Published in final edited form as:

Expert Opin Pharmacother. 2018 June ; 19(8): 883–894. doi:10.1080/14656566.2018.1472767.

Comparing the Efficacy of Benzodiazepines and Serotonergic Anti-Depressants for Adults with Generalized Anxiety Disorder: A meta-analytic review

Angelina F. Gomez, Abigail L. Barthel, and Stefan G. Hofmann

Department of Psychological and Brain Sciences, Boston University

Abstract

Introduction: Generalized anxiety disorder (GAD) is a common form of anxiety disorder. Selective serotonin reuptake inhibitors (SSRIs), serotonin and norepinephrine reuptake inhibitors (SNRIs), and benzodiazepines (BZs) are the most commonly prescribed medications for GAD, but little is known about the relative efficacy of these pharmacological treatments.

Areas Covered: This study provides a meta-analytic review of the efficacy of these medications in the treatment of adults with GAD. A comprehensive literature search yielded 54 articles reporting 56 unique studies with 12,655 participants treated with either pill placebo (6,191 participants), SSRIs (16 trials, 2,712 participants), SNRIs (17 trials, 2,603 participants), or BZs (23 trials, 1,149 participants). The overall combined effect size was modest to moderate (Hedges' $g = 0.37$, $p < 0.0001$). Effect sizes decreased significantly over time. SSRIs (Hedges' $g = 0.33$) and SNRIs (Hedges' $g = 0.36$) demonstrated significantly lower effect sizes than BZs (Hedges' $g = 0.50$).

Expert opinion: These findings were not due to differences in treatment length or publication year. The results of this study suggest that the most common forms of pharmacotherapy for adult GAD are moderately effective, with BZs being the most effective drug.

1. Introduction

Generalized Anxiety Disorder (GAD) is characterized by persistent and uncontrollable worry that is associated with various unpleasant physiological symptoms[1]. The disorder shows an estimated 12-month prevalence between 0.2-2.9%, and an estimated lifetime prevalence between 2.8-6.2%[2]. Numerous clinical trials have tested the efficacy of various pharmacotherapeutic agents for GAD over the past 30 years. Specifically, Selective Serotonin Reuptake Inhibitors (SSRIs), Selective Norepinephrine Reuptake Inhibitors (SNRIs), and benzodiazepines (BZs) have become the most commonly prescribed medications for GAD, due to their overall efficacy and relative tolerability. Among these, SSRIs and SNRIs are considered the first-line medication options for GAD[3]. Randomized-controlled trials suggest that 60-75% of patients respond to SSRIs, compared to 40-60% of patients receiving placebo, with similar response rates for SNRIs and BZs[4]. Current

prescribing guidelines recommend BZs for use only as a second-line or augmentative treatment option, although they are commonly prescribed by primary care physicians[5].

Randomized-controlled trials of SSRIs have demonstrated consistent strong evidence of their efficacy in treating GAD, with one study reporting an overall combined effect size (Hedges' g) of 0.36 ($SD=0.09$, $p<0.0001$)[6]. Side effects of SSRIs during the initial weeks of treatment include nausea, dizziness, headaches, jitteriness, sleep difficulties, and gastrointestinal disturbances. Some of these side effects are similar to the symptoms of anxiety, which can lead to premature termination before the medication has had enough time to reach maximum efficacy[5], though slow titration has been shown to attenuate these initial side effects [7]. SNRIs are very similar to SSRIs in their mechanism of action, side effects, and efficacy rates (Hedges' $g = 0.42$, $SD=0.12$, $p<0.0001$)[6], though there is some evidence that SNRIs are associated with less sexual dysfunction side effects than SSRIs [8].

Randomized placebo-controlled trials have demonstrated that BZs are also efficacious treatments for GAD. The combined overall effect size (Hedges' g) of BZ for GAD was 0.38 ($SD=0.15$, $p<0.0001$)[6]. This effect size is equivalent to the individual effect sizes of SSRIs and SNRIs, as well as the overall effect size for the three treatments combined ($g = 0.39$)[6]. BZs are tolerable and have a rapid onset, unlike serotonergic anti-depressants (ADs). However, BZs can be addictive[9], though the risk for developing addiction varies widely among individuals treated with BZs[10]. BZs have also been associated with over-sedation, cognitive impairment, and psychomotor incoordination, which can compound over time if not properly managed[11–13]. As such, BZs are not recommended for patients with current or past substance abuse, and should not be prescribed for indefinite periods of time[5].

These precautions are warranted, but it is also important to note that BZs do not always cause adverse side effects, and many people find them to be highly tolerable and effective. Current prescribing guidelines only recommend BZs as augmentative agents, particularly to mitigate the adverse side effects during the initial weeks of therapy with an antidepressant[4]. However, some of the literature suggests that both drug classes may be associated with similar discontinuation effects[14]. Furthermore, despite the current prescribing guidelines, physician surveys (including psychiatrists and non-psychiatrists) indicate that BZs are prescribed just as frequently as SSRIs and SNRIs, though primarily by non-psychiatrists[15]. It is therefore important that researchers and clinicians have an accurate understanding the overall efficacy of these commonly prescribed agents among adults with GAD.

The primary aims of this meta-analysis are to: 1) determine the overall effect size estimate of SSRIs, SNRIs, and BZs vs. placebo for GAD, 2) determine effect size estimates for each drug class, 3) examine whether statistically significant differences in treatment efficacy exist based on drug class, and if so, 4) explore whether any moderator variables contribute to this difference.

2. Method

2.1 Search Strategy

A comprehensive literature search for randomized, placebo-controlled, pharmacotherapy trials was conducted using the PsycInfo, PubMed, and Web of Science online databases. Article titles, abstracts, and subjects were searched using the following terms: “pharmacotherapy” or “medication” or “drug”; “trial” or “clinical trial” or “randomized controlled” or “placebo controlled”; and “GAD” or “general anxiety” or “generalized anxiety disorder” or “generalised anxiety disorder”. Additionally, the reference list of relevant review articles and previous meta-analyses were manually reviewed to identify additional studies[2, 16–28].

2.2 Inclusion/Exclusion

Articles were deemed eligible if they met all of the following predetermined criteria: 1) The study was a randomized, placebo-controlled trial exploring the efficacy of SSRIs, SNRIs, or BZs in the treatment of GAD; 2) Participants were adults (above 18 years of age); 3) Participants received a primary diagnosis of GAD (using DSM-III, -IV, or -5 criteria); secondary comorbid diagnoses were allowed to reflect an ecologically valid sample[1]; and, 4) The primary outcome measure was improvement on the Hamilton Anxiety scale (HAM-A)[29]. This measure was selected as it is the most widely used, clinician-administered inventory of general anxiety symptoms. Though other anxiety measures can be used in a meta-analysis to calculate an overall effect size, we only included studies with the HAM-A for a few reasons. We wanted our results to be comparable to other recent meta-analyses of drug treatment for adults with GAD, which have exclusively use the HAM-A[16, 17]. Additionally, the measure has been widely validated across clinical samples, has high interrater reliability, and contains a thorough assessment of both the psychological and somatic symptoms of GAD along with resulting distress and life interference[29]. However, it is worth noting that the HAM-A has been criticized for not clearly separating antidepressant from anxiolytic effects, and for the degree of overlap between symptoms and side effects of somatic complaints[29].

Articles were excluded from analysis if they met any of the following exclusion criteria: 1) Not a clinical drug trial with humans (e.g. CBT, cross-sectional study, literature review, Phase 1/rodent trials; $n = 552$); 2) The research precluded valid pre-post efficacy data (e.g. open-label, no randomization, waitlist/TAU-controlled, crossover design, withdrawal study, augmentation trial; $n = 59$); 3) The treatment agent was not in one of the three identified drug classes ($n = 28$); 4) Participants were children or adolescents ($n = 9$); 5) Participants did not have primary diagnosis of GAD ($n = 763$); 6) Insufficient data were provided to conduct the meta-analysis, or full text of the article could not be accessed ($n = 6$); and 7) Results were already included in the meta-analysis reported in a different article (primary efficacy articles were chosen in the case of two articles reporting results on the same sample; $n = 25$). Articles were examined for eligibility by two independent raters (A.F.G and A.L.B), and the final list of eligible articles was determined by comparing these ratings and reaching a consensus.

2.3 Data Collection/Synthesis

Eligible articles were downloaded and examined for efficacy data, participant demographics, and trial characteristics. The following variables were collected: pharmacotherapy class, number of participants in each condition, length of treatment in weeks, mean age, and sex characteristics. The following efficacy variables (in order of preference) were extracted by the first two authors and entered into the meta-analysis: 1) pre-post means and standard deviations; 2) within-groups adjusted mean change from baseline; 3) post means in each group, or difference in means, with *p*-value between groups; 4) sample size and *p*-values between groups, or 5) other available effect size statistics (e.g. Cohen's *d*). When available, adjusted means and standard deviations were used to provide the most reliable effect size data. In the event of missing data, authors were contacted and if appropriate data were not provided, effect sizes were approximated using sample sizes and *p* values. Effect size data were also identified in effect size tables from other meta-analyses and used accordingly. Some of the reviewed studies (17 trials) required additional imputations (e.g. converting standard errors to standard deviations); formulas were obtained from the Cochrane Handbook for Systematic Reviews of Interventions (open online access). These data extraction steps were necessary to include a sufficient number of studies in the meta-analysis to conduct multiple meta-regression analyses.

2.4 Primary Efficacy and Moderator Analyses

Controlled effect sizes were estimated using Hedges' *g*, which corrects for parameter bias due to small sample size [30]. Effect sizes were estimated using random effects models, which assume significant heterogeneity across the included studies. Moderator analyses examined whether there were significant differences in effect size based on study or participant characteristics. Categorical moderators (drug class, risk of bias) were dummy-coded and included in separate two-way ANOVAs with 'SSRI' and 'low' risk as reference variables. Continuous moderators (mean age, % female, length of treatment, publication year) were included in multiple meta-regressions. Where significant moderation effects were found, follow-up analyses were conducted to further explain these results. All analyses were completed with Comprehensive Meta-Analysis [31].

2.5 Quality Control Analyses

Trial quality was assessed using recommendations from the Cochrane risk of bias tool [32]. Each study was rated by two independent authors (A.F.G. and A.L.B.) as containing a high, low or unclear risk of bias in the following domains: 1) Selection Bias (sequence generation and allocation concealment); 2) Attrition Bias (incomplete outcome data/analyses and information about withdrawal, adverse events, and sample size variations); and 3) Reporting Bias (selective outcome reporting; see Table 1). Studies were rated as 'low risk' if all facets of the Cochrane criteria were met within each domain, 'unclear risk' if Cochrane criteria were mostly present or not explained fully, and 'high risk' if one or more elements of the Cochrane criteria were excluded or missing from the studies (see [32] for criteria). Common sources of error pertained to inadequate details regarding blinding and randomization techniques, insufficient data analyses (e.g. LOCF without supplementary MMRM), and unclear group or outcome data during data collection or analysis. After risk of bias was

assessed within studies, an overall rating of bias was made by the first two authors, with 'low risk' overall pertaining to studies with low risk of publication bias across all three domains, 'unclear risk' overall pertaining to studies with one or more sources of error, and 'high risk' overall for studies that excluded data in any domain or included persistent areas of inadequate information.

To address the file drawer problem, a classic fail-safe N analysis was conducted [33, 34]. This analysis estimates the number of missing studies with an effect size of zero necessary to yield non-significant results. For a more meaningful metric, Orwin's fail-safe N test was also conducted. This analysis estimates the number of missing studies with an effect size of zero necessary to yield a trivial effect size (set at Hedges' $g = 0.05$). Unlike the classic fail-safe N test, Orwin's fail-safe N test allows the researcher to set the 'missing' effect size at any level, including a negative effect (which is conceivable, given the potential for medications to increase anxiety).

Lastly, publication bias was examined visually in a funnel plot, and statistically using the Trim and Fill method [35]. The main function of the funnel plot is to indicate any asymmetry in the distribution of studies around the overall effect size based on sample size. Since smaller samples are more susceptible to bias, if the funnel plot indicates more small studies on the right than on the left, the concern is that these smaller studies are disproportionately impacting the overall effect size estimate. Trim and Fill uses an iterative procedure to remove the most extreme small studies from the right, re-computing the effect size at each iteration until the funnel plot is symmetric about the (new) effect size. This will theoretically yield an unbiased estimate of the effect size. Caveats to this approach are that it is built on strong assumptions that the funnel plot should be symmetric, and is known to perform poorly in the presence of substantial between-study heterogeneity [36, 37]. Additionally, the trim and fill method assumes that funnel plot asymmetry is solely due to publication bias, though there is no way to know the true mechanism for this bias [38].

3. Results

3.1 Study Characteristics

Our initial search yielded 1,965 results (see Figure 1). After removing duplicates, books/conference reports, and non-adult populations, 1,450 titles were deemed relevant to evaluate for initial eligibility. Additionally, 46 unique articles were identified from other published meta-analyses. Of these 1,496 article titles, 93 were downloaded and evaluated for inclusion and exclusion criteria. Of these, 39 did not meet inclusion criteria. The remaining 54 articles included data from 56 unique trials of adults with GAD treated with either pill placebo (6,191 participants), SSRIs (16 trials, 2,712 participants), SNRIs (17 trials, 2,603 participants), or BZs (23 trials, 1,149 participants). Information from these 56 trials comprises 12,655 total individual participants with GAD (6,464 receiving active medications; 6,191 receiving placebo; see Table 1).

3.2 Primary Efficacy and Moderator Analysis Results

Controlled effect sizes, standard errors, and significance values for each study are reported in Table 2, and visually presented in Figure 3. The random effects meta-analysis yielded an overall, controlled effect size of Hedges' $g = 0.37$ ($SE = 0.02$, $p < 0.0001$), which suggests that pharmacotherapy was significantly more effective than placebo in reducing GAD symptoms. Results of moderator analyses revealed a significant effect of drug class and publication year (see below) on overall effect size. There was no significant relation between effect size and mean age ($\beta = 0.0014$, $p = 0.70$), % female ($\beta = 0.0008$, $p = 0.73$), and length of treatment ($\beta = -0.0052$, $p = 0.23$).

3.3 Quality Control Analyses

Across the surveyed 54 articles, 12 were given a high overall risk rating, 9 were given a low overall risk score, and 33 were given an unclear overall risk rating. The 'unclear risk' status was mostly due to reviewed studies using a double-blind, randomized-control trial method without explanation of how this method was adhered to, which is consistent with potential bias according to the Cochrane risk of bias tool[32]. This meta-analysis did not include the results from any unblinded randomized placebo-controlled studies. Results indicated no significant differences in effect sizes based on the quality of trial ($Q = 1.34$, $df = 2$, $p = 0.51$). Trials rated as 'high risk' had the largest effect size ($g = 0.42$, $SE = 0.04$, $p < 0.0001$), followed by 'unclear risk' ($g = 0.37$, $SE = 0.04$, $p < 0.0001$), and then 'low risk' ($g = 0.36$, $SE = 0.03$, $p < 0.0001$).

The classic fail-safe N test reveals the number of studies needed to reduce the effect size to statistical non-significance level. If the fail-safe N exceeds $5K+10$, the effect is considered robust (with $K = \#$ trials)[34]. Results of this test indicated that 6,137 trials with an effect size of zero would be needed to bring the overall p -value to non-significance. Additionally, results of Orwin's fail-safe N test [33] indicated that 354 trials with an effect size of zero would be needed to yield an overall trivial effect of treatment over placebo (with a 'trivial' effect set at $g = 0.05$).

Inspection of the funnel plot revealed effect sizes distributed somewhat asymmetrically around the mean effect size (see Figure 2). Using the Trim and Fill method [35], we determined that 17 studies would need to fall to the left of the mean (*i.e.*, have an effect size smaller than the mean) and 0 studies would need to fall to the right of the mean (*i.e.*, have an effect size larger than the mean) to make the plot symmetrical. The random-effects model for the new imputed mean effect size revealed a Hedges' g of 0.33 (95% CI: 0.28-0.37).

3.4 Efficacy Results by Drug Class

See Figure 3 for the forest plot of study effect sizes by drug class. Results of the two-way ANOVA revealed a significant overall difference in effect size based on drug class ($Q = 10.82$, $df = 2$, $p = 0.004$). SSRIs demonstrated the lowest effect size ($g = 0.33$, $SE = 0.03$, $p < 0.0001$; 16 trials with 5,444 individual participants), which was similar to the effect size for SNRIs ($g = 0.34$, $SE = 0.03$, $p < 0.0001$; 17 trials with 4,993 individual participants). BZs yielded the highest effect size ($g = 0.50$, $SE = 0.04$, $p < 0.0001$; 23 trials with 2,218 individual participants), which was significantly larger than effect sizes for SSRIs ($Q =$

10.33, $df=1$, $p=0.001$) and SNRIs ($Q=6.78$, $df=1$, $p=0.009$). There were no significant differences between the two serotonergic anti-depressants ($Q=0.50$, $df=1$, $p=0.48$). Therefore, they were combined into one group (antidepressants; ADs) showing an effect size of 0.34 ($SE=0.02$, $p<0.0001$; 33 trials with 10,437 individual participants) in all further analyses.

3.5 Efficacy Results by Publication Year

Results of the meta-regression revealed a significant overall difference in effect size based on publication year ($Q=8.71$, $df=1$, $p=0.0032$), such that efficacy results decreased significantly over time ($\beta=-0.009$, $SE=0.003$). To examine whether a particular decade was responsible for this temporal trend, trials were dummy coded into three groups: trials published between 1984 and 1995, trials published between 1996 and 2005, and trials published between 2006 and 2015. Results of a two-way ANOVA revealed a significant overall difference in effect size based on publication decade ($Q=10.81$, $df=2$, $p=0.005$). Trials published between 1984 and 1995 demonstrated the highest effect size ($g=0.53$, $SE=0.06$, $p<0.0001$; 13 trials with 1,070 individual participants). Trials published between 1996 and 2005 demonstrated an effect size decrease of 0.13 from the previous decade ($g=0.40$, $SE=0.03$, $p<0.0001$; 24 trials with 5,068 individual participants). Trials published between 2006 and 2015 demonstrated an effect size decrease of 0.09 from the previous decade ($g=0.31$, $SE=0.03$, $p<0.0001$; 19 trials with 6,517 individual participants).

3.6 Moderator Analyses

Moderator analyses were conducted to explore whether the effect of drug class was explained by any other moderator variable. First, based on the evidence that BZs and ADs have distinct anxiolytic and side-effect profiles during the initial phases of treatment [39], an interaction analysis was conducted to explore whether the significant effect of drug class differed based on treatment length. Results of a meta-regression including treatment length, binary drug class, and the interaction term revealed a non-significant interaction effect (interaction term $\beta=-0.0021$, $SE=0.02$, $p=0.93$), suggesting that BZs yield higher effect sizes than ADs, regardless of the length of treatment.

Additional analyses were conducted to explore whether the effect of drug class held up when including publication year as a covariate. Results of a meta-regression including treatment length and binary drug class revealed that, when holding year published constant, the effect of drug class was no longer significant ($Q=2.95$, $df=1$, $p=0.09$).

Trials published within the first decade (1984-1995) were only BZ trials (13 trials). Trials published within the second decade (1996-2005) contained 9 BZ trials, 7 SNRI trials, and 8 SSRI trials. Trials published within the most recent decade (2006-2015) contained 1 BZ trial, 10 SNRI trials, and 8 SSRI trials. Thus, the majority of BZ trials were published within the first and second decades, indicating that the significant effect of publication year could simply be representative of the significant effect of drug class. To test this hypothesis, the effect of publication year was examined when taking all BZ trials out of the analysis. Results of this meta-regression were no longer significant ($Q=0.82$, $df=1$, $p=0.37$; 33 trials with 10,437 individual participants). These follow-up analyses suggest that the significant effect

of publication year was primarily due to the significant effect of drug class and the decade in which trials examining each drug class were published.

4. Discussion

The current meta-analysis of 56 randomized, placebo-controlled trials of pharmacotherapy for GAD yielded an overall controlled effect size of Hedges' $g = 0.37$, which reflects a small-to-medium effect of SSRIs, SNRIs, and BZs combined over placebo. These results did not differ significantly based on participant age, sex, treatment length, or risk of publication bias (low, medium, high). Moreover, results of this meta-analysis suggest that treatment effects differ significantly based on drug class. Benzodiazepines yielded a significantly larger effect size (Hedges' $g = 0.50$) than serotonergic anti-depressants (Hedges' $g = 0.34$), irrespective of treatment length. These results support the use of SSRIs, SNRIs, and BZs for the treatment of GAD. Moreover, the findings favor BZs over SSRIs and SNRIs. We realize that this is inconsistent with current prescribing guidelines, which do not recommend BZs as a first-line treatment for GAD.

The significant change in overall effect size across the three decades examined in this meta-analysis is notable, and can be interpreted in a few ways. The follow-up analyses reported in this study suggest that the significantly higher overall effect size of BZs, along with the lack of BZ trials published in the two most recent decades (1996-2005 included 9 BZ trials, while 2006-2015 included just 1 BZ trial), could partially explain the significant decrease in overall effect size over time. Another possible explanation for these findings may be the result of changing requirements for the reporting of null or negative drug trial findings. Prior to publishing guidelines implemented within the past decade, companies sponsoring drug trials were not required to make negative/null findings publicly available. This potential bias could therefore explain the larger effect sizes observed in public reports from the 80's and 90's.

A third possible explanation for the effect of publication year found in this meta-analysis is the increase in placebo-response to antidepressants over time, demonstrated in the seminal study published in JAMA [40]. Indeed, we found that the overall effect size continued to decrease between the second and third decades examined in this meta-analysis; studies published between 1996-2005 had an overall effect size of $g = 0.40$ (24 trials with 5,068 individual participants), which decreased by 0.09 in the studies published between 2006-2015 ($g = 0.31$; 19 trials with 6,517 individual participants). Notably, the studies in the last two decades were primarily antidepressant trials, and gathered data from trials conducted after the Walsh 2002 JAMA article was published. It is therefore conceivable that the results reported in this study could be a quasi-replication of Walsh and colleagues' findings; more specifically, the temporal decrease in overall effect size reported in this study could indicate a further increase in placebo-response within antidepressant clinical trials over time.

The thorough quality control assessments reported in this study indicate minimal evidence that the overall effect size estimate is impacted by publication bias. The statistical comparison of trials rated as 'low' 'unclear' or 'high' risk of publication bias revealed no

significant differences in overall effect size based on trial quality ($p = 0.51$). Additionally, the results of the two types of fail-safe N tests suggest that the overall effect size is a robust finding. The classic fail-safe N is the least stringent robustness analysis, as it provides an estimate of the number of trials with an effect size of 0.0 that would be needed to yield a *null* effect of drug vs. placebo ($p > 0.05$). Orwin's fail-safe N is a much more meaningful analysis, as it provides an estimate of the number of trials with an effect size of 0.0 that would be needed to yield a *clinically trivial* effect of drug vs. placebo ($g = 0.05$). Substantive (or clinical) significance is a much higher standard to reach than alpha significance, so Orwin's fail-safe N test yields the most probable estimate of the number of studies needed to meaningfully counteract the reported results. Results of our analysis indicated that 354 trials would be needed to reduce our findings to clinical non-significance, which is more than 6x the number of studies included in this meta-analysis. Lastly, the funnel plot was somewhat asymmetrically distributed, indicating modest evidence that the overall effect size could be influenced by publication bias in studies with small samples. Results of the Trim and Fill analysis indicated that 17 studies with effect sizes less than the overall effect size estimate would be necessary to create a symmetrical funnel plot. If these studies were included in this meta-analysis, the new 'unbiased' overall effect size estimate would be $g = 0.33$ (reduced from 0.37), which is still a clinically significant effect of drug over placebo.

A number of limitations constrain the generalizability of the results presented herein. First, the significant difference in effect size based on drug class should not be conflated with outperformance of one drug class over the other. To determine whether BZs outperform ADs, head-to-head comparisons between the different classes of medication must be conducted beyond those of which have been mentioned previously. Secondly, these results should be interpreted with caution given that there was no examination of patient characteristics that may contraindicate the use of BZs (e.g. current or past substance abuse, cognitive impairments, or tolerance effects). Nonetheless, the robust effect size of BZs for the treatment of GAD should be considered for clinical decision making. Third, most of the BZ trials used compounds with medium-long half-lives (e.g. lorazepam, diazepam), which have different anxiolytic and side-effect profiles than shorter-acting compounds (such as alprazolam). These slower-acting compounds are associated with less risk of dependence relative to alprazolam, but are more frequently associated with future development of cognitive decline[41]. Fourth, this meta-analysis only focused on SSRIs, SNRIs, and BZs, but there are many more pharmacotherapeutic agents found to be effective for the treatment of GAD. Most notably, pregabalin is an alternative first-line treatment for GAD, has a rapid onset of action similar to BZs, and can be combined with antidepressants during the first days of treatment in substitution for a BZ[42, 43]. Finally, these results are limited by variability in the data used to calculate controlled effect sizes, given that additional imputations and effect size estimations were conducted in the absence of complete study data.

5. Conclusions

The results of this meta-analysis provide a novel contribution to the recent interest in reappraising the role of BZs in the treatment of GAD. This topic is currently debated given that BZs have consistently demonstrated superior tolerability, whereas ADs are commonly

associated with GAD-related side-effects. Moreover, this inquiry is especially salient since there are few RCTs directly comparing BZs with ADs, despite evidence suggesting that their efficacy may be comparable. As such, the current recommendation that ADs should be the first-line intervention for GAD may lead to the unsubstantiated assumption that ADs are more efficacious than BZs. This meta-analysis provides evidence that BZs are more efficacious than ADs for reducing GAD symptoms. In light of this evidence and in accordance with other recent expert opinions [4], we recommend that BZs should be considered as a viable treatment option for adults with GAD, especially for the initial treatment phase.

This recommendation does not minimize the appropriateness of ADs as treatments for GAD. However, the emerging evidence suggests that the combined use of ADs and BZs during initial medication treatment could yield the optimal risk-benefit ratio of both medications. BZs reach their maximal efficacy rapidly, but these effects level off after about 4 weeks of regular treatment. Conversely, ADs reach their maximal efficacy within 4-8 weeks (if patients can tolerate the side-effects, that is). It should be noted, however, that BZs are not free of adverse effects; indeed, they have been shown to lead to cognitive impairments and an increased risk of developing dementia when prescribed for long periods of time [11–13]. Given these contrasting temporal profiles, we recommend the combined use of BZs and ADs during the first 4-8 weeks of medication treatment, followed by careful tapering of the BZ dosage, or reduction to use only as-needed (PRN) to augment continued AD use. We believe this type of combined use could minimize the adverse risks of BZs, while potentially improving AD treatment compliance by attenuating initial unpleasant side-effects.

These are of course broad recommendations based on the results of this one meta-analysis, and prescribing decisions are made weighing a number of factors in addition to the risk-benefit assessment. As always, prescribing clinicians should tailor the specific treatment strategy to the individual patient by carefully considering the patient's history, the potential for clinically meaningful improvements in quality of life, as well as any possible adverse effects.

Although this study did not examine psychosocial treatments for GAD, it should be mentioned that cognitive-behavioral therapy (CBT) is also considered a first-line intervention for GAD. The most recent meta-analyses of CBT for GAD found a large pooled effect of psychotherapy versus placebo ($g = 0.84$; 95% CI: 0.71–0.97) with low to moderate heterogeneity[44, 45]. Therefore, clinicians should also consider CBT as another alternative or augmentation strategy for treating GAD.

Acknowledgments:

The authors thank J Curtiss from Boston University for his assistance in formulating the statistical analysis plan.

Funding:

This manuscript was not funded.

Declaration of Interest

SG Hofmann receives financial support from the Alexander von Humboldt Foundation (as part of the Humboldt Prize), as well as grants from the National Institutes of Health (NIH)/National Center for Complementary and Integrative Health (NCCIH) (R01AT007257), NIH/National Institute of Mental Health (NIMH) (R01MH099021, U01MH108168). He is also supported by the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition - Special Initiative and receives compensation for his work as an advisor from the Palo Alto Health Sciences and for his work as a Subject Matter Expert from John Wiley & Sons, Inc. and SilverCloud Health, Inc. He also receives royalties and payments for his editorial work from various publishers. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. One referee declares that they have worked as a consultant for Lundbeck.

References

1. American Psychiatric Association DSM-5 Task Force Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Arlington, VA: American Psychiatric Association, 2013
2. Bandelow B , Michaelis S , Wedekind D . Treatment of anxiety disorders. Dialogues in Clinical Neuroscience 2017;19:93–107 [PubMed: 28867934]
3. Newman MG , Llera SJ , Erickson TM , et al. Worry and generalized anxiety disorder. Annu Rev ClinPsychol2013;9:275–97**A well-written review of the literature on GAD, including: leading theoretical models; cognitive, affective, and neurobiological mechanisms; and the efficacy of various treatment options.
4. Reinhold JA , Rickels K . Pharmacological treatment for generalized anxiety disorder in adults. Expert OpinPharmacother2015;16:1669–81
5. Ravindran LN , Stein MB . The pharmacologic treatment of anxiety disorders. J. Clin. Psychiatry 2010;71:839–54 [PubMed: 20667290] *A well-written historical perspective and review of the literature on medication treatment for GAD.
6. Hidalgo RB , Tupler LA , Davidson JRT . An effect-size analysis of pharmacologic treatments for generalized anxiety disorder. J Psychopharmacol (Oxford) 2007;21:864–72 [PubMed: 17984162]
7. Buoli M , Dell’osso B , Bosi MF , et al. Slow vs standard up-titration of paroxetine in the treatment of panic disorder: a prospective randomized trial. Psychiatry ClinNeurosci2010;64:612–9
8. Delgado PL , Brannan SK , Mallinckrodt CH , et al. Sexual Functioning Assessed in 4 Double-Blind Placebo- and Paroxetine-Controlled Trials of Duloxetine for Major Depressive Disorder. J. Clin. Psychiatry 2005;66:686–92 [PubMed: 15960560]
9. Longo LP , Johnson B . Addiction: Part I. Benzodiazepines-side effects, abuse risk and alternatives. American family physician 2000;61:2121–8 [PubMed: 10779253]
10. Hidalgo RB , Sheehan DV . Benzodiazepines risk, abuse, and dependence: A tsunami in a tea cup. Psychiatry (Edgmont) 2010;6:13–5
11. Barker MJ , Greenwood KM , Jackson M , et al. Cognitive Effects of Long-Term Benzodiazepine Use. CNS Drugs 2004;18:37–48 [PubMed: 14731058]
12. Chan T-T , Leung WC- Y , Li V , et al. Association between high cumulative dose of benzodiazepine in Chinese patients and risk of dementia. Psychogeriatrics2017;17:310–6 [PubMed: 28145025]
13. Stewart SA . The effects of benzodiazepines on cognition. J. Clin. Psychiatry 2005;66 Suppl 2:9–13*A comprehensive review of the risk of cognitive impairments associated with long-term benzodiazepine use.
14. Koen N , Stein DJ . Pharmacotherapy of anxiety disorders. Dialogues in Clinical Neuroscience 2011;13:423–37 [PubMed: 22275848]
15. Ferreira-Garcia R , Mochcovitch M , Costa do Cabo M , et al. Predictors of pharmacotherapy response in generalized anxiety disorder. Harv Rev Psychiatry 2017;25:65–79 [PubMed: 28272131]
16. Baldwin DS , Woods R , Lawson R , et al. Efficacy of drug treatments for generalised anxiety disorder. BMJ 2011;342:d1199 [PubMed: 21398351] **A meta-analysis of 27 drug trials for GAD (primarily trials of SSRI, SNRIs, and pregabalin, plus a few of tigabine and lorazepam). Main findings favored fluoxetine and sertraline in terms of treatment response and tolerability.

17. Bandelow B , Reitt M , Röver C , et al. Efficacy of treatments for anxiety disorders. *IntClinPsychopharmacol*2015;30:183–92**A recent meta-analysis of 234 clinical trials of pharmacological, psychological and combined treatments for panic, GAD, and social phobia. Medication treatment outperformed psychological treatment, favoring SNRIs, followed by BZs and then SSRIs.
18. Cuijpers P , Sijbrandij M , Koole SL , et al. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders. *World Psychiatry* 2013;12:137–48 [PubMed: 23737423]
19. Fu J , Peng L , Li X . The efficacy and safety of multiple doses of vortioxetine for generalized anxiety disorder. *Neuropsychiatr Dis Treat* 2016;12:951–9 [PubMed: 27143896]
20. Gould RA , Otto MW , Pollack MH , et al. Cognitive Behavioral and Pharmacological Treatment of Generalized Anxiety Disorder: A Preliminary Meta-Analysis. *Behavior Therapy* 1997;28:285–305
21. Li X , Zhu L , Su Y , et al. Short-term efficacy and tolerability of venlafaxine extended release in adults with generalized anxiety disorder without depression. *PLOS ONE* 2017;12:e0185865 [PubMed: 28982121]
22. Mitte K , Noack P , Steil R , et al. A Meta-analytic Review of the Efficacy of Drug Treatment in Generalized Anxiety Disorder. *Journal of Clinical Psychopharmacology* 2005;25:141–50 [PubMed: 15738745]
23. Offidani E , Guidi J , Tomba E , et al. Efficacy and Tolerability of Benzodiazepines versus Antidepressants in Anxiety Disorders. *Psychotherapy and Psychosomatics* 2013;82:355–62 [PubMed: 24061211] **A meta-analysis of 22 controlled comparisons of ADs vs. BZs, whose findings favored BZs. This article provided key evidence motivating the comparison between drug classes in the present study.
24. Roest AM , de Jonge P , Williams CD , et al. Reporting Bias in Clinical Trials Investigating the Efficacy of Second-Generation Antidepressants in the Treatment of Anxiety Disorders. *JAMA Psychiatry* 2015;72:500–10 [PubMed: 25806940] **An influential report on various sources of publication bias present in efficacy trials of FDA-approved second-generation antidepressants for anxiety disorders. Results indicate that reporting biases led to significant increases in the number of positive findings in the literature.
25. Sugarman MA , Loree AM , Baltes BB , et al. The Efficacy of Paroxetine and Placebo in Treating Anxiety and Depression. *PLOS ONE* 2014;9:e106337 [PubMed: 25162656] *A well-designed meta-analysis examining the unique effects of paroxetine on anxiety and depression symptoms.
26. de Vries YA , de Jonge P , van den Heuvel E , et al. Influence of baseline severity on antidepressant efficacy for anxiety disorders. *Br J Psychiatry* 2016;208:515–21 [PubMed: 26989093] *A recent comprehensive review presenting clinically relevant findings that antidepressants are effective in treating anxiety disorders, regardless of baseline anxiety severity.
27. Zareifopoulos N , Dylja I . Efficacy and tolerability of vilazodone for the acute treatment of generalized anxiety disorder. *Asian J Psychiatr*2017;26:115–22
28. Zhang Y , Huang G , Yang S , et al. Duloxetine in treating generalized anxiety disorder in adults. *Asia Pac Psychiatry* 2016;8:215–25 [PubMed: 26238298]
29. Maier W , Buller R , Philipp M , et al. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *J Affect Disord*1988;14:61–8 [PubMed: 2963053]
30. Rosenthal R Meta-analytic procedures for social research. Thousand Oaks: Sage, 1991*The seminal report presenting a statistical solution to the problem of null or negative trial findings remaining unpublished.
31. Borenstein M , Rothstein H . Comprehensive meta-analysis: A computer program for research. Englewood: Biostat Inc, 1999*Highly intuitive, truly comprehensive statistical software.
32. Higgins JPT , Altman DG , Gøtzsche PC , et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928 [PubMed: 22008217]
33. Orwin RG . A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics* 1983;8:157
34. Rosenthal R The file drawer problem and tolerance for null results. *Psychological Bulletin* 1979;86:638–41

35. Duval S , Tweedie R . Trim and fill. *Biometrics* 2000;56:455–63 [PubMed: 10877304]
36. Peters JL , Sutton AJ , Jones DR , et al. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med* 2007;26:4544–62 [PubMed: 17476644]
37. Terrin N , Schmid CH , Lau J , et al. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22:2113–26 [PubMed: 12820277]
38. Higgins JPT , Green S , eds. *Cochrane Handbook for Systematic Reviews of Interventions*. 5th ed.: The Cochrane Collaboration, 2011th **This Handbook is an invaluable tool for any author conducting a literature review or meta-analysis.
39. Baldwin DS , Hou R , Huneke NTM , et al. Pharmacotherapy in generalized anxiety disorder. *CNS Drugs* 2017;31:307–17 [PubMed: 28303465] **A recent comprehensive review of new research targets for GAD, including the validation of a laboratory paradigm as a robust experimental model that mirrors the subjective, autonomic, and cognitive features of GAD.
40. Walsh BT , Seidman SN , Sysko R , et al. Placebo response in studies of major depression. *JAMA* 2002;287:1840 [PubMed: 11939870] **An influential report finding that placebo-response in published trials of antidepressant medication for MDD is often substantial, and has increased over time. Researchers should interpret with great caution the results of any antidepressant trial that does not include a proper placebo-control.
41. Shash D , Kurth T , Bertrand M , et al. Benzodiazepine, psychotropic medication, and dementia: A population-based cohort study. *Alzheimers Dement* 2016;12:604–13 [PubMed: 26602630]
42. Dobrea C , Buoli M , Arici C , et al. Tolerability and use in co-administration of pregabalin in affective patients: a 6-month prospective naturalistic study. *Expert Opin Drug Saf* 2012;11:893–9 [PubMed: 22946673]
43. Frampton JE , Foster RH . Pregabalin: in the treatment of generalised anxiety disorder. *CNS Drugs* 2006;20:685–94 [PubMed: 16863276]
44. Carpenter JK , Andrews LA , Witcraft SM , et al. Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depress Anxiety* 2018*A recent meta-analysis of 41 RCTs of CBT for anxiety-related disorders.
45. Cuijpers P , Sijbrandij M , Koole S , et al. Psychological treatment of generalized anxiety disorder: a meta-analysis. *ClinPsychol Rev* 2014;34:130–40
46. Alaka KJ , Noble W , Montejo A , et al. Efficacy and safety of duloxetine in the treatment of older adult patients with generalized anxiety disorder: a randomized, double-blind, placebo-controlled trial. *Int J Geriatr Psychiatry* 2014;29:978–86 [PubMed: 24644106]
47. Allgulander C , Dahl AA , Austin C , et al. Efficacy of sertraline in a 12-week trial for generalized anxiety disorder. *Am J Psychiatry* 2004;161:1642–9 [PubMed: 15337655]
48. Allgulander C , Hackett D , Salinas E . Venlafaxine extended release (ER) in the treatment of generalised anxiety disorder. *Br J Psychiatry* 2001;179:15–22 [PubMed: 11435263]
49. Allgulander C , Hartford J , Russell J , et al. Pharmacotherapy of generalized anxiety disorder: results of duloxetine treatment from a pooled analysis of three clinical trials. *Curr Med Res Opin* 2007;23:1245–52
50. Ansseau M , Olié J- P , Frenckell R von , et al. Controlled comparison of the efficacy and safety of four doses of suriclone, diazepam, and placebo in generalized anxiety disorder. *Psychopharmacology* 1991;104:439–43 [PubMed: 1685793]
51. Baldwin DS , Huusom AKT , Maehlum E . Escitalopram and paroxetine in the treatment of generalised anxiety disorder. *Br J Psychiatry* 2006;189:264–72 [PubMed: 16946363]
52. Bandelow B , Chouinard G , Bobes J , et al. Extended-release quetiapine fumarate (quetiapine XR). *Int J Neuropsychopharmacol* 2010;13:305–20 [PubMed: 19691907]
53. Bose A , Korotzer A , Gommoll C , et al. Randomized placebo-controlled trial of escitalopram and venlafaxine XR in the treatment of generalized anxiety disorder. *Depress Anxiety* 2008;25:854–61 [PubMed: 18050245]
54. Boyer WF , Feighner JP . A placebo-controlled double-blind multicenter trial of two doses of ipsapirone versus diazepam in generalized anxiety disorder. *IntClinPsychopharmacol* 1993;8:173–6
55. Brawman-Mintzer O , Knapp RG , Rynn M , et al. Sertraline treatment for generalized anxiety disorder. *J. Clin. Psychiatry* 2006;67:874–81 [PubMed: 16848646]

56. Brown GG , Ostrowitzki S , Stein MB , et al. Temporal profile of brain response to alprazolam in patients with generalized anxiety disorder. *Psychiatry Res* 2015;233:394–401 [PubMed: 26211623]
57. Ceulemans DLS , Hoppenbrouwers LJA , Gelders YG , et al. The influence of ritanserin, a serotonin antagonist, in anxiety disorders. *Pharmacopsychiatry* 1985;18:303–5 [PubMed: 3931107]
58. Cohn JB , Rickels K , Steege JF . A pooled, double-blind comparison of the effects of buspirone, diazepam and placebo in women with chronic anxiety. *Curr Med Res Opin* 1989;11:304–20
59. Cutler NR , Sramek JJ , Keppel Hesselink JM , et al. A double-blind, placebo-controlled study comparing the efficacy and safety of ipsapirone versus lorazepam in patients with generalized anxiety disorder. *Journal of Clinical Psychopharmacology* 1993;13:429–37 [PubMed: 7907097]
60. Davidson JRT , Bose A , Korotzer A , et al. Escitalopram in the treatment of generalized anxiety disorder: double-blind, placebo controlled, flexible-dose study. *Depress Anxiety* 2004;19:234–40 [PubMed: 15274172]
61. Davidson JRT , DuPont RL , Hedges D , et al. Efficacy, Safety, and Tolerability of Venlafaxine Extended Release and Buspirone in Outpatients With Generalized Anxiety Disorder. *J. Clin. Psychiatry* 1999;60:528–35 [PubMed: 10485635]
62. Enkelmann R Alprazolam versus buspirone in the treatment of outpatients with generalized anxiety disorder. *Psychopharmacology* 1991;105:428–32 [PubMed: 1798836]
63. Feltner DE , Crockatt JG , Dubovsky SJ , et al. A randomized, double-blind, placebo-controlled, fixed-dose, multicenter study of pregabalin in patients with generalized anxiety disorder. *Journal of Clinical Psychopharmacology* 2003;23:240–9 [PubMed: 12826986]
64. Ferreira L , Figueira M-L , Bessa-Peixoto A , et al. Psychomotor and anxiolytic effects of mexazolam in patients with generalised anxiety disorder. *Clinical Drug Investigation* 2003;23:235–43 [PubMed: 17535036]
65. Fontaine R , Chouinard G , Annable L . Bromazepam and diazepam in generalized anxiety. *Psychopharmacology Bulletin* 1984;20:126–7 [PubMed: 6144130]
66. Fontaine R , Mercier P , Beaudry P , et al. Bromazepam and lorazepam in generalized anxiety: a placebo-controlled study with measurement of drug plasma concentrations. *Acta Psychiatr Scand* 1986;74:451–8
67. Fresquet A , Sust M , Lloret A , et al. Efficacy and Safety of Lesopitron in Outpatients with Generalized Anxiety Disorder. *The Annals of Pharmacotherapy* 2000;34:147–53 [PubMed: 10676820]
68. Gelenberg AJ , Lydiard RB , Rudolph RL , et al. Efficacy of Venlafaxine extended-release capsules in nondepressed outpatients with generalized anxiety disorder. *JAMA* 2000;283:3082–8 [PubMed: 10865302]
69. Gommoll C , Durgam S , Mathews M , et al. A double-blind, randomized, placebo-controlled, fixed-dose phase III study of vilazodone in patients with generalized anxiety disorder. *Depress Anxiety* 2015;32:451–9 [PubMed: 25891440]
70. Gommoll C , Forero G , Mathews M , et al. Vilazodone in patients with generalized anxiety disorder: a double-blind, randomized, placebo-controlled, flexible-dose study. *Int Clin Psychopharmacol* 2015;30:297–306
71. Goodman WK , Bose A , Wang Q . Treatment of generalized anxiety disorder with escitalopram: pooled results from double-blind, placebo-controlled trials. *J Affect Disord* 2005;87:161–7 [PubMed: 15982747]
72. Hackett D , Haudiquet V , Salinas E . A method for controlling for a high placebo response rate in a comparison of venlafaxine XR and diazepam in the short-term treatment of patients with generalised anxiety disorder. *European Psychiatry* 2003;18:182–7 [PubMed: 12814852]
73. Hartford J , Kornstein S , Liebowitz M , et al. Duloxetine as an SNRI treatment for generalized anxiety disorder: results from a placebo and active-controlled trial. *Int Clin Psychopharmacol* 2007;22:167–74
74. Hidalgo RB , Tupler LA , Davidson JRT . An effect-size analysis of pharmacologic treatments for generalized anxiety disorder. *J Psychopharmacol (Oxford)* 2007;21:864–72 [PubMed: 17984162]

75. Kasper S , Herman B , Nivoli G , et al. Efficacy of pregabalin and venlafaxine-XR in generalized anxiety disorder: results of a double-blind, placebo-controlled 8-week trial. *IntClinPsychopharmacol*2009;24:87–96
76. Koponen H , Allgulander C , Erickson J , et al. Efficacy of duloxetine for the treatment of generalized anxiety disorder. *J. Clin. Psychiatry* 2007;9:100–7
77. Kragh-Sørensen P , Holm P , Fynboe C , et al. Bromazepam in generalized anxiety. *Psychopharmacology* 1990;100:383–6 [PubMed: 1969172]
78. Lenox-Smith A , Reynolds A . A double-blind, randomised, placebo controlled study of venlafaxine XL in patients with generalised anxiety disorder in primary care. *British Journal of General Practice* 2003;53:772–7 [PubMed: 14601352]
79. Mahableshwarkar AR , Jacobsen PL , Chen Y , et al. A randomised, double-blind, placebo-controlled, duloxetine-referenced study of the efficacy and tolerability of vortioxetine in the acute treatment of adults with generalised anxiety disorder. *Int J ClinPract*2014;68:49–59
80. McLeod DR , Hoehn-Saric R , Porges SW , et al. Effects of alprazolam and imipramine on parasympathetic cardiac control in patients with generalized anxiety disorder. *Psychopharmacology* 1992;107:535–40 [PubMed: 1603897]
81. Merideth C , Cutler AJ , She F , et al. Efficacy and tolerability of extended release quetiapine fumarate monotherapy in the acute treatment of generalized anxiety disorder: a randomized, placebo controlled and active-controlled study. *IntClinPsychopharmacol*2012;27:40–54
82. Möller H- J , Volz H- P , Reimann IW , et al. Opipramol for the Treatment of Generalized Anxiety Disorder: A Placebo-Controlled Trial Including an Alprazolam-Treated Group. *Journal of Clinical Psychopharmacology* 2001;21:59–65 [PubMed: 11199949]
83. Montgomery SA , Tobias K , Zornberg GL , et al. Efficacy and safety of pregabalin in the treatment of generalized anxiety disorder. *J. Clin. Psychiatry* 2006;67:771–82 [PubMed: 16841627]
84. Nicolini H , Bakish D , Dueñas H , et al. Improvement of psychic and somatic symptoms in adult patients with generalized anxiety disorder: examination from a duloxetine, venlafaxine extended-release and placebo-controlled trial. *Psychol Med* 2009;39:267–76 [PubMed: 18485261]
85. Nimatoudis I , Zissis NP , Kogeorgos J , et al. Remission rates with venlafaxine extended release in Greek outpatients with generalized anxiety disorder. A double-blind, randomized, placebo controlled study. *IntClinPsychopharmacol*2004;19:331–6
86. Pande AC , Crockatt JG , Feltner DE , et al. Pregabalin in generalized anxiety disorder: a placebo-controlled trial. *Am J Psychiatry* 2003;160:533–40 [PubMed: 12611835]
87. Pollack MH , Zaninelli R , Goddard A , et al. Paroxetine in the treatment of generalized anxiety disorder. *J. Clin. Psychiatry* 2001;62:350–7 [PubMed: 11411817]
88. Pourmotabbed T , McLeod DR , Hoehn-Saric R , et al. Treatment, Discontinuation, and Psychomotor Effects of Diazepam in Women With Generalized Anxiety Disorder. *Journal of Clinical Psychopharmacology* 1996;16:202–7 [PubMed: 8784650]
89. Power KG , Simpson RJ , Swanson V , et al. A controlled comparison of cognitive-behaviour therapy, Diazepam, and placebo, alone and in combination, for the treatment of generalised anxiety disorder. *Journal of Anxiety Disorders* 1990;4:267–92
90. Rickels K , DeMartinis N , Aufdembrinke B . A Double-Blind, Placebo-Controlled Trial of Abecarnil and Diazepam in the Treatment of Patients With Generalized Anxiety Disorder. *Journal of Clinical Psychopharmacology* 2000;20:12–8 [PubMed: 10653203]
91. Rickels K , Pollack MH , Feltner DE , et al. Pregabalin for treatment of generalized anxiety disorder. *Archives of General Psychiatry* 2005;62:1022–30 [PubMed: 16143734]
92. Rickels K , Pollack MH , Sheehan D , et al. Efficacy of extended-release venlafaxine in nondepressed outpatients with generalized anxiety disorder. *Am J Psychiatry* 2000;157:968–74 [PubMed: 10831478]
93. Rickels K , Schweizer E , DeMartinis N , et al. Gepirone and diazepam in generalized anxiety disorder. *Journal of Clinical Psychopharmacology* 1997;17:272–7 [PubMed: 9241006]
94. Rickels K , Zaninelli R , McCafferty J , et al. Paroxetine treatment of generalized anxiety disorder: a double-blind, placebo-controlled study. *Am J Psychiatry* 2003;160:749–56 [PubMed: 12668365]

95. Rynn M , Russell J , Erickson J , et al. Efficacy and safety of duloxetine in the treatment of generalized anxiety disorder: a flexible-dose, progressive-titration, placebo-controlled trial. *Depress Anxiety* 2008;25:182–9 [PubMed: 17311303]
96. Stein DJ , Ahokas A , Márquez MS , et al. Agomelatine in generalized anxiety disorder: an active comparator and placebo-controlled study. *J. Clin. Psychiatry* 2014;75:362–8 [PubMed: 24569045]
97. Steiner M , Allgulander C , Ravindran A , et al. Gender differences in clinical presentation and response to sertraline treatment of generalized anxiety disorder. *Hum Psychopharmacol* 2005;20:3–13 [PubMed: 15551351]
98. Wilcox CS , Ryan PJ , Morrissey JL , et al. A fixed-dose study of adinazolam-SR tablets in generalized anxiety disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 1994;18:979–93 [PubMed: 7824763]
99. Wu WY , Wang G , Ball SG , et al. Duloxetine versus placebo in the treatment of patients with generalized anxiety disorder in China. *Chinese Medicine Journal (Engl)* 2011;124:3260–8

Figure 1.
Flowchart of study selection.

Figure 2.
Funnel plot with imputed studies.

Figure 3.
Forest plot of effect sizes by drug class.

Study Characteristics

Table 1.

Study (first author)	Pub Year	Active drug	Drug class	% Female (total)	Mean Age (total)	TX (wks)	Outcome	Selection Bias	Attrition Bias	Reporting Bias	Total Bias
Anseau	1991	diazepam	BZ	56	41.0	4	HAMA	?			UNCLEAR
Boyer	1993	diazepam	BZ	NR	NR	4	HAMA	?			UNCLEAR
Brown	2015	alprazolam	BZ	60	37.0	4	HAMA	?	?		UNCLEAR
Ceulemans	1985	lorazepam	BZ	68	42.8	2	HARS	?	?	?	HIGH
Cohn	1989	diazepam	BZ	100	38.0	4	HAMA	?			UNCLEAR
Cutler	1993	lorazepam	BZ	64	NR	6	HAMA	?			UNCLEAR
Enkelmann	1991	alprazolam	BZ	49	35.0	6	HAMA	?	?		UNCLEAR
Feltner	2003	lorazepam	BZ	54	38.0	9	HAMA				LOW
Ferreira	2003	mexazolam	BZ	63	33.0	3	HAMA				LOW
Fontaine	1984	bromazepam	BZ	50	NR	4	HAMA	?	?	X	HIGH
Fontaine	1986	lorazepam	BZ	43	34	4	HAMA	?			UNCLEAR
Fresquet	2000	lorazepam	BZ	52	35.0	6	HAMA	?	?	X	HIGH
Kragh-Sorensen	1990	bromazepam	BZ	75	35	2	HAMA	?	?	?	HIGH
McLeod	1992	alprazolam	BZ	64	40.0	6	HAMA	?	X	X	HIGH
Moller	2001	alprazolam	BZ	67	48.0	4	HAMA	?			UNCLEAR
Pande	2003	lorazepam	BZ	59	35.8	4	HAMA	?	X		HIGH
Pourmotabbed	1996	diazepam	BZ	100	36.1	6	HAMA	?	?	?	HIGH
Power	1990	diazepam	BZ	78	41.0	10	HAMA	?	?		UNCLEAR
Rickels	1993	diazepam	BZ	67	39.0	8	HAMA	?			UNCLEAR
Rickels	1997	diazepam	BZ	61	41.0	8	HAMA	?			UNCLEAR
Rickels	2000	diazepam	BZ	60	39.0	6	HAMA				LOW
Rickels	2005	alprazolam	BZ	64	40.0	4	HAMA	?			UNCLEAR
Wilcox	1994	adinazolam	BZ	NR	NR	4	HAMA	?	?	X	HIGH
Alaka	2014	duloxetine	SNRI	77	71.0	10	HAMA				LOW
Allgulander	2001	venlafaxine	SNRI	61	45.0	24	HRSA	?	?		UNCLEAR
Allgulander	2007	duloxetine	SNRI	64	42.0	10	HAMA	?	?		UNCLEAR
Davidson	1999	venlafaxine	SNRI	61	38.0	8	HAD-A	?	?		UNCLEAR

Study (first author)	Pub Year	Active drug	Drug class	% Female (total)	Mean Age (total)	TX (wks)	Outcome	Selection Bias	Attrition Bias	Reporting Bias	Total Bias
Gelenberg	2000	venlafaxine	SNRI	59	39.0	28	HAMA	?	?		UNCLEAR
Hackett	2003	venlafaxine	SNRI	66	44.0	8	HAMA	?	?		UNCLEAR
Hartford	2007	venlafaxine	SNRI	63	40.8	10	HAMA				LOW
Kasper	2009	venlafaxine	SNRI	59	41.0	8	HARS		?		UNCLEAR
Koponen	2007	duloxetine	SNRI	68	43.8	9	HAMA				LOW
Lenox-Smith	2003	venlafaxine	SNRI	59	47.0	24	HAMA		?	X	HIGH
Mahableshwarkar	2014	duloxetine	SNRI	68	38.6	8	HAMA	?	?	?	HIGH
Montgomery	2006	venlafaxine	SNRI	62	44.1	6	HAMA				LOW
Nicolini	2009	venlafaxine	SNRI	57	42.8	10	HAMA				LOW
Nimatoudis	2004	venlafaxine	SNRI	67	42.0	8	HAMA	?	?		UNCLEAR
Rickels	2000	venlafaxine	SNRI	54	41.0	8	HAMA	?			UNCLEAR
Rynn	2008	duloxetine	SNRI	61	41.0	10	HAMA	?	?		HIGH
Wu	2011	duloxetine	SNRI	51	37.6	15	HAMA	?	?	?	HIGH
Allgulander	2004	sertraline	SSRI	54	41.0	12	HAMA				LOW
Baldwin	2006	escitalopram	SSRI	62	41.0	12	HAMA				LOW
Bandelow	2010	paroxetine	SSRI	51	52.0	8	HAMA		?		UNCLEAR
Bose	2008	escitalopram	SSRI	63	37.0	4	HAMA				LOW
Brawman-Mintzer	2006	sertraline	SSRI	58	40.0	10	HAMA				LOW
Davidson	2004	escitalopram	SSRI	52	39.0	8	HAMA	?	?	?	HIGH
Gommoll, a	2015	vilazodone	SSRI	63	39.0	8	HAMA				LOW
Gommoll, b	2015	vilazodone	SSRI	69	40.0	8	HAMA				LOW
Goodman - #1	2005	escitalopram	SSRI	55	39.0	8	HAMA	?	?	X	HIGH
Goodman - #3	2005	escitalopram	SSRI	55	39.0	8	HAMA	?	?	X	HIGH
Merideth	2012	escitalopram	SSRI	64	38.0	8	HAMA		?		UNCLEAR
Pollack	2001	paroxetine	SSRI	63	40.0	8	HAMA	?		?	UNCLEAR
Rickels	2003	paroxetine	SSRI	56	40.0	8	HAMA		?		UNCLEAR
Stein	2014	escitalopram	SSRI	73	43.6	12	HAMA		?	?	UNCLEAR
Steiner - females	2005	sertraline	SSRI	50*	41.3	12	HAMA	?	X	X	HIGH
Steiner - males	2005	sertraline	SSRI	50*	41.3	12	HAMA	?	X	X	HIGH

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TX: treatment; HARS: Hamilton Anxiety Rating Scale; BZ: benzodiazepine; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor; HAM-A: Hamilton anxiety scale; Pub: Publication; wks: weeks; : low risk of bias; ? : unclear risk of bias; X: high risk of bias.

*The Steiner (2005) study separated their sample by gender so the % female is technically 100% and 0% for each sample, respectively, but was listed as 50% for each, so as to maintain statistical normality.

Table 2.

Effect Sizes and Statistics for Each Study

Study (first author)	Pub Year	Drug class	Hedges's g	Standard error	Variance	Statistics for each study			Sample Sizes		
						Lower limit	Upper limit	Z-Value	p-Value	TX	PL
Anseau	1991	BZ	0.651	0.194	0.037	0.271	1.030	3.361	0.001	54	57
Boyer	1993	BZ	0.504	0.195	0.038	0.121	0.886	2.582	0.010	55	52
Brown	2015	BZ	1.250	0.402	0.162	0.461	2.039	3.106	0.002	19	11
Ceulemans	1985	BZ	0.613	0.310	0.096	0.004	1.221	1.974	0.048	22	20
Cohn	1989	BZ	0.435	0.132	0.017	0.176	0.693	3.294	0.001	113	121
Cutler	1993	BZ	0.358	0.190	0.036	-0.014	0.730	1.885	0.059	61	51
Enkelmann	1991	BZ	0.916	0.291	0.084	0.347	1.486	3.152	0.002	32	21
Feltner	2003	BZ	0.373	0.176	0.031	0.028	0.718	2.121	0.034	64	66
Ferreira	2003	BZ	0.339	0.257	0.066	-0.166	0.843	1.317	0.188	32	28
Fontaine	1984	BZ	1.257	0.379	0.143	0.514	1.999	3.318	0.001	16	16
Fontaine	1986	BZ	0.701	0.320	0.102	0.075	1.328	2.194	0.028	20	20
Fresquet	2000	BZ	0.409	0.287	0.082	-0.154	0.971	1.424	0.154	30	20
Kragh-Sorensen	1990	BZ	0.345	0.176	0.031	0.001	0.689	1.963	0.050	97	49
McLeod	1992	BZ	0.558	0.374	0.140	-0.176	1.292	1.491	0.136	14	14
Moller	2001	BZ	0.270	0.139	0.019	-0.003	0.542	1.937	0.053	102	105
Pande	2003	BZ	0.742	0.176	0.031	0.398	1.086	4.222	0.000	68	69
Pourmotabbed	1996	BZ	0.609	0.430	0.185	-0.234	1.451	1.416	0.157	10	11
Power	1990	BZ	0.487	0.312	0.097	-0.124	1.098	1.562	0.118	22	19

Study (first author)	Pub Year	Drug class	Hedges's g	Statistics for each study				Sample Sizes		
				Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	TX PL
Rickels	1993	BZ	0.512	0.198	0.039	0.123	0.900	2.582	0.010	49 55
Rickels	1997	BZ	0.583	0.177	0.031	0.236	0.929	3.297	0.001	67 65
Rickels	2000	BZ	0.461	0.140	0.020	0.187	0.736	3.295	0.001	104 104
Rickels	2005	BZ	0.425	0.153	0.023	0.125	0.725	2.777	0.005	88 85
Wilcox	1994	BZ	0.933	0.453	0.205	0.045	1.820	2.059	0.040	10 10
Total BZ Random Effects			0.497	0.043	0.002	0.413	0.582	11.530	0.000	1149 1069
Alaka	2014	SNRI	0.604	0.123	0.015	0.362	0.846	4.897	0.000	143 131
Allgulander	2001	SNRI	0.292	0.122	0.015	0.052	0.531	2.389	0.017	140 130
Allgulander	2007	SNRI	0.349	0.060	0.004	0.232	0.466	5.839	0.000	668 495
Davidson	1999	SNRI	0.419	0.149	0.022	0.127	0.711	2.811	0.005	86 97
Gelenberg	2000	SNRI	0.431	0.131	0.017	0.174	0.687	3.294	0.001	115 123
Hackett	2003	SNRI	0.414	0.210	0.044	0.001	0.826	1.965	0.049	68 34
Hartford	2007	SNRI	0.411	0.112	0.013	0.192	0.630	3.672	0.000	164 161
Kasper	2009	SNRI	0.032	0.125	0.016	-0.213	0.278	0.258	0.796	125 128
Koponen	2007	SNRI	0.357	0.108	0.012	0.144	0.569	3.293	0.001	170 175
Lenox-Smith	2003	SNRI	0.251	0.128	0.016	0.000	0.503	1.962	0.050	122 122
Mahabeshwarkar	2014	SNRI	0.375	0.116	0.013	0.149	0.602	3.246	0.001	149 154
Montgomery	2006	SNRI	0.333	0.139	0.019	0.062	0.605	2.404	0.016	110 100
Nicolini	2009	SNRI	0.458	0.113	0.013	0.236	0.679	4.055	0.000	158 163
Nimatoudis	2004	SNRI	1.023	0.309	0.096	0.417	1.629	3.309	0.001	24 22

Study (first author)	Pub Year	Drug class	Hedges's g	Statistics for each study					Sample Sizes		
				Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	TX	PL
Rickels	2000	SNRI	0.323	0.149	0.022	0.032	0.615	2.173	0.030	86	96
Rynn	2008	SNRI	0.252	0.111	0.012	0.035	0.469	2.275	0.023	168	159
Wu	2011	SNRI	0.320	0.139	0.019	0.046	0.593	2.293	0.022	107	100
Total SNRI Random Effects											
Allgulander	2004	SSRI	0.496	0.105	0.011	0.289	0.702	4.705	0.000	182	188
Baldwin	2006	SSRI	0.302	0.122	0.015	0.063	0.541	2.472	0.013	132	138
Bandelow	2010	SSRI	0.249	0.097	0.009	0.060	0.438	2.577	0.010	214	217
Bose	2008	SSRI	0.298	0.140	0.019	0.024	0.572	2.135	0.033	102	104
Brawman-Mintzer	2006	SSRI	0.166	0.110	0.012	-0.050	0.383	1.508	0.132	165	163
Davidson	2004	SSRI	0.573	0.116	0.014	0.346	0.801	4.934	0.000	154	153
Gommoll, a	2015	SSRI	0.247	0.095	0.009	0.060	0.433	2.595	0.009	223	221
Gommoll, b	2015	SSRI	0.220	0.101	0.010	0.023	0.417	2.184	0.029	198	197
Goodman - #1	2005	SSRI	0.308	0.126	0.016	0.060	0.556	2.438	0.015	124	128
Goodman - #3	2005	SSRI	0.573	0.116	0.014	0.346	0.801	4.934	0.000	154	153
Merideth	2012	SSRI	0.193	0.098	0.010	0.000	0.385	1.961	0.050	203	212
Pollack	2001	SSRI	0.287	0.111	0.012	0.069	0.506	2.578	0.010	161	163
Rickels	2003	SSRI	0.256	0.102	0.010	0.056	0.456	2.507	0.012	197	188
Stein	2014	SSRI	0.436	0.123	0.015	0.195	0.677	3.550	0.000	139	131
Steiner - females	2005	SSRI	0.406	0.105	0.011	0.200	0.611	3.868	0.000	182	188
Steiner - males	2005	SSRI	0.295	0.104	0.011	0.090	0.499	2.825	0.005	182	188

Study (first author)	Pub Year	Drug class	Hedges's g	Statistics for each study					Sample Sizes		
				Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	TX	PL
Total SSRI Random Effects				0.032	0.001	0.263	0.387	10.242	0.000	2712	2732
Overall Random Effects				0.020	0.000	0.335	0.414	18.690	0.000	6464	6191

TX: treatment; PL: placebo; Pub: Publication; BZ: benzodiazepine; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.