

## Original Article

**Cite this article:** Ahn-Horst RY, Turner EH (2024). Unpublished trials of alprazolam XR and their influence on its apparent efficacy for panic disorder. *Psychological Medicine* 54, 1026–1033. <https://doi.org/10.1017/S0033291723002830>

Received: 13 May 2023  
Revised: 10 August 2023  
Accepted: 4 September 2023  
First published online: 19 October 2023

**Keywords:**

anxiety disorders/drug therapy; benzodiazepines/therapeutic use; meta-analysis; publication bias; randomized controlled trials; United States Food and Drug Administration

**Corresponding author:**

Erick H. Turner;  
Email: [turnere@ohsu.edu](mailto:turnere@ohsu.edu)

# Unpublished trials of alprazolam XR and their influence on its apparent efficacy for panic disorder

Rosa Y. Ahn-Horst<sup>1,2,3</sup> and Erick H. Turner<sup>4,5</sup> 

<sup>1</sup>Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; <sup>2</sup>Department of Psychiatry, McLean Hospital, Belmont, MA, USA; <sup>3</sup>Department of Psychiatry, Harvard Medical School, Boston, MA, USA; <sup>4</sup>Behavioral Health and Neurosciences Division, Veterans Affairs Portland Health Care System, Portland, OR, USA and <sup>5</sup>Department of Psychiatry, Oregon Health & Science University, Portland, OR, USA

**Abstract**

**Objective.** To test for publication bias with alprazolam, the most widely prescribed benzodiazepine, by comparing its efficacy for panic disorder using trial results from (1) the published literature and (2) the US Food and Drug Administration (FDA).

**Methods.** From FDA reviews, we included data from all phase 2/3 efficacy trials of alprazolam extended-release (Xanax XR) for the treatment of panic disorder. A search for matching publications was performed using PubMed and Google Scholar. Publication bias was examined by comparing: (1) overall trial results (positive or not) according to the FDA *v.* corresponding publications; (2) effect size (Hedges's *g*) based on FDA data *v.* published data.

**Results.** The FDA review showed that five trials were conducted, only one of which (20%) was positive. Of the four not-positive trials, two were published conveying a positive outcome; the other two were not published. Thus, according to the published literature, three trials were conducted and all (100%) were positive. Alprazolam's effect size calculated using FDA data was 0.33 (CI<sub>95%</sub> 0.07–0.60) *v.* 0.47 (CI<sub>95%</sub> 0.30–0.65) using published data, an increase of 0.14, or 42%.

**Conclusions.** Publication bias substantially inflates the apparent efficacy of alprazolam XR.

**Introduction**

Randomized controlled trials play a major role in forming the evidence base and shaping clinical practice. However, clinical decision-making is based on accessible and published studies. Trials with statistically significant results are more likely to be published than trials with non-significant results, thus inflating estimates of drug efficacy and safety (Dickersin, Chan, Chalmers, Sacks, & Smith, 1987; Suñé, Suñé, & Montoro, 2013).

The degree to which published trial results overestimate efficacy can be determined using FDA review documents (Turner, 2013). In the US, drug companies are required to register with the FDA all trials they intend to conduct for purposes of US marketing approval. Results of these trials are documented in FDA medical and statistical reviews. These are publicly available for download at Drugs@FDA (<https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>) for drugs approved by the FDA since 1998 (Turner, 2013). Reviews of drugs approved prior to 1998 require a request through the Freedom of Information Act (<https://www.accessdata.fda.gov/scripts/foi/FOIRequest/requestinfo.cfm>).

To date, research utilizing FDA data to examine reporting biases has been mostly limited to a few psychiatric drug classes, especially drugs approved for the treatment of major depressive disorder (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008), anxiety disorders (Roest et al., 2015) and schizophrenia (Turner, Knoepfmacher, & Shapley, 2012). Similar work on additional drug classes is needed.

Benzodiazepines are prescribed to over 5% of US adults (Bachhuber, Hennessy, Cunningham, & Starrels, 2016) and are involved in nearly 14% of fatal opioid overdoses (National Institute on Drug Abuse, 2022), potentiating their lethality (Boon, van Dorp, Broens, & Overdyk, 2020). While their potential have been acknowledged, the consensus view appears to be that benzodiazepines are 'highly effective' (Silberman et al., 2021) and 'efficacious for the short- and long-term treatment of anxiety disorders (Nardi & Quagliato, 2022)'. The American Psychiatric Association Practice Guideline for the treatment of panic disorder recommends the use of benzodiazepines (along with SSRIs, SNRIs, TCAs, or CBT) as initial therapy, citing 'demonstrated efficacy in numerous controlled trials' (Stein et al., 2010). Also, multiple meta-analyses have reported that benzodiazepines may be superior to placebo in the treatment of panic disorder (Bighelli et al., 2016; Boyer, 1995; Breilmann et al., 2019; Wilkinson, Balestrieri, Ruggeri, & Bellantuono, 1991). Several of these

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

meta-analyses acknowledged the possibility of publication bias, but few have searched for unpublished trial data. A Cochrane review of 24 trials comparing various benzodiazepines to placebo (Breilmann et al., 2019) included data from just one unpublished trial (however, please see Discussion). Notably, the authors questioned the superiority of benzodiazepines over placebo, in part because they inferred ‘probable publication bias’ based on funnel plot asymmetry. To our knowledge, no studies have formally investigated publication bias with benzodiazepines by searching for unpublished benzodiazepine trial data in regulatory documents and reassessing their efficacy.

Among all benzodiazepines, the one with the highest prescribing rate in the US – and with the highest frequency for all measures of nonmedical use, abuse, and related harms – is alprazolam (Ait-Daoud, Hamby, Sharma, & Blevins, 2018). In the current study, we examined efficacy data for alprazolam, specifically its extended-release formulation (Xanax XR), for the treatment of panic disorder. Because it was approved in 2003, its FDA reviews are available for download from Drugs@FDA. By contrast, data on the original immediate-release formulation of alprazolam, and all other benzodiazepines, are much less accessible, since they were approved many years before the 1997 launch of Drugs@FDA. Our objective was to compare alprazolam XR’s efficacy according to the published literature with its efficacy according to the FDA. Efficacy was examined in terms of overall trial outcome (positive or not) and meta-analytic effect size.

## Methods

### Data from FDA reviews

At Drugs@FDA (<http://www.accessdata.fda.gov>), we downloaded the medical and statistical reviews for alprazolam XR. From these we identified all phase 2/3 randomized, double-blind, placebo-controlled trials of alprazolam XR for the treatment of panic disorder with or without agoraphobia; patients included were adult male and female outpatients ages 18–65. The FDA reviews of these trials identified multiple primary outcome measures, among which, the following five were common to all trials: (1) mean change from baseline in total number of panic attacks, (2) percentage of patients achieving zero panic attacks, (3) mean change from baseline in Clinical Global Impression–Severity, (4) Clinical Global Impression–Change in Condition, and (5) overall phobia state. For each of these, summary statistics were extracted for purposes of meta-analysis (see below). For each trial, we also extracted the FDA’s regulatory decision, that is, whether, for purposes of approval, the trial was judged to be positive (supportive of efficacy).

Regarding one trial (Study 5, #0032), the FDA reported its overall conclusion but no summary statistics other than its sample size. Since this trial was unpublished, we established email contact with the sponsor, but our follow-up emails were not answered. We also requested additional data on this trial from the FDA’s Freedom of Information Office, but they estimated that our ‘complex track request’ would not be processed until fall 2024.

### Data from journal articles

For each FDA-registered premarketing trial, we searched PubMed, bibliographies of review articles (Breilmann et al., 2019; Perugi, Frare, & Toni, 2007) and Google Scholar. In

PubMed, we used the following search syntax: (alprazolam [Title/Abstract] OR xanax[Title/Abstract]) AND (XR[Title/Abstract] OR extended release[Title/Abstract] OR sustained release) AND placebo[Title/Abstract]. In Google Scholar, we combined ‘alprazolam XR’ (and ‘alprazolam extended release’) with ‘placebo’ and ‘panic disorder’ and reviewed the first ten pages of search results. Author RA identified the best match between FDA-registered clinical trials and journal articles based on drug name, comparator, dosage groups, associated sample sizes, trial duration, and investigator names. Stand-alone publications (i.e. full article reporting the results of a single trial) were preferred. For any given trial, if no stand-alone publication could be found, we included publications aggregating the results of multiple trials, as long as they did not cover other trials separately published in stand-alone format (Turner et al., 2012).

Each trial publication conclusion was classified as positive or not positive based on the overall conclusion expressed in the abstract. If no abstract was available, we used the overall conclusion expressed in the conclusion/discussion section. We then noted whether the two data sources (trial publication v. FDA review) (dis)agreed regarding the trial’s outcome.

For purposes of meta-analysis (below), both authors independently identified and extracted summary data on the apparent primary outcome, that is, the drug-placebo comparison reported first in the text of the results section or in the table or figure first cited in the text (Turner et al., 2008).

### Statistical analysis

We conducted two meta-analyses, one based on the FDA reviews and one based on corresponding publications. For outcomes based on continuous data, the measure of effect size was Hedges’s  $g$  (Hedges, 1982), calculated as:

$$g = t \times \sqrt{\frac{1}{n_{\text{drug}}} + \frac{1}{n_{\text{placebo}}}}$$

(This method is algebraically equivalent to using means and standard deviations (Rosenthal, 1991) which, in our experience, are inconsistently provided in FDA reviews.) The  $t$  statistic was calculated from  $p$  and  $N$  using the TINV function in Excel, multiplying  $t$  by  $-1$  if drug underperformed placebo. If a precise  $p$  value was unavailable (e.g.  $p < 0.05$ ),  $t$  was calculated from other summary statistics (SDs, SEs, 95% CIs), if available; otherwise, we calculated  $t$  from the upper limit of the  $p$  value range. There were two multiple-dose trials for which we calculated a single study-level effect size, as done previously (Roest et al., 2015; Turner et al., 2008, 2012; Turner, Cipriani, Furukawa, Salanti, & de Vries, 2022), avoiding a spuriously low standard error by counting the shared placebo  $N$  only once.

For one outcome based on count data (percentage of patients achieving zero panic attacks), we used the `csi` command in Stata 11 (StataCorp LP, 2009) to calculate the odds ratio (OR), which we converted to Cohen’s  $d$  using the method of Cox and Snell (Anzures-Cabrera, Sarpatwari, & Higgins, 2011; Thorlund, Walter, Johnston, Furukawa, & Guyatt, 2011).

$$d = \frac{\ln \text{OR}}{1.65}$$

(We considered the method of Hasselblad and Hedges, which differs only in that it employs a slightly larger denominator (1.81), but this would have led to smaller FDA-based *d* values, increasing the gap between them and the corresponding journal-based *d* values.)

The variance of *d* was calculated as:

$$v_d = \frac{((1/a) + (1/b) + (1/c) + (1/c))}{1.65^2}$$

where *a*, *b*, *c*, and *d* are the cell frequencies in the 2-by-2 contingency table (Wilson, 2017). Cohen's *d* was multiplied by the correction factor *J*, where

$$J = 1 - \frac{4}{4df - 1}$$

to obtain Hedges's *g* (Cooper, Hedges, & Valentine, 2019) and the standard error of *g* was calculated as the square root of *v<sub>g</sub>*, where

$$v_g = J^2 \times v_d$$

For each FDA-reviewed clinical trial, we calculated a composite effect size using the arithmetic mean of the five outcome-level effect sizes and of their variances (López-López, Page, Lipsey, & Higgins, 2018).

Study 5 (trial #0032) could not be incorporated into the meta-analysis due to insufficient statistical summary data (above). Statistical analyses were independently performed by RA using R 4.0.2 (R Core Team, 2020) and by ET using Stata 11 (StataCorp LP, 2009); their results were compared and reconciled.

**Results**

The numerical results extracted from the FDA review and from the published literature are presented in Table 1.

**Study 1 (2000/0271)**

*Results per FDA*

Study 1 was conducted at three centers in two countries (US and Canada) from July 1986 to March 1989 (M-53/56). In this study, 208 male and female adult outpatients with panic disorder with or without agoraphobia were treated, 69 with alprazolam XR, 70 with alprazolam immediate release (IR, not included in this article's tables or figures), and 69 with placebo (M-17/56 and M-53/56) (Abbreviations here refer to pages 17 and 53 of the 56-page FDA medical (M) review. Henceforth, 'S-' refers to the FDA statistical review). The FDA statistician stated Study 1 was 'positive in 5 of 7 primary endpoints' (S-4/26). However, because the FDA deems studies positive only if all primary endpoints achieve statistical significance (FDA, 2022), the FDA medical review included Study 1 as one of four that had 'failed on face' (M-5/56 and M-15/56).

*Results published*

The article on Study 1 (Pecknold, Luthe, Munjack, & Alexander, 1994) presented it as a positive trial and gave no indication that, overall, it should be considered negative. The abstract stated, 'On global measures, Hamilton Rating Scale for Anxiety, phobia rating, and work disability measures, both active treatment groups

**Table 1.** Summary statistical data extracted from FDA review and journal articles

Study number (Sponsor trial number)	Dose, mg/day	Duration, treatment (discontinuation) periods	Results from FDA review on primary outcomes					Results from journal article (apparent primary outcome)
			CGI-Mean change in condition	CGI-Severity, mean change from baseline	Overall phobia state, mean change from baseline	# patients achieving zero <sup>a</sup>	Panic attacks, mean change in # v. baseline	
Study 1 (2000/0271)	1-10 mg (mean 4.4 ± 2.3)	6 wks (16 wks)	<i>p</i> = 0.0008 <sup>b</sup> [68;65]	<i>p</i> = 0.009 [68;65]	<i>p</i> = 0.002 [67;67]	36[32 34] 33 OR 1.09 (0.56, 2.14)	<i>p</i> = 0.1104 <sup>b</sup> [67;66]	<i>p</i> < 0.001 <sup>c</sup> [56;45]
Study 2 (2000/0369)	1-10 mg (mean 4.7 ± 2.5)	6 wks (5 wks)	<i>p</i> = 0.0001 [85;78]	<i>p</i> = 0.0001 [84;78]	<i>p</i> = 0.0005 [83;74]	61[24 32 44 OR 3.49 (1.81, 6.71)	<i>p</i> = 0.0026 [85;76]	<i>p</i> < 0.01 <sup>c</sup> [102;92]
Study 3 (2002/0002)	4 mg 6 mg	10 wks (4 wks)	<i>p</i> = 0.0872 <sup>b</sup> [152;73]	<i>p</i> = 0.3351 <sup>b</sup> [152;73]	<i>p</i> = 0.9765 <sup>b</sup> [149;73]	77[75 35 38 OR 1.11 (0.64, 1.94)	<i>p</i> = 0.5391 <sup>b</sup> [152;73]	<i>p</i> = 0.014 [76;67] <i>p</i> = 0.005 [65;67]
Study 4 (2002/0003)	4 mg 6 mg	10 wks (4 wks)	<i>p</i> = 0.2996 <sup>b</sup> [173;79]	<i>p</i> = 0.07 [173;79]	<i>p</i> = 0.2159 <sup>b</sup> [172;79]	78[95 36 43 OR 0.98 (0.58, 1.67)	<i>p</i> = 0.26 [173;79]	Not published
Study 5 (2002/0032)	4 mg	8 wks (4 wks)	NS [23;24]	NS [23;24]	NS [23;24]	NS [23;24]	NS [23;24]	Not published

CGI, Clinical Global Impression rating scale. Sample sizes, shown as [M<sub>drug</sub>; M<sub>placebo</sub>], represent number of patients analyzed, which may be slightly less than number treated, as given in text.

<sup>a</sup>Count data reported in FDA review (alprazolam yes|alprazolam no|placebo yes|placebo no), from which authors calculated odds ratios (OR) ± 95% confidence intervals.

<sup>b</sup>Precise *p* value calculated from means, SDs, and Ns; dose groups pooled.

<sup>c</sup>Upper limit of *p* value range used to calculate Hedges's *g*. Brackets: [M<sub>drug</sub>; M<sub>placebo</sub>]. Gray shading indicates nonsignificant outcomes. NS, outcomes assumed nonsignificant because FDA stated trial had 'failed on face'.

were equally effective and significantly more efficacious than the placebo cell on endpoint MANOVA analysis. On analysis of the panic factor...significantly more effective than the placebo group'. The Results section, Outcome Measures subsection, began, 'Both the physician-rated and patient-treated scales estimating overall improvement showed that [sic] both CT [compressed tablet] alprazolam and the XR alprazolam to be superior to placebo. For the patient-rated scale... $p < 0.001$  on all weeks' for the patient-rated scale. The remainder of the Results section presented several additional statistically significant results. The Comment section began: 'During this 6-week trial, both the CT and the XR alprazolam patients had significantly more improved ratings than the placebo group on most of the outcome variables'.

### Study 2 (2000/0369)

#### Results per FDA

Study 2 was conducted at three centers in the United States from June 1988 to January 1990. In this study, 200 male and female adult outpatients with panic disorder were treated, 104 patients with alprazolam XR and 96 with placebo (M-17 and -53/56). Of the five alprazolam XR trials, Study 2 was the only study the FDA deemed clearly positive: 'All 7 co-primary efficacy endpoints were statistically significantly superior in the group treated with alprazolam XR compared to the group treated with placebo' (M-47/56).

The fact that just one positive study was deemed sufficient for alprazolam XR's approval was apparently not the case earlier: 'This study was first submitted [~33 characters redacted], but the NDA [new drug application] was [~31 characters redacted] due to the fact that there was only one positive study and per guideline at that time required two positive studies for approval of the submission (S-5/26)'.

The FDA review also reported an irregularity at one of the three study sites: '...[site] inspection findings were as follows: '28 of 37 subject records were not available for review, as Dr Rosenthal had destroyed these records in March 1999...validity of the data reported could not be verified...it is recommended that the Review Division should consider excluding all data generated at this site and reanalyzing efficacy data in support of this NDA' (M-11/56). After doing so, the FDA review division found, 'these results do not change the conclusion' (S-12/26).

#### Results published

Study 2 was published as positive (Schweizer, Patterson, Rickels, & Rosenthal, 1993). However, the above-mentioned investigator was included as a co-author, as were the data from his site.

### Study 3 (2002/0002)

#### Results per FDA

Study 3 was conducted at 15 centers in the United States from June 1990 to October 1991. In this study, 231 male and female adult outpatients with panic disorder were treated, 155 with alprazolam XR and 76 with placebo (M-17 and -53/56). According to the FDA, the primary analyses involved comparisons for primary endpoints at week eight for all patients, using the last-observation-carried-forward (LOCF) method of handling dropouts (Committee for Proprietary Medicinal Products, 2001; Leon et al., 2006). As reported in Table 1 and the FDA review (S-24/26), the results were nonsignificant on 5 of 5 primary outcomes.

#### Results published

In the corresponding journal publication, Study 3 (Alexander, 1993) was presented as positive, contrary to the FDA report. The first half of the results section (there was no abstract) was an 'Initial Analysis', which began, 'Statistically significant improvements...with both doses of alprazolam XR within the first week of treatment'. The second sentence referred to Table 1, whose title indicated that the results were observed values, that is, obtained via complete case analysis, which, unlike the (primary) LOCF method, violates the intention to treat principle by simply omitting the data from patients who drop out (Committee for Proprietary Medicinal Products, 2001; Leon et al., 2006). Table 1 displayed Week 1 results of  $p = 0.014$  and  $p = 0.0005$  for the two doses, respectively. This table displayed  $p$  values only for one other time point, Week 6, both with  $p < 0.05$ . Results for Week 8, the primary analysis time point (see FDA section above), were not shown. Later, in the fourth paragraph or the results section, LOCF results were mentioned, but two of its three sentences reported on the relative performance of the two active dose groups; its last sentence acknowledged that 'Similar reductions were seen in the placebo group...' but not the corresponding nonsignificant  $p$  values.

The second half of the Results section was entitled 'Analysis Excluding High Placebo-Response Centers', for which, judging from the FDA review, there was no prespecified plan. This section's first sentence reported, 'excluding four [of 15] centers with unusually high placebo responses, revealed clear differences between alprazolam-XR and placebo at every time point and on all efficacy measures'. This section devoted two tables and two figures to these post hoc results. Table 2 displayed ten  $p$  values (2 doses  $\times$  5 time points), including seven with  $p < 0.05$ ; Table 3 displayed 16  $p$  values, including 14 with  $p < 0.05$ .

The Conclusion section began, 'The results of this study indicate that both 4 and 6 mg alprazolam-XR are effective in reducing panic attacks and producing clinical improvements in patients with panic disorder'. The Conclusion section ended, '...active treatment with alprazolam-XR clearly demonstrates that twice-daily dosing with an extended-release preparation is an effective treatment for panic disorder'.

### Study 4 (2002/0003)

#### Results per FDA

Study 4 was conducted at 15 centers in the United States from May 1990 to October 1991. In this study, 261 male and female adult outpatients with panic disorder were treated, 178 patients with alprazolam XR and 83 with placebo (M-17 and -53/56). As reflected in Table 1, the FDA review (S-21/26) reported nonsignificant on all five primary endpoints.

#### Results (un)published

There was no stand-alone paper for Study 4; it was mentioned briefly in a review publication (Stahl, 1993) also covering Studies 1–3, which were separately published in stand-alone format. Consequently, Study 4 was considered not transparently published.

### Study 5 (2002/0032)

#### Results per FDA

Study 5 was conducted at one center in the United States from 1994 to 1995. In this study, 50 male and female outpatients

**Table 2.** Effect size (ES) values (Hedges's  $g \pm 95\%$  CI) calculated from primary outcome results extracted from FDA review and shown in Table 1

Trial ID number	CGI-Mean change in condition	CGI-Severity, mean change from baseline	Overall phobia state, mean change from baseline	Panic attacks, # patients achieving zero	Panic attacks, mean change in # v. baseline	Composite across the 5 primary outcomes
Study 1 (2000/0271)	0.59 (0.24–0.94)	0.50 (0.15–0.84)	0.56 (0.21–0.90)	0.05 (–0.36 to 0.46)	0.28 (–0.07 to 0.62)	0.40 (0.04–0.75)
Study 2 (2000/0369)	0.81 (0.49–1.13)	0.95 (0.62–1.27)	0.54 (0.22–0.86)	0.75 (0.36–1.15)	0.48 (0.17–0.79)	0.71 (0.37–1.04)
Study 3 (2002/0002)	0.24 (–0.04 to 0.52)	0.17 (–0.11 to 0.45)	0.00 (–0.28 to 0.28)	0.07 (–0.27 to 0.40)	0.06 (–0.22 to 0.34)	0.11 (–0.18 to 0.40)
Study 4 (2002/0003)	0.15 (–0.12 to 0.41)	0.25 (–0.02 to 0.52)	0.17 (–0.10 to 0.43)	–0.01 (–0.34 to 0.31)	0.15 (–0.12 to 0.42)	0.14 (–0.14 to 0.42)
Study 5 (2002/0032)	Study 5 omitted from meta-analysis due to lack of summary statistics					
Overall	0.43 (0.13–0.74)	0.46 (0.12–0.79)	0.30 (0.04–0.57)	0.21 (–0.13 to 0.54)	0.23 (0.05–0.41)	0.33 (0.06–0.60)

Gray shading indicates outcomes with nonsignificant results. Columns represent results for the five primary outcomes and a composite ES across outcomes; rows represent the various studies. CGI refers to Clinical Global Impression rating scale.

with panic disorder with or without agoraphobia were treated, 23 patients with alprazolam XR and 24 were treated with placebo (M-17 and -53/56). Patients in both treatment groups also received cognitive-behavioral therapy. Beyond sample sizes and the fact the results were nonsignificant, the FDA provided no statistical results. Referring to this plus three other studies, the FDA medical review (M-15/56) stated, '[A]s these 4 studies failed on face, the efficacy data were not reviewed in detail. Furthermore, the Division had required that the sponsor submit data from only one positive well-controlled trial for the purpose of establishing efficacy for this NDA [new drug application]'.

#### Results (un)published

Study 5 was neither published as a stand-alone article nor acknowledged in the above-mentioned review article (Stahl, 1993).

#### Summary comparison of FDA v. journal reporting for the five studies

Overall, in the five studies, a total of 950 male and female adult outpatients with panic disorder were treated, 531 patients with alprazolam XR, 70 with alprazolam IR, and 349 with placebo (M-17 and -53/56). As shown in Fig. 1, the FDA review indicated five efficacy trials of alprazolam XR, only one of which (Study 2) was positive, with the remaining four trials 'failed on face'. These were not transparently published: Two – Studies 1 and 3 – were published as positive (Alexander, 1993; Pecknold et al., 1994), contrary to the FDA's conclusion, while the other two – Studies 4 and 5 – were not published.

#### Meta-analyses comparing published v. FDA data

Table 2 shows, for each study, an FDA-based effect size (ES) value for each primary outcome and a composite ES. For each study, the FDA-based composite ES value is shown in Table 2 and in Fig. 2, where it is compared to the journal-based ES. Across the studies, the overall FDA-based composite ES was 0.33 ( $CI_{95\%}$  0.06–0.60) while the journal-based ES was 0.47 ( $CI_{95\%}$  0.29–0.65), representing an increase of 0.14 or 42%.

## Discussion

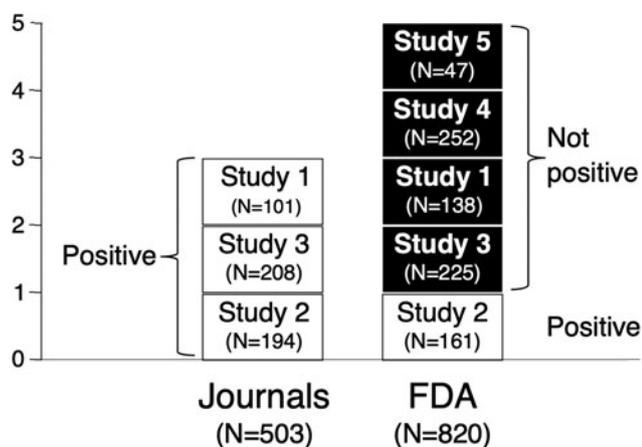
### Key findings

We found that alprazolam XR may be less effective than the published literature would suggest. According to the published literature, every trial of alprazolam XR found it to be effective. By contrast, according to the FDA, only one of five trials was positive. Consequently, the effect size derived from FDA data (0.33) was substantially lower than the effect size derived from the published literature (0.47). The FDA-based effect size was similar to the FDA-based effect size reported previously for antidepressants in the treatment of panic disorder (0.28) (Roest et al., 2015). These findings arguably alter the risk-benefit ratio for the prescribing of this benzodiazepine, especially in the light of recent attention to their contribution to the opioid crisis (Park, Saitz, Ganoczy, Ilgen, & Bohnert, 2015) and the availability of safer alternatives.

### Strengths and Limitations

Prior studies have compared trial data from the FDA v. journal articles for other drug classes (introduction). This is the first study, to our knowledge, to quantify the effect of selective publication on the apparent efficacy of a benzodiazepine, and it arguably provides a more realistic estimate of its efficacy.

A strength of this study is its use of FDA reviews to discover unpublished clinical trial data. Unfortunately, this resource is seldom utilized in Cochrane systematic reviews (Schroll, Bero, & Gøtzsche, 2013). Indeed, two Cochrane reviews on various benzodiazepines (Bighelli et al., 2016; Breilmann et al., 2019) found only one unpublished alprazolam trial (GSK), but here alprazolam served as an active control v. the SSRI paroxetine, for which GSK was seeking FDA approval for panic disorder. (Our group previously reported this trial's paroxetine data using the FDA review for that drug-indication combination (Roest et al., 2015). In that trial, neither drug demonstrated superiority to placebo.) The latter Cochrane review (Breilmann et al., 2019), more relevant because it focused on benzodiazepines v. placebo, missed the two unpublished trials we found (Studies 4 and 5) plus one of the published trials (Study 3) (Alexander, 1993). As the authors acknowledged, they were unable to rule out overestimation of treatment effects



**Figure 1.** Alprazolam XR trial outcomes as presented by journal articles v. FDA. The FDA reviewed five alprazolam XR trials for efficacy and deemed only one of them to be positive. The FDA deemed the other four studies to have ‘failed on face’. Of these four trials, two were published as positive, conflicting with the conclusion of the FDA, while the other two were not transparently published. Sample sizes shown refer to treated patients whose data were analyzed; for Study 2, FDA *N* < journal *N* due to exclusion of one questionable site (see text); patients treated with alprazolam immediate release excluded.

due to sponsorship bias; by contrast, FDA statistical reviewers, with access to patient-level data and the original protocol, conduct independent analyses (Turner, 2004).

This study has several limitations. FDA reviews are limited to premarketing trials, so postmarketing trials are excluded. However, we know of no reason to suspect that drug performance should change after v. before marketing approval; and postmarketing trials can be susceptible to sponsorship-based reporting biases (Heres et al., 2006). This study is restricted to one benzodiazepine for one anxiety-related indication, and it is restricted to issues of efficacy, as opposed to ‘real-world’ effectiveness (Singal, Higgins, & Waljee, 2014). Examination of safety/harm

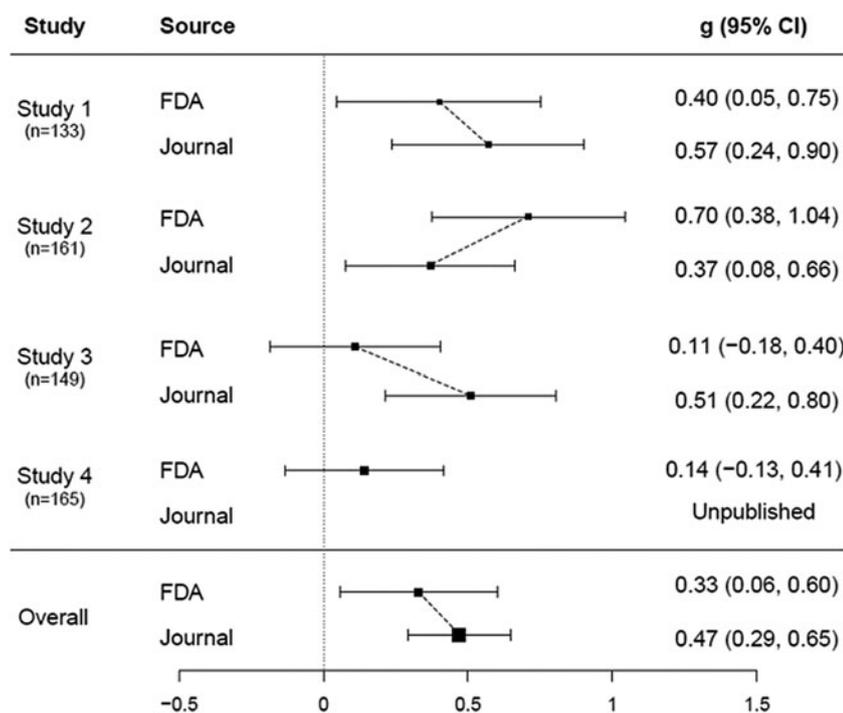
outcomes, which would impact the overall risk-benefit ratio, was beyond the scope of this study. We lacked summary data on one small nonsignificant trial. Lastly, it is possible that trials could have been misclassified as unpublished; however, given our literature search methods, it seems unlikely such publications would be discoverable by most interested clinicians.

**Implication**

This study brings to light unpublished trial data and provides a more balanced and realistic view of the efficacy of alprazolam XR, compared to what has been previously reported. It is unknown whether the discrepancy between FDA and journal trial data is greater or smaller for other benzodiazepines. This adds to the literature on publication bias in clinical trials for drugs for psychiatric conditions, including major depressive disorder (Melander, Ahlqvist-Rastad, Meijer, & Beermann, 2003; Turner et al., 2008), bipolar disorder (Ghaemi, 2009), anxiety disorders (Roest et al., 2015), and schizophrenia (Heres et al., 2006; Turner et al., 2012).

While one might expect that trial registration would prevent the type of reporting bias documented here, ClinicalTrials.gov did not come into existence until the year 2000, many years after all other currently marketed benzodiazepines were approved. If alprazolam XR were approved in the current era, its trials might well have been reported more transparently. Indeed, an increase in the transparent reporting has been found with trials involving, for example, antidepressants (Turner et al., 2022). This is arguably due to policy changes, such as the advent of ClinicalTrials.gov and its later augmentation with required results reporting (Zarin, Tse, Williams, & Carr, 2016), increased awareness of unpublished negative studies (Eyding et al., 2010; Jureidini, McHenry, & Mansfield, 2008; Melander et al., 2003; Turner et al., 2008; Whittington et al., 2004), and recent incentives to increase transparency (Miller, Wilenzick, Ritcey, Ross, & Mello, 2017).

Our findings are consistent with prior studies that have compared clinical trial data in journal publications with those in FDA



**Figure 2.** Forest plots of efficacy of alprazolam XR for panic disorder based on data from FDA v. published literature. The effect sizes of the FDA trials and corresponding published studies are compared in the figure above. FDA-based composite effect sizes (from Table 2) are shown as solid squares and journal-based effect sizes are shown as open squares. Values for effect size are expressed as Hedges’s (the difference between two means divided by their pooled standard deviation). Horizontal lines indicate 95% confidence intervals. The overall effect size based on published trial data was higher than the effect size based on FDA data, with an increase of 0.14 or 42%.

reviews. Selective reporting of clinical trials undermines the integrity of the evidence base and deprives clinicians, patients, researchers, and policymakers of accurate data critical for decision-making. Our study highlights the value of regulatory data to the public health.

**Data.** All data used in this paper – extracted from journal articles and FDA review documents (Drugs@FDA, <https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>) – are publicly accessible.

**Funding statement.** This study was not funded.

**Competing interests.** RA has no competing interests to declare. ET declares that he formerly worked at the FDA as a clinical reviewer and has no other competing interests to declare.

**Ethical standards.** Approval by an ethics board was not required, as only summary data in the public domain were used.

## References

- Ait-Daoud, N., Hamby, A. S., Sharma, S., & Blevins, D. (2018). A review of alprazolam use, misuse, and withdrawal. *Journal of Addiction Medicine, 12*(1), 4–10. doi: 10.1097/adm.0000000000000350
- Alexander, P. (1993). Alprazolam-XR in the treatment of panic disorder: Results of a randomized, double-blind, fixed-dose, placebo-controlled multicenter study. *Psychiatric Annals, 23*(10), 14–18. doi: 10.3928/0048-5713-19931002-06
- Anzures-Cabrera, J., Sarpatwari, A., & Higgins, J. P. (2011). Expressing findings from meta-analyses of continuous outcomes in terms of risks. *Statistics in Medicine, 30*(25), 2967–2985. doi: 10.1002/sim.4298
- Bachhuber, M. A., Hennessy, S., Cunningham, C. O., & Starrels, J. L. (2016). Increasing benzodiazepine prescriptions and overdose mortality in the United States, 1996–2013. *American Journal of Public Health, 106*(4), 686–688. doi: 10.2105/ajph.2016.303061
- Bighelli, I., Trespici, C., Castellazzi, M., Cipriani, A., Furukawa, T. A., Giralda, F., ... Barbui, C. (2016). Antidepressants and benzodiazepines for panic disorder in adults. *Cochrane Database of Systematic Reviews, 9* (9), Cd011567. doi: 10.1002/14651858.CD011567.pub2
- Boon, M., van Dorp, E., Broens, S., & Overdyk, F. (2020). Combining opioids and benzodiazepines: Effects on mortality and severe adverse respiratory events. *Annals of Palliative Medicine, 9*(2), 542–557. doi: 10.21037/apm.2019.12.09
- Boyer, W. (1995). Serotonin uptake inhibitors are superior to imipramine and alprazolam in alleviating panic attacks: A meta-analysis. *International Clinical Psychopharmacology, 10*(1), 45–49. doi: 10.1097/00004850-199503000-00006
- Breilmann, J., Giralda, F., Guaiana, G., Barbui, C., Cipriani, A., Castellazzi, M., ... Koesters, M. (2019). Benzodiazepines versus placebo for panic disorder in adults. *Cochrane Database of Systematic Reviews, 3*(3), Cd010677. doi: 10.1002/14651858.CD010677.pub2
- Committee for Proprietary Medicinal Products. (2001). Points to consider on missing data (CPMP/EWP/1776/99). *The European agency for the evaluation of medicinal products*, p. 3. Retrieved from [https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-missing-data\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-missing-data_en.pdf)
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation, p. 213.
- Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., & Smith, Jr H. (1987). Publication bias and clinical trials. *Controlled Clinical Trials, 8*(4), 343–353. doi: 10.1016/0197-2456(87)90155-3
- Eyding, D., Lelgemann, M., Grouven, U., Härter, M., Kromp, M., Kaiser, T., ... Wieseler, B. (2010). Reboxetine for acute treatment of major depression: Systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *British Medical Journal, 341*, c4737. doi: 10.1136/bmj.c4737
- FDA (2022). Multiple endpoints in clinical trials guidance for industry (FDA-2016-D-4460). Center for biologics evaluation and research and center for drug evaluation and Research. Retrieved from <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry>
- Ghaemi, S. N. (2009). The failure to know what isn't known: Negative publication bias with lamotrigine and a glimpse inside peer review. *Evidence-based Mental Health, 12*(3), 65–68. doi: 10.1136/ebmh.12.3.65
- GSK. A Double-Blind, Multicentered, Flexible-Dose Study of Paroxetine, Alprazolam and Placebo in the Treatment of Panic Disorder. GSK Study ID: 29060/223. Retrieved from <https://www.gsk-studyregister.com/en/trial-details/?id=29060/223>
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*(2), 490–499. doi: 10.1037/0033-2909.92.2.490
- Heres, S., Davis, J., Maino, K., Jetzinger, E., Kissling, W., & Leucht, S. (2006). Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: An exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry, 163*(2), 185–194. doi: 10.1176/appi.ajp.163.2.185
- Jureidini, J., McHenry, L., & Mansfield, P. (2008). Clinical trials and drug promotion: Selective reporting of study 329. *International Journal of Risk & Safety in Medicine, 20*, 73–81. doi: 10.3233/JRS-2008-0426
- Leon, A. C., Mallinckrodt, C. H., Chuang-Stein, C., Archibald, D. G., Archer, G. E., & Chartier, K. (2006). Attrition in randomized controlled clinical trials: Methodological issues in psychopharmacology. *Biological Psychiatry, 59*(11), 1001–1005. doi: 10.1016/j.biopsych.2005.10.020
- López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. T. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods, 9*(3), 336–351. doi: 10.1002/jrsm.1310
- Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. (2003). Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. *British Medical Journal, 326*(7400), 1171–1173. doi: 10.1136/bmj.326.7400.1171
- Miller, J. E., Wilenzick, M., Ritcey, N., Ross, J. S., & Mello, M. M. (2017). Measuring clinical trial transparency: An empirical analysis of newly approved drugs and large pharmaceutical companies. *BMJ open, 7*(12), e017917. doi: 10.1136/bmjopen-2017-017917
- Nardi, A. E., & Quagliato, L. A. (2022). Benzodiazepines are efficacious and safe for long-term use: Clinical research data and more than sixty years in the market. *Psychotherapy and Psychosomatics, 91*(5), 300–303. doi: 10.1159/000524730
- National Institute on Drug Abuse. (2022). *Benzodiazepines and opioids*. Gaithersburg, Maryland USA: National Institute on Drug Abuse. Retrieved from <https://www.drugabuse.gov/drugs-abuse/opioids/benzodiazepines-opioids>
- Park, T. W., Saitz, R., Ganoczy, D., Ilgen, M. A., & Bohnert, A. S. (2015). Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics: Case-cohort study. *British Medical Journal, 350*, h2698. doi: 10.1136/bmj.h2698
- Pecknold, J., Luthe, L., Munjack, D., & Alexander, P. (1994). A double-blind, placebo-controlled, multicenter study with alprazolam and extended-release alprazolam in the treatment of panic disorder. *Journal of Clinical Psychopharmacology, 14*(5), 314–321.
- Perugi, G., Frare, F., & Toni, C. (2007). Diagnosis and treatment of agoraphobia with panic disorder. *CNS Drugs, 21*(9), 741–764. doi: 10.2165/00023210-200721090-00004
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Roest, A. M., de Jonge, P., Williams, C. D., de Vries, Y. A., Schoevers, R. A., & Turner, E. H. (2015). Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: A report of 2 meta-analyses. *JAMA psychiatry, 72*(5), 500–510. doi: 10.1001/jamapsychiatry.2015.15
- Rosenthal, R. (1991). *Meta-Analytic procedures for social research*. Newbury Park, California: Sage.
- Schroll, J. B., Bero, L., & Gøtzsche, P. C. (2013). Searching for unpublished data for Cochrane reviews: Cross sectional study. *British Medical Journal, 346*, f2231. doi: 10.1136/bmj.f2231
- Schweizer, E., Patterson, W., Rickels, K., & Rosenthal, M. (1993). Double-blind, placebo-controlled study of a once-a-day, sustained-release

- preparation of alprazolam for the treatment of panic disorder. *American Journal of Psychiatry*, 150(8), 1210–1215. doi: 10.1176/ajp.150.8.1210
- Silberman, E., Balon, R., Starcevic, V., Shader, R., Cosci, F., Fava, G. A., ... Sonino, N. (2021). Benzodiazepines: It's time to return to the evidence. *British Journal of Psychiatry*, 218(3), 125–127. doi: 10.1192/bjp.2020.164
- Singal, A. G., Higgins, P. D., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, 5(1), e45. doi: 10.1038/ctg.2013.13
- Stahl, S. M. (1993). Alprazolam-XR: Dosage considerations. *Psychiatric Annals*, 23(10), 27–31. doi: 10.3928/0048-5713-19931002-08
- StataCorp LP. (2009). *Stata statistical software: Release 11*. College Station, TX. Retrieved from <https://www.stata.com/support/faqs/resources/citing-softwaredocumentation-faqs/>
- Stein, M., Pollack, M., Roy-Byrne, P., Sareen, J., Simon, N., & Campbell-Sills, L. (2010). Practice guideline for the treatment of patients with panic disorder. *American Psychiatric Association*.
- Suñé, P., Suñé, J. M., & Montoro, J. B. (2013). Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One*, 8(1), e54583. doi: 10.1371/journal.pone.0054583
- Thorlund, K., Walter, S. D., Johnston, B. C., Furukawa, T. A., & Guyatt, G. H. (2011). Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. *Research Synthesis Methods*, 2(3), 188–203. doi: 10.1002/jrsm.46
- Turner, E. H. (2004). A taxpayer-funded clinical trials registry and results database. *PLoS medicine*, 1(3), e60. doi: 10.1371/journal.pmed.0010060
- Turner, E. H. (2013). How to access and process FDA drug approval packages for use in research. *British Medical Journal*, 347, f5992. doi: 10.1136/bmj.f5992
- Turner, E. H., Cipriani, A., Furukawa, T. A., Salanti, G., & de Vries, Y. A. (2022). Selective publication of antidepressant trials and its influence on apparent efficacy: Updated comparisons and meta-analyses of newer versus older trials. *PLoS medicine*, 19(1), e1003886. doi: 10.1371/journal.pmed.1003886
- Turner, E. H., Knoepflmacher, D., & Shapley, L. (2012). Publication bias in antipsychotic trials: An analysis of efficacy comparing the published literature to the US Food and Drug Administration database. *PLoS Medicine*, 9(3), e1001189. doi: 10.1371/journal.pmed.1001189
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3), 252–260. doi: 10.1056/NEJMsa065779
- Whittington, C. J., Kendall, T., Fonagy, P., Cottrell, D., Cotgrove, A., & Boddington, E. (2004). Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. *Lancet (London, England)*, 363(9418), 1341–1345. doi: 10.1016/s0140-6736(04)16043-1
- Wilkinson, G., Balestrieri, M., Ruggeri, M., & Bellantuono, C. (1991). Meta-analysis of double-blind placebo-controlled trials of antidepressants and benzodiazepines for patients with panic disorders. *Psychological Medicine*, 21(4), 991–998. doi: 10.1017/s0033291700029986
- Wilson, D. (2017). Formulas Used by the “Practical Meta-analysis Effect Size Calculator”. Retrieved from <https://mason.gmu.edu/~dwilsonb/downloads/esformulas.pdf>
- Zarin, D. A., Tse, T., Williams, R. J., & Carr, S. (2016). Trial reporting in ClinicalTrials.gov – The final rule. *New England Journal of Medicine*, 375(20), 1998–2004. doi: 10.1056/NEJMs1611785