Journal of Clinical Epidemiology 106 (2019) 80–87

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Extremely large outlier treatment effects may be a footprint of bias in trials from less developed countries: randomized trials of gabapentinoids

Karishma Desai[a], Ian Carroll[b], Steven Asch[c,d], Tina Hernandez-Boussard[a,e,f], John P.A. Ioannidis[d,e,g,h,i,*]

[a]*Department of Medicine (Biomedical Informatics Research Center), Stanford University, Stanford, California 94305, USA*
[b]*Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University, Stanford, California 94305, USA*
[c]*Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, California 94025, USA*
[d]*Department of Medicine, Stanford University, Stanford, California 94305, USA*
[e]*Department of Biomedical Data Science, Stanford University, Stanford, California 94305, USA*
[f]*Department of Surgery, Stanford University, Stanford, California 94305, USA*
[g]*Department of Health Research and Policy, Stanford University, Stanford, California 94305, USA*
[h]*Department of Statistics, Stanford University, Stanford, California 94305, USA*
[i]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California 94305, USA*

Accepted 14 October 2018; Published online 23 October 2018

Abstract

Objectives: Court documents have proven that a manufacturer-orchestrated strategy tried to promote gabapentin by distorting evidence in randomized trials. Given this background, we aimed to assess whether implausibly large treatment effects for gabapentin and for a similar gabapentinoid, pregabalin may have been published.

Study Design and Setting: We identified meta-analyses on gabapentin or pregabalin on any outcome from Google Scholar, PubMed, and EMBASE. We explored excess of significance in meta-analyses and whether outlier studies with extreme results (differing >0.8 standard deviations from the summary effect of the meta-analysis) were scrutinized.

Results: All 10 evaluated meta-analyses showed statistically significant favorable findings. Heterogeneity I^2 estimates exceeding 90% were noted in four meta-analyses of postoperative pain. In these four meta-analyses, 77 studies had estimates differing >0.8 standard deviations from the summary estimate. Thirty-nine of 77 represented extremely favorable results, and 33 of them came from less developed countries with no tradition of clinical research, 22 reported no information on funding, and 20 reported no conflicts of interest. Conversely, 27 of 38 studies with unfavorable results came from more developed countries.

Conclusion: Extremely favorable outlier studies in the meta-analyzed literature of gabapentin and pregabalin may be a footprint of bias in studies done in less developed countries.

Keywords: Bias; Gabapentin; Pregabalin; Meta-analysis; Excess significance; Outlier; Less developed countries; Industry; Funding

1. Introduction

In 1993, Neurontin was originally approved by the US Food and Drug Administration (FDA) for adults for partial seizures. However, the drug was increasingly used for off-label treatment of pain and psychiatric conditions. In 1996, a lawsuit was filed for illegal marketing of gabapentin for off-label use [1]. In 2004, Warner-Lambert, now a subsidiary of Pfizer, pleaded guilty to these charges and paid about $430 million dollars for criminal and civil liability [2].

Internal company documents released as part of the litigation [3] show that Pfizer and Parke-Davis executed a strategy to conduct trials and disseminate trial findings for off-label uses, influence publication content, select findings for presentation/publication, and have their staff author articles for others without attribution. Using these internal documents, an evaluation of 11 published gabapentin trials [4,5] demonstrated extensive manipulation and selective reporting of outcomes. Other authors have dissected

Conflicts of interest: None.

* Corresponding author. 1265 Welch Road, Medical School Office Building, Room X306, Stanford, CA 94305, USA. Tel.: 650 704-5584; fax: 650 725-6247.

E-mail address: jioannid@stanford.edu (J.P.A. Ioannidis).

https://doi.org/10.1016/j.jclinepi.2018.10.012

What is new?

Key findings

- We found a large number of published studies with implausible, extremely favorable results for gabapentin and pregabalin.
- Most of these studies had been done in less developed countries and reported no information on funding or conflicts of interest.
- Most of the studies with unfavorable results came from more developed countries.

What this adds to what was known?

- Our analysis suggests strong bias in the global literature on a topic with an already well-documented manufacturer-orchestrated strategy that distorted evidence in randomized trials.

What is the implication and what should change now?

- Trials with extreme, implausible results should be scrutinized and raise suspicion for the integrity of the evidence included in a meta-analysis.

gabapentin seeding trials, that is, trials run for marketing rather than real clinical purposes [6].

Since gabapentin was released in 1993, many trials have assessed the utility of gabapentin for diverse off-label indications; perhaps, most commonly for the reduction of postoperative pain where many dozens of relatively small studies have been published. This evidence has been assessed by several meta-analyses that try to summarize these findings [7–11]. Even when the meta-analysts have no conflicts, they still synthesize primary data that may have been generated by a biased, skewed process that causes inflated summary results. One may be particularly worried about studies with very extreme effects because these would cause the greatest inflation of the summary result. Reported extremely large effects are implausible. Very few interventions in medicine have truly huge effects [12,13]. Furthermore, the documented history of gabapentin marketing may add further suspicion when extremely large effects are reported in specific trials.

The present study aims to explore gabapentin meta-analyses for the presence of studies with implausibly favorable results. We first sought evidence of excess significance (having too many studies with significant results), which may reflect publication bias, selective analysis, and outcome reporting bias [14]. Then we focused specifically on studies with extremely large effect sizes and tried to examine their provenance, reported funding, and reported conflicts of interest. We considered both gabapentin and pregabalin, given their similarity in terms of drug class (gabapentinoids), mechanism of action, and manufacturer [15].

2. Methods

2.1. Data sources and search strategy

We aimed to identify meta-analyses on gabapentin or pregabalin on any outcome. The last search was conducted in June 2016. Google Scholar, PubMed, and Embase were used for searches. We searched for relevant articles with the strings: gabapentin AND meta-analysis, pregabalin AND meta-analysis, gabapentin AND systematic review, pregabalin AND systematic review, gabapentin AND pooled analysis, and pregabalin AND pooled analysis.

2.2. Study selection and evaluation

The full text of potentially eligible articles was evaluated. Articles were retained if they were in English for any indication/condition, and there was enough information provided to use study-level data, particularly the number of participants in each of the two compared groups and the standardized effect size (expressed as Cohen's d, Hedges' g, or other similar standardized metrics) for continuous outcomes or the 2×2 table for binary outcomes.

When more than one meta-analyses on the same condition were eligible, only the most recent more inclusive article (the one with the largest number of studies) was retained. We assessed the primary outcomes within each meta-analysis, and when it was not clearly stated which outcome was the primary one, we selected the first reported outcome with available data presented per study. For the two largest meta-analyses (effects of gabapentin and of pregabalin on postoperative pain), we considered secondarily all outcomes that presented forest plots or tabulated data with study-level information, so that we can examine better the prevalence of studies with outlier extremely large effects.

2.3. Excess significance testing

A single investigator (K.D.) extracted data from forest plots and tables of relevant outcomes for each meta-analysis. Standardized effect sizes were extracted in each data set, and it was noted whether the results were statistically significant ("positive," $P < 0.05$) or not. All effect sizes were calculated as standardized mean differences (SMD); for dichotomous binary outcomes, odds ratios were calculated and converted to SMD using the Chinn transformation [16]. For each dataset in each meta-analysis, we estimated the power to detect at $\alpha = 0.05$ a plausible effect. We used a previously developed method to test for excess of significance in meta-analyses [14,17]. The sum of the power estimates gives the number of expected "positive" datasets. The expected number of "positive" datasets was

then compared against the observed number of "positive" studies for each outcome using chi-square exact tests.

In the speculated presence of major bias, we used different approaches and estimates on what a plausible effect for gabapentin or pregabalin would be for these outcomes. Typically, one wants to minimize the impact of publication and other reporting biases on the calculation of the plausible effect using the results of the largest studies [14,18]. However, in these meta-analyses, no trials stood out as being particularly large, and several trials had extremely large effect sizes. Therefore, first, we estimated the summary effect size according to fixed effects, and we identified outlier studies with extreme effects, defined a priori as those with effect size >0.8 standard deviations away from the summary fixed effect size. The rationale for our "rule of 0.8" is that large effects in the standardized scale are typically considered those that are >0.8. Then we estimated the fixed effects summary of the three largest studies, excluding any outliers, and we calculated the expected number of "positive" studies based on that summary effect. Furthermore, because this estimate is also likely to be biased (perhaps even heavily so for these drugs), we calculated also expected numbers of "positive" studies with different assumed effect sizes ranging from 0.2 (small effect) to 0.8 (large effect).

2.4. Evaluation of outlier studies

Finally, we identified all outlier studies (as defined previously) for all eligible meta-analyses and examined their primary publications in more depth. We explored whether outliers with more extremely favorable results than the summary effect (favorable outliers) differed from unfavorable outliers in country of origin (whether they came from less developed countries without long-standing tradition in clinical research, as defined in the study by Ioannidis et al. [19]). When we documented a strong difference in this regard, we performed additional exploratory analyses to try to explain the observed patterns. Specifically, we explored whether favorable and unfavorable outliers differed in reported funding and reported conflicts of interest. We do not report *P* values for these exploratory analyses because they were done post hoc. Moreover, given the observed patterns of potential bias, we also performed meta-analyses summarizing the results of studies that came from more developed countries and that also had no hints of bias in funding of conflicts of interest (i.e., they reported funding only from nonprofit sources or declared specifically that there was no funding and no conflicts of interest).

3. Results

3.1. Eligible meta-analyses

The literature search yielded 82 items that required a detailed screening, of which 34 were excluded after perusing the full-length articles. The remaining 48 articles were potentially eligible. Twenty-four were for gabapentin,10 were for pregabalin, and 14 were for both gabapentin and pregabalin. All articles were categorized based on indication, and within each category, the articles were examined for potential overlap. The most recent and comprehensive meta-analyses within each category were included in the analytical sample. Therefore, eventually, data from seven articles were eligible for in-depth analysis.

Table 1 summarizes the seven articles evaluated in-depth. They included 296 distinct datasets spread across seven different conditions, measuring 10 different outcomes that were subjected to meta-analysis. The size of each meta-analysis varied with the smallest including 647

Table 1. Summary table for included meta-analyses

Drug, article, condition	Outcome	Study data set, no.	Experimental/ controls, no.	Effect size, standardized mean difference (95% CI)	I^2, %
Gabapentin					
Al-Bachari, partial epilepsy	50% reduction in seizure frequency	6	705/501	0.41 (0.22, 0.60)	0
Doleman, postoperative pain	24-h morphine consumption	66	2,852/2,405	−1.08 (−1.14, −1.02)	93.5
Achuthavan, postoperative nausea and vomiting	Nausea	10	324/323	0.29 (0.08, 0.49)	12
Finnerup, neuropathic pain	50% pain intensity reduction	14	1,963/1,738	0.13 (0.11, 0.15)	55.5
Pregabalin					
Lam, postoperative pain	2-hr pain scores	60	2,514/2,104	−0.97 (−1.19, −0.74)	92.3
Lam, postoperative pain	24-hr pain scores	56	2,894/2,043	−0.44 (−0.66, −0.22)	91.5
Lam, postoperative pain	24-hr morphine consumption	46	2,307/1,658	−0.93 (−1.2, −0.65)	92.5
Boshcen, anxiety	Total score on Hamilton Anxiety Rating Scale	7	719/633	0.36 (0.26, 0.47)	58
Costa, epilepsy	≥50% reduction in seizure frequency	6	1,300/567	0.71 (0.43, 0.98)	68.4
Finnerup, neuropathic Pain	50% pain intensity reduction	25	4,000/5,024	0.13 (0.11, 0.15)	68.4

participants and the largest including 9,024. The largest meta-analyses for both drugs were on postoperative pain, including between 46 and 66 study datasets each. All meta-analyses showed statistically significant findings in favor of gabapentin and pregabalin for the selected outcomes.

Three outcomes from two different articles had large point estimates ($d > 0.8$ in absolute magnitude), and another five effects were also substantial ($d = 0.29–0.71$ in absolute magnitude). Heterogeneity I^2 estimates exceeding 50% were noted in eight of the 10 outcomes in the meta-analyses, and four (the ones on postoperative pain) had extreme I^2 estimates exceeding 90%; all of which also had large summary effects (absolute $d > 0.8$).

The extreme heterogeneity was driven by the presence of multiple studies with outlier extreme effects in these four meta-analyses of postoperative pain (15–26 outlier studies per meta-analysis, total of 77 outlier study estimates; Table 2). These four meta-analyses were the ones with the largest number of studies, that is, 24-hour morphine consumption for gabapentin in postoperative pain, and three outcomes for pregabalin in postoperative pain (2-hour pain [primary outcome], 24-hour pain, and 24-hour morphine consumption). For gabapentin in postoperative pain, 26 of the 66 studies were outliers. For pregabalin in postoperative pain, 39 studies had an outlier in at least one of the three outcomes. Of the other six meta-analyses, one had only one outlier study, and the other five meta-analyses had no outliers (Table 2). Of the 77 outliers, 39 showed larger benefits for gabapentin or pregabalin than the summary effect would suggest (favorable outliers), and 38 showed smaller benefits or no benefit (unfavorable outliers).

3.2. Expected and observed "positive" datasets

Table 2 also shows the effect size that we estimated based on the three largest studies in each meta-analysis (excluding outliers). As shown, these three studies typically were still not very large (total sample size 150–1,475 per meta-analysis). For five of 10 outcomes, the observed number of "positive" datasets was significantly larger than the expected using this plausible effect (Table 2), but as explained in the Section 2, we felt that even this plausible effect is probably inflated.

Table 3 depicts the expected number of "positive" studies assuming different plausible effect sizes. For a small plausible effect ($d = 0.2$), the total number of expected "positive" datasets would be 55 vs. 146 observed. For $d = 0.2$, all 10 meta-analyses showed excess significance ($P < 0.05$ one-tailed) for the excess of observed vs. expected "positive" results. The total number of expected "positive" datasets became 96, 137.8, and 179.2 with plausible effects of $d = 0.3$, $d = 0.4$, and $d = 0.5$. Therefore, there was excess significance for a plausible $d = 0.3$, but not for $d = 0.4$ or higher.

3.3. Evaluation of outlier studies

Details on the outlier studies for the two topics with the largest number of studies appear in Supplementary Table 1. In the gabapentin meta-analyses for postoperative pain, 14 trials had favorable outliers, whereas 12 trials had unfavorable outliers. All 14 trials with favorable outliers, but only four of the 12 trials with unfavorable outliers were from less developed countries (exact $P = 0.0003$). In the

Table 2. Plausible effect, observed, and expected positive data

Drug, article, condition	Outcome	Number of outliers	Total sample size in the three studies used for plausible effects (experimental/control)	Plausible effect	*P* values	Observed positive data sets, no.	Expected positive data sets, no.
Gabapentin							
Al-Bachari, partial epilepsy	50% reduction in seizure frequency	0	490/335	0.40	0.55	3	4.3
Doleman, postoperative pain	24-h morphine consumption	26	227/179	0.84	0.006	41	58
Achuthavan, postoperative nausea and vomiting	Nausea	1	125/125	0.23	0.27	3	1.5
Finnerup, neuropathic pain	50% pain intensity reduction	0	634/505	0.11	0.006	9	2.0
Pregabalin							
Lam, postoperative pain	2-hr pain scores	17	216/162	0.63	<0.001	25	42.9
Lam, postoperative pain	24-hr pain scores	15	694/542	0.42	0.17	16	23.4
Lam, postoperative pain	24-hr morphine consumption	19	293/203	0.60	0.29	25	29.9
Boshcen, anxiety	Total score on Hamilton Anxiety Rating Scale	0	408/325	0.34	0.09	6	3.2
Costa, epilepsy	≥50% reduction in seizure frequency	0	655/294	0.15	0.08	4	0.8
Finnerup, neuropathic pain	50% pain intensity reduction	0	620/855	0.04	<0.001	14	1.8

Table 3. Expected positive data sets with varying plausible effect size

Drug, article	Expected positive data sets (plausible effect = 0.2), no.	Expected positive data sets (plausible effect = 0.3), no.	Expected positive data sets (plausible effect = 0.4), no.	Expected positive data sets (plausible effect = 0.5), no.	Expected positive data sets (plausible effect = 0.6), no.	Expected positive data sets (plausible effect = 0.7), no.	Expected positive data sets (plausible effect = 0.8), no.
Gabapentin							
Al-Bachari, partial epilepsy	1.7	3.1	4.3	5	5.4	5.6	5.7
Doleman, postoperative pain	9.2	16.5	25.8	35.2	44.5	51.6	56.1
Achuthavan, postoperative nausea	1.3	2.0	3.5	4.9	6.4	7.6	8.7
Finnerup, neuropathic pain	4.8	7.8	10.2	11.4	12.1	11.6	12.9
Pregabalin							
Lam, postoperative pain, 2-h pain	8.3	14.7	23.1	32.2	40.7	47.7	52.8
Lam, postoperative pain, 24-h pain	8.1	14.5	22.4	30.8	38.6	44.8	49.3
Lam, postoperative pain, 24-h morphine consumption	6.4	11.6	17.5	23.9	29.9	35.1	39.1
Boshcen, anxiety	1.9	3.7	5.3	6.4	6.8	6.9	7.0
Costa, epilepsy	2.1	3.9	5.2	5.8	6.0	6.0	6.0
Finnerup, neuropathic pain	11.2	18.2	20.5	23.6	24.3	24.5	24.7

pregabalin meta-analyses for postoperative pain, there were 51 outliers for the three outcomes from 39 different trials. Of these 51 outliers, 25 were favorable, and 26 were unfavorable. Nineteen of the 25 favorable outliers, but only nine of the 26 unfavorable ones were from less developed countries (exact $P = 0.0047$). For gabapentin and pregabalin combined, most unfavorable outliers (27 of 38) were in trials from more developed countries, whereas very few favorable outliers (6 of 39) were in trials from more developed countries (exact $P < 0.0001$).

The median sample size of trials with favorable outliers and those with unfavorable outliers was similar (data not shown). Outliers from more vs. less developed countries also had similar sample sizes (median 63.5 vs. 60).

Thirteen favorable outliers of gabapentin for postoperative pain did not mention any sponsor/funding source, and one stated that they received no funding. Nine favorable outliers of pregabalin did not mention any sponsor/funding, three acknowledged industry funding, three were funded by other not-for-profit organizations, and six stated that they received no funding. Of the trials with unfavorable outliers with gabapentin or pregabalin, 10 did not mention any funding, three acknowledged funding by the industry, eight were funded by nonprofit organizations, and five stated that they received no funding.

Among gabapentin trials with favorable outliers, 13 did not mention anything about conflicts of interest, and one declared no conflicts of interest. Among pregabalin trials with favorable outliers, seven did not mention anything about conflicts of interest, 11 declared no conflict of interest, and three reported conflict of interest because of industry affiliations. Of the trials with unfavorable outliers with gabapentin or pregabalin, eight did not mention anything about conflicts of interest, 12 declared no conflict of interest, and five reported conflict of interest because of industry affiliation.

Meta-analysis of the trials from more developed countries and no hints of bias showed a summary effect of

−0.22 (95% confidence interval [CI], −0.53 to 0.07), that is, a nonsignificant benefit, for gabapentin for postoperative pain (two trials, $n = 181$ participants). The meta-analysis of trials from more developed countries and no hints of bias for pregabalin had a summary effect of −0.55 (95% CI, −0.81 to −0.28) for the outcome pain after 2 hours (four trials, $n = 255$ participants), 0.22 (95% CI, 0.02–0.43) for pain at 24 hours (five trials, $n = 385$ participants), and −0.44 (95% CI, −0.71 to −0.18) for morphine consumption at 24 hours (four trials, $n = 255$ participants). Results should be viewed cautiously, given the imprecision because of few remaining studies.

4. Discussion

Our evaluation of meta-analyses of gabapentin and pregabalin on 10 different outcomes for different conditions showed strong evidence for excess significance bias in this literature. The number of statistically significant results in the component studies included in these meta-analyses was much larger than could be reasonably expected, given that the vast majority of these studies have been of very small sample size. For the largest meta-analyses (those on postoperative pain management), we identified both favorable and unfavorable outliers for gabapentin and pregabalin. Favorable and unfavorable outliers were markedly different. Favorable outliers almost always came from less developed countries with no well-established tradition of clinical research, and they typically reported no information on funding or conflicts of interest. Conversely, unfavorable outliers came mostly from more developed countries and reported funding sources and conflicts of interest. When limited to the data from studies from more developed countries and no hints of bias, the limited evidence showed no benefit from gabapentin for postoperative pain and either harm or much more modest benefit from pregabalin for the same indication (three outcomes), whereas the meta-analyses of all studies had suggested very large benefits.

Studies with outlier, extremely favorable effect sizes were always of small or very small sample size. An empirical assessment of very large effects (odds ratios >5) across the entire Cochrane Library [19] has shown that very large effects are seen occasionally in very small studies, but almost never get validated when a second trial is done on the same topic. Other empirical evaluations [13] show that genuine, very large effects are extremely rare. It is unlikely that the pain literature should be much different in this regard. Even if we assume that some pain treatments may achieve very good results, treatment effects exceeding by far the threshold of large effects ($d = 0.8$), as we observed here, are suspicious. Most of the favorable outliers represented logically implausible standardized effects of $d = 2–4$ or even larger. Thus, we presume that the favorable outliers reported are unlikely to reflect the underlying true effect (if any) of gabapentin or pregabalin. Studies from countries with well-established clinical research tradition and no hints of bias are few, but they clearly contradict the results of the favorable outlier studies.

A previous evaluation has shown that studies done in less developed countries with no tradition in clinical research are reporting on average stronger effects than those done on the same topic in more developed countries [20]. However, in the previous evaluation, the difference in effect sizes was small (relative odds ratio of ≤ 1.15). This may also reflect the fact that mortality and major hard clinical outcomes were assessed in that previous evaluation. Here, we have documented extremely large, outlying effects in studies from such countries. It is possible that very small studies need to report extreme results to reach significance and have some perceived "positive" message to get published. Authors may think that this is a way to attract attention and allow their publication. Our data suggest that reviewers and editors should be wary of such extreme effects in the absence of strong evidence for plausibility.

One may speculate whether these extremely favorable outlier effects conceal hidden industry conflicts of interest. Of the 11 published trials that were analyzed because of being present in the court documents [5], two were included in the meta-analyses discussed in this article, and both were not found to be outliers. This very low rate of reporting of industry funding in these trials does not exclude industry involvement, however. Several empirical studies have shown the large underreporting of conflicts of interest in medical research in a wide variety of settings [21,22]. With improved reporting over time [23], it seems that industry sponsoring and conflicts are more prevalent than previously reported. It is unknown whether trials funded by the industry in less developed countries may be particularly prone not to report their funding. Alternatively, it may be that these extremely favorable results with no mention of industry funding did indeed receive no industry support. In this scenario, trials in less developed countries offered implausible results because of various flaws in the design, conduct, analysis, and reporting without any industry interference.

One may speculate whether differences between trials done in more vs. less developed countries may also reflect methodological limitations in the latter group. Two systematic evaluations of gabapentin that were published recently (after our search dates) [24,25] found that the large majority had high risk of bias. Among 74 randomized trials of gabapentin for postoperative pain, only eight had low risk of bias; and thus, the reviewers were careful to avoid strong conclusions [24]. Another publication assessed 132 trials of gabapentin. The analysis of reduction in 24-hour opioid consumption could include only 13 trials with low risk of bias, and the analysis of serious adverse events could include only nine trials with low risk of bias. The reviewers concluded that the evidence, also appraised on the Grading of Recommendations Assessment, Development and Evaluation scale, was very tenuous [25]. Both meta-analyses on postoperative pain that we included had also used Cochrane

risk of bias assessments for the included trials. Only three outliers from less developed countries were assessed as having high risk of bias on any item. However, this may reflect missing information, optimistic evaluation by the data extractors (especially in the pregabalin meta-analysis where only four trials had any item assessed as high risk of bias), or even an effort by the trialists to hide bias.

Some genuine differences in the management of pain may also exist between more and less developed countries. Doleman et al. [26] have observed that trials with higher baseline pain or higher morphine consumption in their participants have higher treatment effects. One may speculate whether trials in less developed countries use less intensive concurrent analgesic strategies after surgery and thus have higher baseline pain. Nevertheless, the use of standardized treatment effect helps reduce this as a potential issue, as the higher baseline trials tend to have a large SDs.

Some limitations should be discussed. First, excess significance tests offer hints of bias but are not offering definitive proof for bias or its exact mechanism. Similarly, a negative test for excess significance, especially for meta-analyses with a limited number of studies, does not exclude the potential for bias. Second, the exact estimation of excess statistical significance is influenced by the choice of plausible effect size. We performed analyses using different plausible effect sizes, and under most scenarios, there is evidence for substantial excess significance. Even if the plausible effect for these drugs in the assessed conditions is large, the multiple outlier studies are still presenting a strong paradox that is difficult to explain in the absence of bias. Finally, our choice of how to define the outlier studies is using a cut-off of 0.8 standard deviations away from the summary effect. Different cut-off definitions would give a different number of outliers. In all, the stark contrast of the results of these outliers distinguishes them from the other studies included in the meta-analysis. However, even some/many of the other studies may still be biased and may be subject to some of the same conflicts of interest or other flaws. A literature that has such a well-documented history of known subversion needs to be seen with substantial skepticism even for more modest results.

In conclusion, we document the high prevalence of extreme effects with a large number of outlier studies in the meta-analyzed literature of gabapentin and pregabalin, in particular, for the most commonly studied topic of postoperative pain. Extreme heterogeneity and the high prevalence of extreme effects in outlier studies should be hallmarks that raise suspicion in meta-analyses of medical interventions and may be the footprint of major biases.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2018.10.012.

References

[1] US Ex Rel, Franklin V, Parke D. 147 F. USA: U.S. District Court for the District of Massachusetts; 2004.
[2] Steinman MA, Bero LA, Chren MM, Landefeld CS. Narrative review: the promotion of gabapentin: an analysis of internal industry documents. Ann Intern Med 2006;145:284−93.
[3] Vedula SS, Goldman PS, Rona IJ, Greene TM, Dickersin K. Implementation of a publication strategy in the context of reporting biases. A case study based on new documents from Neurontin® litigation. Trials 2012;13(1):1.
[4] Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. N Engl J Med 2009;361:1963−71.
[5] Vedula SS, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of gabapentin. PLoS Med 2013;10(1):e1001378.
[6] Krumholz SD, Egilman DS, Ross JS. Study of neurontin: titrate to effect, profile of safety (STEPS) trial: a narrative account of a gabapentin seeding trial. Arch Intern Med 2011;171:1100−7.
[7] Al-Bachari S, Pulman J, Hutton JL, Marson AG. Gabapentin add-on for drug-resistant partial epilepsy. Cochrane Database Syst Rev 2013;(7):1−43.
[8] Doleman B, Heinink T, Read D, Faleiro R, Lund J, Williams J. A systematic review and meta-regression analysis of prophylactic gabapentin for postoperative pain. Anaesthesia 2015;70(10):1186−204.
[9] Toulis KA, Tzellos T, Kouvelas D, Goulis DG. Gabapentin for the treatment of hot flashes in women with natural or tamoxifen-induced menopause: a systematic review and meta-analysis. Clin Ther 2009;31:221−35.
[10] Backonja M, Glanzman RL. Gabapentin dosing for neuropathic pain: evidence from randomized, placebo-controlled clinical trials. Clin Ther 2003;25:81−104.
[11] Chadwick D, Anhut H, Greiner M, Alexander J, Murray GH, Garofalo EA, et al. A double-blind trial of gabapentin monotherapy for newly diagnosed partial seizures. International Gabapentin Monotherapy Study Group 945-77. Neurology 1998;51:1282−8.
[12] Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. JAMA 2012;308:1676−84.
[13] Nagendran M, Pereira TV, Kiew G, Altman DG, Maruthappu M, Ioannidis JP, et al. Very large treatment effects in randomised trials as an empirical marker to indicate whether subsequent trials are necessary: meta-epidemiological assessment. BMJ 2016;355:i5432.
[14] Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials 2007;4:245−53.
[15] Bockbrader HN, Wesche D, Miller R, Chapel S, Janiczek N, Burger P. A comparison of the pharmacokinetics and pharmacodynamics of pregabalin and gabapentin. Clin Pharmacokinet 2010;49:661−9.
[16] Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 2000;19:3127−31.
[17] Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ 2014;348:g2035.
[18] Ioannidis JP. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Math Psychol 2013;57(5):184−7.
[19] Ioannidis JP, Pereira TV, Horwitz RI. Emergence of large treatment effects from small trials—reply. JAMA 2013;309:768−9.
[20] Panagiotou OA, Contopoulos-Ioannidis DG, Ioannidis JP. Comparative effect sizes in randomised trials from less developed and more developed countries: meta-epidemiological assessment. BMJ 2013;346:f707.
[21] Krimsky S, Rothenberg L. Financial interest and its disclosure in scientific publications. JAMA 1998;280:225−6.

[22] Krimsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. Sci Eng Ethics 2001;7(2):205−18.

[23] Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible research practices and transparency across the biomedical literature. PLoS Biol 2016;14(1):e1002333.

[24] Fabritius ML, Geisler A, Petersen PL, Wetterslev J, Mathiesen O, Dahl JB. Gabapentin in procedure-specific postoperative pain management - preplanned subgroup analyses from a systematic review with meta-analyses and trial sequential analyses. BMC Anesthesiol 2017;17(1):85.

[25] Fabritius ML, Geisler A, Petersen PL, Nikolajsen L, Hansen MS, Kontinen V, et al. Gabapentin for post-operative pain management - a systematic review with meta-analyses and trial sequential analyses. Acta Anaesthesiol Scand 2016;60(9):1188−208.

[26] Doleman B, Sutton AJ, Sherwin M, Lund JN, Williams JP. Baseline morphine consumption may explain between-study heterogeneity in meta-analyses of adjuvant analgesics and improve precision and accuracy of effect estimates. Anesth Analg 2018;126:648−60.