.28 (NS)	–	–	0.27 (NS)	0.12	−0.10 to 0.35
Relaxation vs. waiting list	6	1.31	0.92–1.71	< 0.0001	41.3	114	0.054*	1.21	0.80–1.63
Non-face-to-face therapies vs. waiting list	20	0.91	0.73–1.09	< 0.0001	53.6	1108	0.43 (NS)	0.91	0.73–1.09

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; I^2, heterogeneity; *n*, number of study arms.
*Trend towards statistical significance.

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

Drugs	*n*	*d*	CI low	*P*	I^2	Fail-safe *N*	Egger *P*	Adjusted *d*	CI
Alprazolam	9	0.66	0.37–0.95	< 0.0001	64.4	123	0.24	0.49	0.18–0.81
Buspirone	5	0.35	0.05–0.64	0.02	52.7	12	0.29	0.34	0.05–0.64
Citalopram	2	0.23	−0.13 to 0.59	0.20 (NS)	75.8	–	–	–	–
Clomipramine	5	0.87	0.43–1.31	< 0.0001	77.3	78	0.29	0.529	0.04–1.02
Clonazepam	4	0.63	0.37–0.90	< 0.0001	48.2	51	0.20	0.581	0.31–0.85
Diazepam	7	0.76	0.38–1.14	< 0.0001	69.3	83	0.31	0.611	0.18–1.04
Duloxetine	3	0.53	0.42–0.65	< 0.0001	0.0	60	0.003	0.53	0.42–0.65
Escitalopram	6	0.45	0.26–0.63	< 0.0001	77.5	148	0.64	0.45	0.26–0.63
Fluoxetine	4	0.56	0.07–1.05	0.025	80.1	42	0.001	0.56	0.07–1.05
Fluvoxamine	12	0.60	0.43–0.78	< 0.0001	58.3	342	0.18	0.56	0.37–0.75
Hydroxyzine	4	0.79	0.41–1.17	< 0.0001	77.8	72	0.32	0.49	0.07–0.91
Imipramine	4	0.74	0.46–1.01	< 0.0001	0.0	22	0.38	0.74	0.46–1.01
Lorazepam	3	0.87	0.13–1.61	0.022	92.6	49	0.25	0.87	0.13–1.61
Moclobemide	5	0.17	−0.23 to 0.56	0.4 (NS)	86.5	–	–	–	–
Opipramol	1	0.20	−0.07 to 0.48	0.14 (NS)	–	–	–	–	–
Paroxetine	19	0.56	0.41–0.70	< 0.0001	85.6	1603	0.06	0.42	0.26–0.57
Phenelzine	6	0.96	0.60–1.31	< 0.0001	49.5	79	0.90	0.96	0.60–1.31
Pregabalin	8	0.55	0.37–0.74	< 0.0001	78.0	314	0.92	0.55	0.37–0.74
Quetiapine	3	0.56	0.47–0.66	< 0.0001	0.0	106	0.37	0.56	0.47–0.66
Sertraline	6	0.54	0.38–0.70	< 0.0001	60.0	169	0.23	0.54	0.38–0.70
Venlafaxine	19	0.50	0.31–0.69	< 0.0001	90.6	1305	0.55	0.33	0.14–0.52

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; I^2, 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 antipsychotics, 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 PDTh, 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 an 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

Borwin Bandelow: manuscript writing, data collection, meta-analysis, statistics. Markus Reitt: data collection, meta-analysis, statistics. Christian Röver and Yvonne Görlich: meta-analysis, statistics. Sophie Michaelis: data collection, meta-analysis. Dirk Wedekind: manuscript writing.

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

Baldwin DS, Ajel K, Masdrakis VG, Nowak M, Rafiq R (2013). Pregabalin for the treatment of generalized anxiety disorder: an update. *Neuropsychiatr Dis Treat* **9**:883–892.

Baldwin DS, Anderson IM, Nutt DJ, Allgulander C, Bandelow B, den Boer JA, *et al.* (2014). Evidence-based pharmacological treatment of anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder: a revision of the 2005 guidelines from the British Association for Psychopharmacology. *J Psychopharmacol* **28**:403–439.

Bandelow B, Seidler-Brandler U, Becker A, Wedekind D, Rüther E (2007). Meta-analysis of randomized controlled comparisons of psychopharmacological and psychological treatments for anxiety disorders. *World J Biol Psychiatry* **8**:175–187.

Bandelow B, Zohar J, Hollander E, Kasper S, Möller HJ, Zohar J, *et al.*, WFSBP Task Force on Treatment Guidelines for Anxiety, Obsessive-Compulsive and Post-Traumatic Stress Disoders (2008). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the pharmacological treatment of anxiety, obsessive-compulsive and post-traumatic stress disorders – first revision. *World J Biol Psychiatry* **9**:248–312.

Borenstein M, Hedges L, Higgins J, Rothstein H (2009). *Introduction to meta-analysis*. New York: Wiley.

Borkovec TD, Sibrava NJ (2005). Problems with the use of placebo conditions in psychotherapy research, suggested alternatives, and some strategies for the pursuit of the placebo phenomenon. *J Clin Psychol* **61**:805–818.

Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry* **196**:173–178.

Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF III (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry* **12**:137–148.

Dunlap WP, Cortina JM, Vaslow JB, Burke MJ (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods* **1**:170–177.

Duval S, Tweedie R (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**:455–463.

Fedoroff IC, Taylor S (2001). Psychological and pharmacological treatments of social phobia: a meta-analysis. *J Clin Psychopharmacol* **21**:311–324.

Flint J, Cuijpers P, Horder J, Koole SL, Munafo MR (2014). Is there an excess of significant findings in published studies of psychotherapy for depression? *Psychol Med* 1–8.

Furukawa TA, Watanabe N, Churchill R (2007). Combined psychotherapy plus antidepressants for panic disorder with or without agoraphobia. *Cochrane Database Syst Rev* **1**:CD004364.

Hamilton M (1959). The assessment of anxiety states by rating. *Br J Med Psychol* **32**:50–55.

Hoge EA, Bui E, Marques L, Metcalf CA, Morris LK, Robinaugh DJ, *et al.* (2013). Randomized controlled trial of mindfulness meditation for generalized anxiety disorder: effects on anxiety and stress reactivity. *J Clin Psychiatry* **74**: 786–792.

Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, *et al.* (2013). Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* **185**:E201–E211.

Jacobson E (1938). *Progressive relaxation*. Chicago: University Press.

Jin H, Meyer JM, Jeste DV (2004). Atypical antipsychotics and glucose dysregulation: a systematic review. *Schizophr Res* **71**:195–212.

Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF (2013). Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ* **347**:f6104.

Knijnik DZ, Kapczinski F, Chachamovich E, Margis R, Eizirik CL (2004). Psychodynamic group treatment for generalized social phobia. *Rev Bras Psiquiatr* **26**:77–81.

Leichsenring F, Salzer S, Beutel ME, Herpertz S, Hiller W, Hoyer J, *et al.* (2013). Psychodynamic therapy and cognitive-behavioral therapy in social anxiety disorder: a multicenter randomized controlled trial. *Am J Psychiatry* **170**:759–767.

Liebowitz MR (1987). Social phobia. *Mod Probl Pharmacopsychiatry* **22**:141–173.

Lipsitz JD, Gur M, Vermes D, Petkova E, Cheng J, Miller N, *et al.* (2008). A randomized trial of interpersonal therapy versus supportive therapy for social anxiety disorder. *Depress Anxiety* **25**:542–553.

Mitte K (2005). A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *J Affect Disord* **88**:27–45.

Moher D, Altman DG, Liberati A, Tetzlaff J (2011). PRISMA statement. *Epidemiology* **22**:128. author reply 128.

National Institute for Health and Clinical Excellence (NICE) (2011). *Anxiety: management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care*. London: The British Psychological Society and The Royal College of Psychiatrists. Available at: *http://www.nice.org.uk*. [Accessed 12 April 2015].

National Institute for Health and Clinical Excellence (NICE) (2013). *Social anxiety disorder: recognition, assessment and treatment (full guideline)*. London: The British Psychological Society and The Royal College of Psychiatrists. Available at: *http://www.nice.org.uk*. [Accessed 12 April 2015].

Ost LG (1988). Applied relaxation vs progressive relaxation in the treatment of panic disorder. *Behav Res Ther* **26**:13–22.

Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG (2009). Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* **118**:1–8.

Roshanaei-Moghaddam B, Pauly MC, Atkins DC, Baldwin SA, Stein MB, Roy-Byrne P (2011). Relative effects of CBT and pharmacotherapy in depression versus anxiety: is medication somewhat better for depression, and CBT somewhat better for anxiety? *Depress Anxiety* **28**:560–567.

Scottish Intercollegiate Guidelines Network (SIGN) (2012). Methodology checklist 2: controlled trials. Available at: *http://www.sign.ac.uk*. [Accessed 1 April 2013].

Stangier U, Schramm E, Heidenreich T, Berger M, Clark DM (2011). Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: a randomized controlled trial. *Arch Gen Psychiatry* **68**:692–700.

Sterne JA, Gavaghan D, Egger M (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* **53**:1119–1129.

van Apeldoorn FJ, van Hout WJ, Mersch PP, Huisman M, Slaap BR, Hale WW III, *et al.* (2008). Is a combined therapy more effective than either CBT or SSRI alone? Results of a multicenter trial on panic disorder with or without agoraphobia. *Acta Psychiatr Scand* **117**:260–270.

Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, *et al.* (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* **21**:655–679.