Chapter 3

Bias in the reporting of harms in clinical trials of second-generation antidepressants for depression and anxiety

Ymkje Anna de Vries, Annelieke M. Roest, Lian Beijers,
Erick H. Turner, Peter de Jonge

European Neuropsychopharmacology (2016), 26, 1752 - 1759

Abstract

Background: Previous research has shown that reporting bias has inflated the apparent efficacy of antidepressants. We investigated whether apparent safety was also affected.

Methods: We included 133 trials, involving 31,296 patients, of second-generation antidepressants for the treatment of major depressive disorder (MDD) or anxiety disorders, obtained from Food and Drug Administration (FDA) reviews. We extracted data on overall discontinuation, discontinuation due to adverse events, and serious adverse events (SAEs). Meta-analysis was used to compare discontinuation rates between FDA reviews and matching journal articles, while SAEs were compared qualitatively.

Results: The odds ratio for overall discontinuation, comparing drug to placebo, was 1.0 for both sources, while that for discontinuation due to adverse events was 2.4 for both sources. Seventy-seven of 97 (79%) journal articles provided incomplete information on SAEs; sixty-one (63%) articles made no mention of SAEs at all. Of 21 articles which could be compared to the FDA, only 6 (29%) had full reporting without discrepancies. Nine (43%) articles reported a discrepant number of SAEs. Descriptions were absent or discrepant in 6 (29%) additional articles, even for important SAEs such as suicide attempts.

Conclusions: Reporting bias has not affected average discontinuation rates over trials. However, SAE reporting is not only very poor, with over half of articles failing to discuss SAEs altogether, but discrepancies between the FDA and articles were common and often led to a more favorable drug-placebo comparison. These findings suggest that journal articles are an unreliable source of data on SAEs in antidepressant trials.

Introduction

A significant fraction of all studies are never published in peer-reviewed journals [38]. Even within the subset of studies that are published, the (primary) analyses and outcomes reported in journal articles frequently deviate from the protocol [103, 186]. As a consequence, statistically significant (positive) studies or outcomes are more likely to be published than non-significant (negative) studies [38] or outcomes [103].

While it is often difficult to assess the presence of reporting bias, the United States Food and Drug Administration (FDA) maintains an independent results database for drug trials, which can be used to examine the presence of reporting bias within a set of trials [105]. This database has previously been used to assess reporting bias in trials of antipsychotics for schizophrenia [104] and antidepressants for major depressive disorder (MDD) [19] and anxiety disorders [20].

Second-generation antidepressants have been found to be effective for MDD [19] and anxiety disorders [113, 116, 119, 122, 123, 187]. They are considered to have a favorable risk-benefit profile and hence are widely prescribed [109]. While both studies examining the FDA database of antidepressant trials confirmed their efficacy for MDD and anxiety disorders, they also revealed substantial reporting bias [19, 20]. Although nearly all published trials (94 – 96%) reported positive results, only 51% of all submitted trials for MDD, and 72% of those for anxiety disorders, were judged to be positive by the FDA. As a consequence of reporting bias, the effect size of antidepressant treatment was overestimated by 32% and 15% for MDD and anxiety disorders, respectively.

An accurate assessment of the risk-benefit ratio of antidepressants requires an unbiased understanding of safety as well as efficacy, but this other side of the coin has not, thus far, been examined as comprehensively. Previous research has indicated that reporting of harms in journal articles is incomplete and inadequate in various medical fields [54, 55, 56], including psychiatry [53].

The case of reboxetine demonstrates the impact that reporting bias can have on apparent safety as well as efficacy: inclusion of unpublished data not only shifted the difference in efficacy between reboxetine and placebo from significant to non-significant, but it also showed that reboxetine was significantly inferior to placebo in terms of selected harm outcomes, while the published trials suggested they were equivalent [101]. Poor reporting of harms has also been found in trials of two other antidepressants (sertraline and duloxetine) and several antipsychotics, with serious adverse events (SAEs) not always reported fully or accurately in journal articles [57, 58].

The work on antidepressant trials was limited to relatively recent trials of three antidepressants, and only the reboxetine study quantified the possible impact of bias on an important harm outcome, discontinuation from the trial. In the present study, we assessed the presence of reporting bias, and its impact on several harm outcomes, within

a comprehensive set of trials of second-generation antidepressants for both MDD and anxiety disorders.

Methods

Data from FDA reviews and journal articles

We previously obtained FDA reviews of second-generation antidepressants approved for MDD [19] and/or anxiety disorders [20] (specifically generalized anxiety disorder (GAD), social anxiety disorder (SAD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD) and panic disorder (PD)). We defined second-generation antidepressants as including selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), as well as other antidepressants (specifically mirtazapine, bupropion, and nefazodone) approved between 1987 and 2008.

From these reviews, we identified all phase 2/3 short-term clinical trials registered with the FDA and conducted in pursuit of marketing approval. For MDD, we identified 74 trials of 12 drugs; for GAD, 11 trials of 4 drugs; for SAD, 11 trials of 5 drugs; for OCD, 13 trials of 5 drugs; for PTSD, 7 trials for 2 drugs; and for PD, 17 trials for 5 drugs. Two of the PD trials were not included in our previous analysis [20], as we did not receive the FDA review containing these trials in time. Hence, we included a total of 133 trials, consisting of data from 31,296 participants, of whom 18,904 were treated with antidepressants and 12,392 with placebo.

We conducted an extensive search of the published literature to identify journal articles corresponding to these FDA-registered trials, as described previously [19, 20]. A total of 97 publications were identified, covering 102 (77%) of 133 trials: 51 for MDD (including 1 publication covering 2 trials), 9 for GAD (1 publication covering 2 trials), 11 for SAD, 9 for OCD (1 publication covering 2 trials), 5 for PTSD, and 12 for PD (1 publication covering 3 trials).

For each trial, we extracted the following data from FDA reviews and corresponding journal articles, separately for each treatment group: sample size, number and proportion of patients discontinuing, number and proportion of patients discontinuing due to adverse events specifically, and the number and nature of serious adverse events (SAEs). SAEs are defined as any adverse event that results in death, hospitalization, disability or permanent damage, a birth defect, or any other life-threatening situation. Individual trial protocols may, however, define additional adverse events as serious adverse events. Both SAEs occurring during the administration of a drug and those occurring within a specified period (usually 30 days) after the last dose (post-therapy SAEs) must be reported to the FDA.

For trials which subdivided discontinuation due to adverse events into subcategories, we counted all participants who discontinued due to side effects, laboratory abnormalities, test findings, suicide attempts, suicide, and other causes of death, whether considered drug-related or not. When the journal article stated that no SAEs attributable to the drug were observed, we counted the article as reporting zero SAEs for the drug group.

We extracted data from (in order of preference) the safety population (all randomized patients who took at least one dose of medication or placebo), the randomized population, or the intention-to-treat efficacy population (all patients who took at least one dose of medication or placebo and who had at least one post-baseline efficacy evaluation). Extraction was performed independently by two investigators (YV and AR for depression and YV and LB for anxiety), with disagreements resolved by consensus. All remaining discrepancies between journal articles and FDA reviews were double-checked for possible errors.

Statistical analysis – discontinuation rates

For fixed-dose (multiple dose) studies, we calculated a combined sample size, and a combined number of patients discontinuing, for the various antidepressant arms. We then calculated the weighted mean discontinuation rate (overall [i.e., for any reason] and due to adverse events) for the drug and placebo group over all trials, as well as specifically per disorder.

We conducted four random-effects meta-analyses, as we had two data sources (FDA and journal articles) and two discontinuation rates (overall and due to adverse events). Restricted maximum likelihood (REML) was used as the estimation method. In case of empty cells (zero patients discontinuing in a group), 0.5 was added to the empty cell. Meta-regression was used to further investigate the impact of data source (FDA versus journal). We also conducted subgroup analyses for the different disorders. The metafor package (version 1.9-5) in R (version 3.1.3) was used for these analyses.

Analyses of serious adverse events

We first examined the availability of SAE data for the journal articles and the FDA reviews. For the journal articles that mentioned SAEs, and for which a comparison with the FDA review could be made, we determined whether there was a discrepancy in the number or description of SAEs in either the placebo or the drug group. In case of discrepancies, we further examined the nature of the discrepancy and whether it would lead to a more favorable drug-placebo comparison.

Results

Discontinuation rates

The average discontinuation rate (weighted by sample size) in the placebo group was 30.5% according to the FDA versus 30.1% according to the journal articles. For the antidepressant groups, the average discontinuation rate was 31.9% according to the FDA and 30.4% according to the journal articles (Table 3.1).

Random-effects meta-analysis of the FDA data yielded an odds ratio (OR) of discontinuation for the drug group compared with placebo of 1.02 (95% CI: 0.96 – 1.08), indicating no significant difference in odds of overall discontinuation between the drug and placebo groups. Similarly, random-effects meta-analysis of the journal data gave an OR of 0.98 (95% CI: 0.92 – 1.06). Lack of significant bias in journal articles compared to FDA was confirmed by meta-regression (p = 0.47) (Table 3.2).

For discontinuation due to adverse events, the average discontinuation rate was 5.2% for placebo according to the FDA and 5.0% according to the journal articles, while it was 12.6% for antidepressants according to the FDA and 12.3% according to journal articles. Random-effects meta-analysis of the FDA data yielded an odds ratio of 2.39 (95% CI: 2.11 – 2.70), indicating a significantly higher risk of discontinuation due to adverse events in the drug group. The journal articles gave similar results, with an OR of 2.42 (95% CI: 2.11 – 2.78), which was not significantly different from the FDA result (p = 0.94) (Table 3.2).

Table 3.1: *Discontinuation rates*

	Overall discontinuation (%)				Discontinuation due to AEs (%)			
	FDA		Journal		FDA		Journal	
	Placebo	Drug	Placebo	Drug	Placebo	Drug	Placebo	Drug
MDD	35.0	34.5	34.5	32.7	5.6	13.1	5.0	12.7
GAD	24.2	28.6	25.5	29.8	4.4	13.7	4.9	14.1
SAD	30.7	33.3	31.2	32.8	4.0	14.5	3.4	14.8
OCD	24.5	28.3	25.6	25.1	5.4	12.0	5.6	11.3
PTSD	32.1	32.5	31.3	34.3	7.4	11.5	6.9	12.0
PD	25.5	26.6	24.8	25.1	5.4	8.9	5.4	8.5
Total	30.5	31.9	30.1	30.4	5.2	12.6	5.0	12.3

Sample-size weighted average rates of overall discontinuation and discontinuation due to adverse events (AEs) for the placebo and drug groups according the Food and Drug Administration (FDA) and journal articles. GAD: generalized anxiety disorder; MDD: major depressive disorder; OCD: obsessive-compulsive disorder; PD: panic disorder; PTSD: post-traumatic stress disorder; SAD: social anxiety disorder.

Table 3.2: *Odds ratio of discontinuation*

	Overall discontinuation		Discontinuation due to AEs	
	FDA	Journal	FDA	Journal
MDD	0.92 (0.85-0.99)	0.87 (0.79-0.96)	2.29 (1.93-2.73)	2.42 (1.98-2.96)
GAD	1.25 (1.07-1.46)	1.26 (1.07-1.50)	3.06 (2.32-4.03)	2.85 (2.15-3.76)
SAD	1.14 (0.91-1.44)	1.08 (0.87-1.31)	3.75 (2.47-5.70)	4.33 (2.97-6.31)
OCD	1.24 (0.96-1.60)	0.99 (0.77-1.27)	2.28 (1.62-3.19)	2.23 (1.55-3.19)
PTSD	0.99 (0.82-1.21)	1.09 (0.87-1.36)	1.67 (1.11-2.51)	1.72 (1.18-2.52)
PD	1.06 (0.88-1.26)	0.99 (0.80-1.22)	1.57 (1.18-2.14)	1.51 (1.04-2.17)
Total	1.02 (0.96-1.08)	0.98 (0.92-1.06)	2.39 (2.11-2.70)	2.42 (2.11-2.78)

Odds ratio (95% confidence interval) of overall discontinuation and discontinuation due to adverse events (AEs) for the drug group compared to the placebo group according the FDA and journal articles. GAD: generalized anxiety disorder; MDD: major depressive disorder; OCD: obsessive-compulsive disorder; PD: panic disorder; PTSD: post-traumatic stress disorder; SAD: social anxiety disorder.

When we examined the disorders separately, no bias was apparent for any of the included disorders (all p-values >0.23). For GAD, overall discontinuation was significantly higher for the drug group than for the placebo group (OR = 1.25, 95% CI: 1.07 – 1.46, p = 0.004), while it was significantly lower for the drug group for MDD (OR = 0.92, 95% CI: 0.85 – 0.99, p = 0.026).

Discontinuation due to adverse events was significantly higher in the drug group for all disorders, with the magnitude of the difference ranging from an OR of 1.57 (for PD) to 3.75 (for SAD) on the basis of FDA data (Table 3.2).

Serious adverse events

Data on serious adverse events was frequently missing from both FDA reviews and journal articles. Out of 133 trials, SAE data was missing in the FDA review for 57 trials (43%). Nearly all of these were older trials for MDD, where data for 56 out of 74 (76%) trials was missing. For an additional 24 trials (18%), trial-level data on SAEs was not available in the FDA review, as the data was pooled over all pivotal trials included in the review.

Of 97 journal articles (covering 102 trials), only 36 (37%) mentioned SAEs at all: complete information was provided in 20 (21%) articles, while 16 (16%) articles had incomplete reporting (e.g. mentioning the number of SAEs without providing descriptions, or giving information for the drug group only).

For 15 of the 36 articles, there was insufficient information in the FDA review to perform a direct comparison. Of the remaining 21 articles, for which study-level SAE data was provided in the FDA reviews, only 6 (29%) articles had full reporting with no discrepancy.

Table 3.3: *Discrepancies in number of reported SAEs*

Disorder	Trial	FDA reporting	Journal reporting
MDD	84023 [188]	2 suicides in mirtazapine group vs. none in placebo group	"No clinically important SAEs ... attributable to mirtazapine were seen"
GAD	HMDT [135]	Severe anxiety and post-therapy death due to asphyxiation (choking)	Anxiety only
SAD	387 [158]	Event listed as "other event of clinical interest"	Event listed as SAE
	393 [159]	Event listed as "other event of clinical interest"	Event listed as SAE
	454 [153]	1 SAE in paroxetine group (brain edema due to car accident)	"No SAEs attributable to paroxetine treatment"
	94-004 [156]	Pregnancy with post-treatment abortion	Not mentioned
PD	399 [146]	1 SAE in placebo group and 5 SAEs in venlafaxine groups	8 SAEs in placebo group and 5 SAEs in venlafaxine groups
	391 [148]	4 SAEs in placebo group and 6 SAEs in drug group	Additional SAE of anxiety in placebo group
	495 [142]	Seizure	Not mentioned

FDA: Food and Drug Administration; GAD: generalized anxiety disorder; MDD: major depressive disorder; PD: panic disorder; SAE: serious adverse event.

In 9 out of 21 (43%) articles there were discrepancies in the reported number of SAEs, which led to a smaller or reversed drug-placebo difference and thus a more favorable drug-placebo comparison in 7 cases (Table 3.3).

Two articles reported additional SAEs in the drug group, which had been classified as "other clinical events of interest" by the FDA. Additional SAEs in the placebo group, which were not noted in the FDA review, were reported by two articles. Two other articles noted that "no SAEs attributable to [drug]" were seen, even though SAEs did occur in the drug group according to the FDA, including two suicides. Post-therapy SAEs were omitted in two other articles, while one article omitted an SAE of seizure in the drug group that had been noted by the FDA.

In 6 of 21 (29%) articles, the reported numbers of SAEs agreed, but a description of the SAEs was either (partly) missing in the journal article (5 articles) or differed from the FDA description of the SAE (1 article) (Table 3.4). The discrepancy in description involved an SAE that was described only as "emotional lability" in the journal article,

Table 3.4: *Missing or discrepant SAE descriptions*

Disorder	Trial	FDA reporting	Journal reporting
MDD	448, 449 [189]	5 SAEs in placebo group (4 somatic SAEs; accidental overdose) and 15 SAEs in drug groups (5 somatic SAEs; 3 cases of unintended pregnancy; abortion; post-therapy depression; 3 cases of emotional lability; depression and emotional lability; manic reaction)	No description
SAD	3108 [150]	1 SAE in placebo group (nasal septum disorder)	No description
	502 [151]	"Emotional lability/intentional OD of paracetamol and aspirin"	"Emotional lability"
OCD	3103 [169]	2 SAEs in placebo group (neoplasm, unintended pregnancy) and 5 in drug group (suicide attempt, accidental injury, asthma, hostility, depression)	No description
PTSD	651 [163]	9 SAEs in paroxetine group (headache; accidental overdose; benign neoplasm; uterine neoplasm; unintended pregnancy; 2 cases of emotional lability; manic reaction)	9 SAEs noted, but only those thought possibly related to treatment specified (headache, accidental overdose)
	671 [162]	1 SAE in placebo group (hives) and 1 SAE in drug group (post-treatment suicide attempt)	No description

FDA: Food and Drug administration; MDD: major depressive disorder; OCD: obsessive-compulsive disorder; OD: overdose. PTSD: post-traumatic stress disorder; SAD: social anxiety disorder; SAE: serious adverse event.

while it was described as "emotional lability/intentional OD [overdose] of paracetamol & aspirin" by the FDA.

With regard to missing descriptions, one article specified only those SAEs thought to be related to drug, while four articles did not provide descriptions for any SAEs. In four of these five articles with missing descriptions, the corresponding FDA reviews showed that there was a preponderance of psychiatric SAEs specifically in the drug groups, including suicide attempt, hostility, depression, emotional lability, and manic reaction, while there were almost no such SAEs in the placebo group.

Discussion

Principal findings

We found no evidence for bias in the reporting of overall discontinuation rates or discontinuation rates due to adverse events. However, discontinuation rates were high, averaging approximately 30% in both placebo and drug groups, especially given the short (6 – 12 weeks) duration of most of the included trials. Furthermore, participants receiving an antidepressant were 2.4 times more likely to discontinue the trial due to adverse events than participants receiving placebo.

Reporting of SAEs was very poor. In 79% of all journal articles, SAE data was incomplete or missing. Almost two-thirds failed to mention SAEs entirely, and an additional 16% of articles provided incomplete information regarding SAEs. For instance, some articles provided only the number of SAEs, without any description. Given the idiosyncratic nature of SAEs, such numbers have little meaning. Other articles provided the number of SAEs for the drug group only. However, lacking information about the base rate of SAEs in the placebo group, readers cannot ascertain whether the rate (or nature) of SAEs in the drug group should be cause for concern. In some of these cases, the FDA review showed that the number of SAEs in the placebo group was zero, but this was not always the case and hence cannot be assumed.

Furthermore, where a direct comparison with the FDA data was possible, discrepancies were frequent and usually led to a more favorable comparison between drug and placebo in the journal article. Discrepancies in the number of reported SAEs most commonly originated from the omission of post-therapy SAEs or the omission of events that were judged to be unrelated to treatment. However, such judgments are made subjectively by site investigators. As they are blind to treatment assignment, and the safety profile of a new drug cannot be known until the entire clinical trials program has been completed, it is impossible for investigators to determine causality.

As evidence that such judgments lead to undue omission of SAEs, in the current study, one article concluded that "no clinically important serious adverse events or side effects attributable to mirtazapine were seen" [188], even though, according to the FDA review, 2 (of 59) patients treated with mirtazapine died by suicide. Other articles neglected to provide a description for some or all SAEs, while the FDA data revealed a greater number of psychiatric adverse events specifically in the drug group. Such psychiatric adverse events are of particular concern, since antidepressants have been most controversially associated with suicidality (especially in children and young adults) [190, 191, 192] and aggression or violence [192, 193, 194, 195].

Because SAEs are infrequent, it is often difficult to reliably determine whether they are due to a drug on the basis of an entire clinical trials program, much less a single

antidepressant trial. Given the small numbers, an excess of SAEs in the drug group may be a chance finding, so many authors may choose to omit these data from journal articles to avoid alarming clinicians. Although meta-analysis of a large set of trials can help compensate for the small sample size of individual trials, poor reporting hampers meta-analysis of harm outcomes by independent authors. Outcome reporting bias is said to be a major threat to the validity of systematic reviews of harm outcomes [196]. Space limitations imposed by journals might also be a reason to omit SAEs entirely or SAE descriptions specifically, particularly when combined with a generally greater interest in efficacy results than in safety results on the part of editors and readers.

Comparison with previous literature

Several studies have found that, compared to ClinicalTrials.gov, SAE reporting in corresponding journal articles is incomplete [197, 198, 199]. Discrepancies are common, and journal articles generally report fewer SAEs. Similar to our results, two of these studies found that one reason for discrepancies was that journal articles reported only SAEs they judged to be drug-related [197, 198].

Regarding antidepressants specifically, Maund and colleagues examined the reporting of harms and benefits in 9 trials of duloxetine for MDD (eight of which were also included in the current study) [57]. Consistent with our results for duloxetine for MDD specifically, they found no discrepancies in the reporting of discontinuations because of adverse events or SAEs, although SAE reporting was very incomplete.

Hughes and colleagues, examining trials of several psychotropic medications (including duloxetine and sertraline) also found that nearly half of all SAEs reported in trial summaries were not reported in the associated journal articles, including over half of all cases of death and suicide [58]. Where SAEs were reported, discrepancies were common. Among the reasons for discrepancies were not reporting SAEs occurring during follow-up and only reporting "drug-related" SAEs, both of which we also found.

"Emotional lability" was one of the most frequently encountered SAE terms in this study. In clinical trials, narrative descriptions of adverse events are coded to a preferred term by use of a medical dictionary [200], but it is often unclear what this term actually means. A recent re-analysis of patient-level data from a trial of paroxetine for MDD in children and adolescents showed that this term was used to code for suicidal ideation, self-harm, or suicide attempts [33]. Consistent with this, we observed a case in which the journal article mentioned only emotional lability, while the FDA additionally reported intentional overdose. Hughes and colleagues also found cases in which, compared to trial summaries, journal articles provided markedly less informative SAE descriptions (e.g. "worsening of the illness" compared to "suicidal ideation") [58]. Hence, using vague descriptions may be an additional way in which SAE reporting can be biased.

Strengths and limitations

Among the strengths of our study is our use of the independent FDA database, which permitted us to identify a complete cohort of pre-marketing trials. As a consequence, and in contrast to previous research, we were able to include a wider range of antidepressants. We also examined important harm outcomes beyond SAEs, providing a more complete overview of reporting on harms.

Our study was limited by the information missing from FDA reviews, particularly with regard to SAEs, which hampered our ability to perform direct comparisons between journal articles and the FDA. This was especially true for older trials of antidepressants for MDD. Although SAEs were almost certainly examined during the drug approval process, this information was missing from the drug application packages provided to us. Therefore, although some information might still be found on clinical trial registries, data on SAEs is unavailable to the public for a significant fraction of all trials performed to obtain marketing approval for second-generation antidepressants, which are currently prescribed to approximately 10% of the US population [109].

Another limitation is that we did not examine common adverse events. Previous work has found that many common adverse events are not reported in journal articles, as these often only report those above a certain frequency threshold (e.g. >10% incidence in the drug group) [57].

Furthermore, we examined specifically whether the same information could be extracted from journal articles as from FDA reviews, without examining more subtle biases. Previous studies, for example, have shown that much less space is devoted to reporting on harms than is devoted to reporting on efficacy [53, 54]. In this study, we found that reporting of the actual discontinuation rates was unbiased, but many journal articles conclude, in their abstract, that the antidepressant was "safe", "well-tolerated", or both, even though antidepressant-treated participants were, on average, 2.4 times more likely to discontinue due to adverse events than placebo-treated patients. This might be considered a form of spin, particularly if the abstract makes no mention of the actual discontinuation rates or occurrence of adverse events.

Conclusions

While reporting of discontinuation rates showed no bias, reporting of SAEs was very poor, and inconsistencies between journal articles and FDA reviews were common. Previous research has shown that adoption of the CONSORT checklist leads to better reporting of the design, participant flow and results for efficacy outcomes in a trial [201]; compliance with the CONSORT extension for harms should similarly improve reporting of harm outcomes [202]. However, information on harms in previously completed trials can only become accessible if pharmaceutical companies or investigators choose to release it. Our

results thus show that an accurate assessment of the risk-benefit ratio of many widely prescribed antidepressants is hampered by poor and biased reporting of SAEs.