International Journal of Risk & Safety in Medicine 24 (2012) 221–231
DOI 10.3233/JRS-2012-0571
IOS Press

The paroxetine 352 bipolar trial: A study in medical ghostwriting

Jay D. Amsterdam[a,*] and Leemon B. McHenry[b]
[a]*Depression Research Unit, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[b]*Department of Philosophy, California State University, Northridge, CA, USA*

Abstract. *Background:* The problem of ghostwriting in corporate-sponsored clinical trials is of concern to medicine, bioethics, and government agencies. We present a study of the ghostwritten archival report of an industry-sponsored trial comparing antidepressant treatments for bipolar depression: GlaxoSmithKline (GSK) paroxetine study 352. This analysis is based upon publicly available evidence presented in a complaint of research misconduct filed with the Office of Research Integrity of the Department of Health and Human Services.

Objectives: We performed a deconstruction of the published study to show how primary and secondary outcome analyses were conflated, turning a 'negative' clinical trial into a 'positive' study – with conclusions and recommendations that could adversely affect patient health.

Methods: The paroxetine 352 study was a randomized, double-blind, placebo-controlled, 19-site trial comparing paroxetine and imipramine in 117 patients with bipolar type I major depressive episode which was unresponsive to prior lithium carbonate therapy.

Results: Analysis of the primary outcome measures found no statistically significant difference between paroxetine or imipramine versus placebo. However, the published article concluded that both drugs were efficacious versus placebo for a *post hoc* subgroup of patients.

Conclusions: Few industry-sponsored studies gain public scrutiny. It is important to make these articles transparent to the scientific and medical community.

Keywords: Ghostwriting, key opinion leaders, depression, bipolar, SSRI, paroxetine, litigation, industry sponsorship

1. Introduction

The problem of truth and transparency in published scientific reports of corporate-sponsored clinical trials has been an on-going concern in the medical and bioethics literature. The difference between what a trial should report and what is actually reported in the medical journals in the past 30 years is so alarming that some editors have declared a crisis of credibility [1]. Details of selective data reporting, misrepresentation of results, and ghostwriting of manuscripts have been revealed from the fragmentary release of documents publicly disclosed in litigation in the United States [2]. Critical evaluation of these practices, however, is sporadic at best because of inherent non-disclosure or inaccessibility of

*Address for correspondence: Jay D. Amsterdam, MD, Depression Research Unit, University Science Center – 3rd Floor, 3535 Market Street, Philadelphia, PA 19104, USA. Tel.: +1 215 662 3462; Fax: +1 215 662 6443; E-mails: enopath@aol.com, jamsterd@mail.med.upenn.edu.

information that remains under judicial seal by the courts. Even the most egregious cases of corporate fraud in the pharmaceutical industry are protected from public disclosure and only become available when attorneys challenge the confidentiality status of the documents. Another, but less common, source of information about unethical reporting and publishing comes from whistleblower complaints where honest investigators refuse to compromise scientific accuracy for commercial or institutional objectives [3].

This report describes how a 'negative' clinical trial was published as a 'positive' study with unsubstantiated claims of efficacy and safety. It is based upon public evidence presented in a complaint of research misconduct filed with the Office of Research Integrity (ORI) of the Department of Health and Human Services (HHS) [4]. Additional supporting documents are available, but remain under current court seal. The article, entitled "Double-Blind, Placebo-Controlled Comparison of Imipramine and Paroxetine in the Treatment of Bipolar Depression," was published in the *American Journal of Psychiatry* in June 2001 [5]. It appears to be part of a marketing effort by SmithKline Beecham (SKB), now GlaxoSmithKline (GSK), to have paroxetine (Paxil®/Seroxat®) outsell all of the competition in the selective serotonin reuptake inhibitor (SSRI) antidepressant market for all indications [6]. This strategic plan included the ghostwriting of articles by medical communication companies that would 'spin' the data in favor of the sponsor's product. For several of these unapproved, but much-sought after indications, we believe that ghostwritten articles were placed in leading medical journals to facilitate off-label prescriptions by clinicians [7]. A key component of this *sub rosa* plan was for prominent academic researchers, known in industry as 'key opinion leaders,' to lend their names to these publications as authors, which would give the appearance of scientific objectivity.

It is difficult to believe that the publication of the paroxetine 352 study was for any other purpose than to facilitate the prescription of paroxetine for the treatment of the depressive phase of manic depressive, or bipolar, disorder (i.e., bipolar depression). This area of treatment represented a natural extension of the already approved indication for paroxetine of unipolar major depressive disorder (or non-bipolar depression). However, despite the potentially lucrative aspect of this indication for marketing paroxetine, it was also well known in the psychiatry community that antidepressants (including SSRIs) could worsen bipolar disorder by precipitating manic episodes [8–10] and increasing the risk of suicide in patients [11–13].

In this report, we deconstruct the paroxetine 352 study and show how the ghostwriters conflated the primary and secondary outcome analyses that turned a 'negative' study into a 'positive' endorsement of antidepressant therapy (in particular paroxetine) for patients with bipolar depression.

2. Methods

2.1. Initial study design

Deconstruction of the paroxetine 352 study is based upon information obtained from the GSK Clinical Trials Website Result Summary for Study 29060/352 updated 09 March 2005 [14], GSK Paroxetine Protocol PAR-29060/352 (amended 22 July, 1994), Drug Industry Document Archives Website [15], and evidence presented in a complaint of research misconduct filed with the federal ORI filed on July 8, 2011 by Dr. Jay D. Amsterdam, Professor of Psychiatry at the University of Pennsylvania [4]. The paroxetine 352 study was conducted between February 1994 and March 1996. It was originally designed as a 10-week, 18-site, randomized, double-blind, placebo-controlled comparison of paroxetine versus imipramine in subjects with bipolar type I disorder and was designated a Phase IV (i.e., post-marketing,

non-indication) study with a projected duration of 2 years. Its objective was "to compare the efficacy and safety of paroxetine and imipramine to [placebo] in the treatment of bipolar depression in subjects stabilized on lithium therapy [14]". The primary efficacy measures, as described in the study protocol, were the change from baseline in the 17-item Hamilton Rating Scale for Depression (HAM-D) total score [16], and the change from baseline in the Clinical Global Impression Severity of Illness (CGI-S) score [17] for paroxetine versus placebo and for imipramine versus placebo. The stated secondary efficacy measures were the proportion of subjects with a final HAM-D score ≤ 7 or a final CGI-S score ≤ 2. Additional secondary outcomes included the proportion of subjects experiencing adverse events, premature treatment discontinuation, and manic or hypomanic reactions as determined by the DSM-III-R Mania/Hypomania Assessment and the Young Mania Rating Scale (YMRS) [18]. A post-treatment safety assessment was performed after subjects were discontinued from double-blind medication to determine the proportion of subjects in each treatment group with discontinuation symptoms due to adverse events (e.g., withdrawal and/or manic symptoms).

The study population consisted of outpatient subjects ≥ 18 years old, with a lifetime diagnoses of bipolar type I disorder and a history of at least one prior manic or major depressive episode within the preceding 5 years. Subjects had a current major depressive episode who failed to respond to lithium carbonate therapy (or lithium carbonate plus sodium valproate or carbamazepine therapy) for ≥ 7 weeks with a serum lithium level of 0.5–1.2 mEq/L for ≥ 6 weeks prior to starting the study. Treatment procedures for the study have been published [5, 14].

2.2. Sample size estimate

The original protocol called for a sample size of 62 subjects per treatment group (or a total of 186 subjects). However, during the course of the study the protocol sample size estimate was formally amended downward to 46 subjects per treatment group (or a total of 138 subjects) [14].

2.3. Statistical methods

A summary of the statistical plan in the amended study protocol called for analyses to be performed on two sets of efficacy data (i.e., interim visit-wise data and endpoint data), with endpoint data considered the primary data set. Unspecified statistical procedures were to be applied to the outcome data to remove potential rating bias that might occur in subjects with concurrent depressive and manic symptoms. Separate analyses were to be performed on the entire subject population, and on two subgroups of subjects: (i) those who experienced a manic or hypomanic episode during the study; and, (ii) those who did not experience a manic or hypomanic episode during the study. The YMRS was to be used to assess severity of manic and/or hypomanic symptoms across treatment groups, and the relationship between change from baseline in YMRS scores and HAM-D scores was to be specifically examined.

Factors that might influence treatment outcome were to be examined via the use of interaction terms in the regression models (i.e., treatment, investigator site, strata of "high" or "low" baseline lithium level). However, it was specifically noted that any interaction term that was not statistically significant (i.e., $p > 0.1$) in the primary analysis would be dropped from all subsequent analyses. The protocol noted that the comparison of primary interest is paroxetine versus placebo across (regardless of) lithium strata, performed at a two-tailed significance level of $p = 0.05$. The protocol also required that differences among subject groups in baseline demographic and diagnostic variables would be checked and, if differences existed for variables predictive of response, their impact on the results would be investigated.

Finally, mania and hypomania were to be analyzed using logistic regression models that included the effect terms of 'treatment', 'investigator', and 'treatment × investigator' interaction. The protocol noted that if the interaction was not significant, it would be dropped from the model.

3. Results

3.1. Changes in original study methodology and reporting

The original protocol sample size estimate of 0.9 (1-β) or 62 subjects per treatment group was officially amended downward to 0.8 (1-β) or 46 subjects per group during the study. The latter value was the sample size described in the GSK Clinical Trials Website Result Summary [14]. No explanation was provided for this change in sample size in the amended protocol. However, we suspect that this reduction in power might have resulted from exceedingly slow subject recruitment into the study, which ultimately led GSK to add a 19th investigative site. By the time GSK decided to halt subject enrollment prematurely and terminate the study, only 117 (of the originally projected 186 subjects) were enrolled, resulting in final sample sizes for paroxetine ($n = 35$), imipramine ($n = 39$), and placebo ($n = 43$). By the time the study was published in June 2001 in the *American Journal of Psychiatry*, however, the declared sample size estimate had again changed with the article stating: "The study was designed (*sic*) to enroll 35 patients per arm, which would allow 70% power to detect a 5-point difference on the Hamilton depression scale score (SD = 8.5) between treatment groups [5]".

Although the published article noted that statistical power was estimated at only 70%, the article did not inform the reader that this value represented an unconventionally low power for a clinical trial. The article did not inform the reader that the original power estimate was 62 subjects per group or that the original power estimate had been officially reduced during the study. Moreover, the article made no mention of the fact that the final power estimate was determined after the study was completed, and that this *post hoc* power estimate most likely occurred as an 'extra-regulatory' protocol change in order to allow the final sample size estimate of 35 subjects per group to comport with the final sample size of the paroxetine group (i.e., $n = 35$). The published article failed to acknowledge clearly that the study failed to recruit the projected sample size necessary to test the primary study hypothesis, and only hinted by its published sample size estimate that the study had insufficient statistical power to test the primary study aims.

The statistical plan described in the GSK Clinical Trials Website Results Summary [14] was considerably less detailed and somewhat different than that of the original or amended study protocol. It presented the primary comparison of interest as paroxetine versus placebo (regardless of baseline lithium level strata). It noted that comparisons of secondary interest were imipramine versus placebo and imipramine versus paroxetine. The statistical plan concluded with the statement: "In addition, analyses were performed within lithium strata, which included only an effect for treatment". The GSK Clinical Trials Website provided no indication that these secondary analyses were performed as *post hoc* procedures, or that they were not statistically warranted – as the primary outcome analyses showed no significant treatment × lithium level strata interaction effect for any of the treatment groups [14].

The analysis plan of the published article played down the statistical procedures used for the primary efficacy analyses and, instead, emphasized the procedures used for analyzing the *post hoc* lithium level strata efficacy analyses. No mention was made in the published article indicating that the sample size of 35 subjects per group represented insufficient statistical power to test adequately for differences among

lithium level subgroups [5]. Moreover, the statistical plan presented in the published article noted that "no adjustments for multiple comparisons were made [5]" which, if properly applied, would have nullified the statistical significance of the only two 'positive' comparisons found in the study.

Finally, in contrast to the amended study protocol that called for the analysis of safety measures to examine manic symptoms (i.e., DSM-III-R Mania/Hypomania Assessment and YMRS), no mention of these measures was made in the GSK Clinical Trials Web-site Results Summary or in the published article, and no safety data acquired from these measures were presented in either published venue.

3.2. Conflation of primary and secondary outcome measures

We believe that GSK conflated primary and secondary study aims, and then presented 'positive' *post hoc* lithium level strata analyses as if they were the primary analyses of interest (Table 1) [14]. Of the more than 30 separate efficacy analyses reported in the GSK Clinical Trials Web-site Results Summary [14], only two *post hoc* comparisons showed statistical significance (i.e., paroxetine versus placebo for change in HAM-D score in the "low" lithium level group ($p = 0.049$) and imipramine versus placebo for change in HAM-D score in the "low" lithium level group ($p = 0.038$) (Table 1).

The published article provided a cursory statement indicating that the primary outcome analysis was 'negative'. However, we believe that the article then conflated the primary and secondary analyses by attributing the 'negative' primary finding to an excessive placebo response rate in the "high" lithium level subgroup (although there is no discernible evidence in the study data to support this conclusion). After conflating the primary and secondary outcome measures, the published article then emphasized the only 'positive' efficacy finding for paroxetine as if it was the primary study aim: "... among the low serum lithium level patients, paroxetine and imipramine were superior to placebo in terms of mean change from baseline in scores on the Hamilton depression scale and CGI severity of illness scale... [5]".

In addition, we believe that the published article conflated efficacy and adverse event data to favor paroxetine over imipramine. For example, the article presented only selected safety data on treatment-emergent manic episodes that favored paroxetine. In contrast to the study protocol that called for the analysis of manic and hypomanic symptoms as measured by the DSM-III-R Mania/Hypomania Assessment and YMRS, no data from these safety measures were presented in either the GSK Clinical Trials Website Results Summary or in the published article. Instead, the GSK Clinical Trials Website Results Summary and the published article simply noted the number of manic episodes that were clinician-reported during the trial and appeared to emphasize the lack of manic episodes reported with paroxetine (versus imipramine and placebo).

We believe that the published article also favored the adverse event profile of paroxetine by portraying it as having virtually no sexual side effects (versus imipramine) with the article stating: "Patients treated with imipramine reported a higher incidence of abnormal ejaculation (18.8%) and impotence (25.0%) than did patients receiving paroxetine (0.0 and 6.3%, respectively) or placebo (5.0 and 0.0%, respectively) [5]". By conflating the 'positive' efficacy and favorable side effect profile of paroxetine, the article suggested that paroxetine was the obvious drug of choice for physicians to prescribe – especially when the article emphasized the greater sexual side effect burden and manic switch rate with imipramine and placebo.

However, closer examination of the available data suggests that this conclusion may be unwarranted. For example, although the published article noted that the most common side effects occurring with paroxetine (at a frequency $\geq 10\%$) were insomnia and somnolence, the published article failed to report a higher frequency of paroxetine-induced treatment-emergent depression (versus imipramine) – as listed in the GSK Clinical Trials Website Results Summary (Table 2). The higher frequency of insomnia and

Table 1
Primary efficacy results derived from the GSK Clinical Trials Website Result Summary [14]

Primary efficacy results (Intent to treat analysis): Baseline mean and change from baseline mean for the first 17-items of HAMD-21 LOCF	Paroxetine (Par)	Imipramine (Imp)	PBO
Total response dataset	*N* = 33	*N* = 36	*N* = 43
Baseline, *n* mean HAMD-21 (se)	33 20.38 (0.68)	36 20.71 (0.65)	43 21.57 (0.59)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.191	Imp vs. PBO 0.332	Par vs. Imp 0.725
Endpoint, *n* change from baseline mean (se)	33 –10.2 (1.27)	36 –10.1 (1.21)	43 –8.06 (1.11)
Endpoint pairwise comparison treatment *p*-values	Par vs. PBO 0.199	Imp vs. PBO 0.220	Par vs. Imp 0.932
High lithium level dataset	*N* = 14	*N* = 17	*N* = 21
Baseline, *n* mean (se)	14 20.29 (1.01)	17 21.35 (0.91)	21 21.95 (0.82)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.206	Imp vs. PBO 0.628	Par vs. Imp 0.436
Endpoint, *n* change from baseline mean (se)	14 –9.79 (1.90)	17 –9.35 (1.72)	21 –10.4 (1.55)
Endpoint pairwise comparison treatment *p*-values	Par vs. PBO 0.809	Imp vs. PBO 0.659	Par vs. Imp 0.867
Low lithium level dataset	*N* = 19	*N* = 19	*N* = 22
Baseline, *n* mean (se)	19 20.37 (0.92)	19 20.11 (0.92)	22 21.18 (0.85)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.519	Imp vs. PBO 0.394	Par vs. Imp 0.840
Endpoint, *n* change from baseline mean (se)	19 –10.4 (1.67)	19 –10.7 (1.67)	22 –5.82 (1.56)
Endpoint pairwise comparison treatment *p*-values	Par vs. PBO 0.049	Imp vs. PBO 0.038	Par vs. Imp 0.912
Baseline mean and change from baseline mean for CGI-S LOCF	Paroxetine	Imipramine	PBO
Total response dataset	*N* = 33	*N* = 36	*N* = 43
Baseline, *n* mean (se)	33 4.21 (0.12)	36 4.31 (0.11)	43 4.33 (0.10)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.476	Imp vs. PBO 0.897	Par vs. Imp 0.573
Endpoint, *n* mean (se)	33 –1.33 (0.24)	36 –1.28 (0.23)	43 –0.91 (0.21)
Endpoint Pairwise comparison treatment *p*-values	Par vs. PBO 0.196	Imp vs. PBO 0.245	Par vs. Imp 0.876
High lithium level dataset	*N* = 14	*N* = 17	*N* = 21
Baseline, *n* mean (se)	14 4.21 (0.16)	17 4.35 (0.15)	21 4.29 (0.13)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.738	Imp vs. PBO 0.739	Par vs. Imp 0.535
Endpoint, *n* change from baseline mean (se)	14 –1.14 (0.38)	17 –0.94 (0.35)	21 –1.24 (0.31)
Endpoint pairwise comparison treatment *p*-values	Par vs. PBO 0.848	Imp vs. PBO 0.528	Par vs. Imp 0.698
Low lithium level dataset	*N* = 19	*N* = 19	*N* = 22
Baseline, *n* mean (se)	19 4.21 (0.17)	19 4.26 (0.17)	22 4.36 (0.16)
Baseline pairwise comparison treatment *p*-values	Par vs. PBO 0.516	Imp vs. PBO 0.670	Par vs. Imp 0.829
Endpoint, *n*	19	19	22
Change from baseline mean (se)	–1.47 (0.31)	–1.58 (0.31)	–0.59 (0.29)
Endpoint pairwise comparison treatment *p*-values	Par vs. PBO 0.040	Imp vs. PBO 0.022	Par vs. Imp 0.810

depression in the paroxetine-treated subjects could have represented drug-induced manic, hypomanic, or mixed manic and depressive symptoms. However, the notable lack of data analyses presented on manic and hypomanic symptoms from the DSM-III-R Mania/Hypomania Assessment and the YMRS in the published article make this possibility difficult to resolve. Moreover, the complete absence of manic or hypomanic symptoms reported with paroxetine stands in stark contrast to the vast majority of published research articles on antidepressant-induced mood conversion symptoms in bipolar depression.

Table 2
Number (%) of reported and elicited adverse events derived from the GSK Clinical Trials Website Result Summary [14]

Most frequent adverse events	Paroxetine ($n=35$)	Imipramine ($n=39$)	Placebo ($n=43$)
Any adverse event	33 (94.3)	36 (92.3)	40 (93.0)
Tremor	14 (40.0)	15 (38.5)	4 (9.3)
Insomnia	13 (37.1)	5 (12.8)	10 (23.3)
Somnolence	12 (34.3)	13 (33.3)	11 (25.6)
Diarrhea	10 (28.6)	6 (15.4)	7 (16.3)
Nausea	9 (25.7)	12 (30.8)	5 (11.6)
Headache	9 (25.7)	16 (41.0)	17 (39.5)
Infection	6 (17.1)	2 (5.1)	1 (2.3)
Dizziness	5 (14.3)	11 (28.2)	1 (2.3)
Constipation	5 (14.3)	11 (28.2)	4 (9.3)
Asthenia	5 (14.3)	5 (12.8)	2 (4.7)
Dry Mouth	4 (11.4)	24 (61.5)	3 (7.0)
Depression	4 (11.4)	0	4 (9.3)
Dyspepsia	2 (5.7)	7 (17.9)	1 (2.3)
Sinusitis	2 (5.7)	2 (5.1)	4 (9.3)
Myalgia	1 (2.9)	0	5 (11.6)
Vomiting	0	6 (15.4)	1 (2.3)
Back pain	0	2 (5.1)	5 (11.6)
Serious adverse events*	0	2 (5.1)	4 (9.3)
Mania	0	1 (2.6)	0
Aggression/homicidal ideation	0	1 (2.6)	0
Manic episode	0	0	2 (4.7)
Paranoia, hallucinations, delusions	0	0	1 (2.3)
Re-emergence of depression	0	0	1 (2.3)
Subjects with fatal SAEs, *n* (%)	0	0	0

*Includes both fatal and non-fatal events.

4. Discussion

4.1. Misrepresentation and bias in reporting study outcomes

The recruitment of only 117 (of the originally projected 186) subjects resulted in a failed (i.e., non-informative) trial with inconclusive results. As a consequence, we believe that the published article had to rely on conflated *post hoc* analyses of data subsets to portray a favorable result for paroxetine. This favorable result was identified by subdividing the already insufficient sample size of each treatment group into smaller subgroups of "high" (i.e., 0.8–1.2 mEq/L) versus "low" (i.e., 0.4–0.79 mEq/L) lithium level strata. This procedure resulted in 6 treatment subgroups: three "high" lithium level subgroups of paroxetine ($n=14$), imipramine ($n=17$), and placebo ($n=21$) and three "low" lithium level subgroups of paroxetine ($n=19$), imipramine ($n=19$), and placebo ($n=22$). A pairwise comparison of treatments within these limited subgroups then identified two 'positive' findings: paroxetine superior to placebo in the "low" lithium level subgroup ($p=0.049$) and imipramine superior to placebo in the "low" lithium level subgroup ($p=0.038$). We believe that these 'positive' findings were then presented as if they were the

main finding of the study. A conflation of the 'positive' efficacy finding for paroxetine with the favorable side effect profile presented for paroxetine provided a positive 'spin' for paroxetine and a negative 'spin' for imipramine in the published article.

The published article failed to disclose that the statistical power of the study was insufficient to determine whether or not paroxetine (or imipramine) was actually superior to placebo in the "low" lithium level subgroup. The article failed to acknowledge that the study was not designed to test whether or not paroxetine (or imipramine) was superior to placebo in subjects with "high" or "low" lithium levels. To test this hypothesis, study subjects would have needed to be maintained in separate lithium level cohorts of "high" or "low" lithium level ranges throughout the study. In fact, the study merely stratified subjects into "high" or "low" lithium level ranges based upon a single baseline lithium level determination for statistical purposes. There were no discrete "high" or "low" lithium levels subgroups in the 352 study, as all study subjects were maintained (according to protocol) within a lithium level range of 0.4–1.2 mEq/L. We believe that it was the conflation of primary and secondary analyses that allowed GSK to present the lithium subgroup analyses as if they were distinct and clinically meaningful entities.

The article failed to inform the reader that the published power estimate of 35 subjects per group was not part of the original study design. Rather, in our opinion, the published power estimate was most likely contrived by GSK to comport with the final sample size of the paroxetine treatment group (i.e., 35 subjects). All of the academic authors named on the published article were also listed as clinical investigators on the trial. Thus, they should have been aware of the unconventional power estimate used in the published article, and that it differed from that of the study protocol. Authors need to exercise due diligence when reviewing a draft manuscript. Had they the opportunity to do so in this case, they should have been able to recognize that the final power estimate was substantially lower than that of the study protocol. Moreover, they would have recognized that the power estimate changed after the trial was completed. This suggests to us that the final change in power estimate was made by GSK without regulatory approval or oversight. If so, this change would likely represent a substantial departure from Good Clinical Practice Guidelines policies for the conduct of clinical trials in humans [19].

By setting forth the fact that there was a clinically meaningful distinction between "high" and "low" lithium level subgroups, the published article provided the reader with a false impression that patients with "low" lithium levels might be uniquely responsive to paroxetine. In addition, the article also asserts a high rate of sexual side effects with imipramine, while reporting almost no sexual side effects with paroxetine. This claim, even if true in the 352 study, is misleading and implies that paroxetine is likely to cause few, if any, sexual side effects in bipolar patients. To enhance this proposition, the article selectively cited only side effect literature on imipramine, while omitting any citation on sexual side effects with paroxetine.

Finally, in our opinion the published article is misleading in its assertion that there is no therapeutic advantage to using antidepressant therapy in bipolar depressed patients with "high" lithium levels. The 352 study was not designed to test this hypothesis and this conclusion does not appear to be supported by the data. Moreover, with the exception of the paroxetine 352 study, there are no other published studies reporting a lack of antidepressant efficacy in patients with "high" lithium levels. Conversely, the assertion that antidepressants may be more effective in patients with "low" lithium levels is potentially dangerous and inconsistent with most published practice guidelines for treating bipolar depression [20, 21]. In this regard, patients with "low" lithium levels would be at greater risk for developing antidepressant-induced mania and suicidal ideation [8–13]. By downplaying the well-known side effect profile of paroxetine and portraying it as being effective for bipolar depression without manic episodes, the published article was able to conflate successfully efficacy and side effect aims to favor paroxetine over imipramine. It is our

opinion that the article's recommendation that "patients with bipolar depression who maintain high serum lithium levels may not require additional antidepressant medications... [while] patients with low serum lithium levels or those who cannot tolerate high serum lithium levels may benefit from augmentation therapy with either paroxetine or imipramine" is speculative, misleading and is not supported by the study results.

4.2. The ghostwriting of paroxetine study 352

The paroxetine 352 study manuscript was ghostwritten by Scientific Therapeutic Information, Inc. (STI) with funds provided by GSK; however, neither STI, the ghostwriter, Sally Laden, nor GSK's role in the production of the manuscript was acknowledged in the published article. Prominent academic researchers (with financial ties to GSK) and GSK employees were designated by GSK as 'authors' on the manuscript. According to available evidence that was used in support of the ORI Complaint and subsequently disclosed in a follow up letter to ORI [22], the manuscript's authors were chosen by GSK in consultation with STI. This practice, and specifically STI's role in creating ghostwritten manuscripts, was criticized in an editorial in the *Journal of the American Medical Association* in relation to Merck & Co., Inc.'s effort to promote sales of Rofecoxib (Vioxx®) [23].

Many of the named authors on the published article had little or no direct involvement in the design, daily conduct, data analysis, or writing of the initial manuscript drafts. In fact, some of the authors were only selected for this role once the ghostwriters began to draft the manuscript from the final study report or a summary provided by GSK. It appears from the available evidence that GSK and STI had originally chosen Dr. Laszlo Gyulai, then Assistant Professor at the University of Pennsylvania, as the paper's first author [22]. However, Dr. Gyulai was subsequently removed from this position by GSK and replaced by two other authors who were assigned by GSK to the first and second positions on the paper [22]. The evidence also indicates that the final GSK-assigned authors on the published article never reviewed or even saw preliminary drafts of the paper, and only saw the final edited manuscript just prior to final acceptance by the *American Journal of Psychiatry* [22]. Conduct of that sort does not comport with the International Committee of Medical Journal Editors' standards of genuine authorship [24]. As we see it, GSK, STI, the named authors on the published article and, indeed, the *American Journal of Psychiatry* are all morally culpable in various degrees for complicity in publishing this article that we believe to be drug promotion disguised as science.

The 352 study invites comparison with another GSK-sponsored study (i.e., the paroxetine 329 study) in which paroxetine was compared to imipramine and placebo in children and adolescents with major depression [25]. Other than the *ad hoc* assignment of authors in the paroxetine 352 study, we see little difference in the manner in which both study reports were ghostwritten and published in the same year. In each case, the article failed to report all the relevant data resulting in a 'spin' in favor of paroxetine. A more detailed description of the ghostwriting of study 329 is now possible via documents de-classified in litigation [26], whereas in the case of the 352 study, GSK and STI have refused to release the available documents into the public domain on the basis of the alleged necessity of protecting trade secrets.

5. Conclusion

Corporate malfeasance in misrepresenting the results of clinical trials, especially where ghostwriting is involved, is of particular concern in the field of psychiatry where outcome measures of clinical trials

are more subjective and lend themselves to manipulation. We are convinced by our analysis that the GSK-sponsored paroxetine 352 study is one such instance of this practice, although prior examples have also come to light [27–30]. Because few industry-sponsored studies gain public scrutiny and even fewer are ever formally retracted [31], it is important to make these articles as transparent as possible in order to correct the scientific record and inform the medical community of potential harm.

Acknowledgments

The authors would like to thank Dr. Bernard Carroll and an anonymous reviewer for the journal for helpful suggestions to an earlier draft of this paper, and to Ronald Goldman and Skip Murgatroyd for legal review. The views expressed herein are those of the authors alone and not necessarily those of any other person, firm or entity.

Conflict of interest

Dr. Amsterdam was a clinical investigator at the University of Pennsylvania for paroxetine study 352 that was supported by a grant from GSK from 1994 to 1996. Dr. Amsterdam is currently not a member of any industry-sponsored advisory board or speaker's bureau, and has no financial interest in any pharmaceutical or medical device company. He has received legal support for his ORI Complaint from the law firm of Baum, Hedlund, Aristei & Goldman of Los Angeles, California.

Dr. McHenry is a member of Healthy Skepticism and research consultant for Baum, Hedlund, Aristei & Goldman.

References

[1] Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. PLoS Med. 2005;2(5):e138.
[2] Kesselheim AS, Avorn J. The role of litigation in defining drug risks. JAMA. 2007;297(3):308–31.
[3] Lenzer J. What can we learn from medical whistleblowers? PLoS Med. 2005;2(7):e209.
[4] Complaint of Scientific Misconduct against Dwight Evans, Laszlo Gyulai, Charles Nemeroff, Gary S. Sachs and Charles Bowden, http://psychrights.org/Research/Digest/Science4Sale/110708EthicsComplaintAgainstEvansGyualaiNemeroffSachsBowdenetal.pdf (Accessed May 2012).
[5] Nemeroff CB, Evans DL, Gyulai L, Sachs GS, Bowden CL, Gergel IP, Oakes R, Pitts CD. Double-blind, placebo-controlled comparison of imipramine and paroxetine in the treatment of bipolar depression. Am J Psychiatry. 2001;158:906–12.
[6] http://abcnews.go.com/images/Primetime/paxil_bpg.pdf (Accessed May 2012).
[7] Nemeroff CB, editor. Advancing the treatment of mood and anxiety disorders: The first 10 years' experience with paroxetine. Psychopharm Bul. 2003;37/1.
[8] Wehr TA, Sack DA, Rosenthal NE, Cowdry RW. Rapidcycling affective disorder: Contributing factors and treatment responses in 51 patients. Am J Psychiatry. 1988;145(2):179-84.
[9] Sachs GS, Lafer B, Stoll AL, Banov M, Tibault AB, Tohen M, Rosenbaum JF. A double-blind trial of bupropion versus desipramine for bipolar depression. J Clin Psychiatry. 1994;55:391-3.
[10] Leverich GS, Altshuler LL, Frye MA, Suppes T, McElroy SL, Keck PE Jr, Kupka RW, Denicoff KD, Nolen WA, Grunze H, Martinez MI, Post RM. Risk of switch in mood polarity to hypomania or mania in patients with bipolar depression during acute and continuation trials of venlafaxine, sertraline, and bupropion as adjuncts to mood stabilizers. Am J Psychiatry. 2006;163:232-9.
[11] Dunner DL. Sub-types of bipolar affective disorder with particular regard to bipolar II. Psychiat Developments. 1983;1:75-85.

[12] Goodwin FK, Jamison KR. Manic-Depressive Illness. Bipolar Disorders and Recurrent Depression. 2nd ed. New York: Oxford University Press; 2007.
[13] Coryell W, Scheftner W, Keller M, Endicott J, Maser J, Klerman GL. The enduring psychosocial consequences of mania and depression. Am J Psychiatry. 1993;150:720-7.
[14] http://www.gsk-clinicalstudyregister.com/result_detail.jsp?protocolId=29060%2F352&studyId=F1D83A94-2628-4C9C-83A5-11D00A2D30AC&compound=paroxetine) (Accessed April 2012).
[15] http://dida.library.ucsf.edu/ (Accessed May 2012).
[16] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56-62.
[17] Guy W, editor. ECDEU Assessment Manual for Psychopharmacology: Publication ADM 76-338, Washington, DC: US Department of Health, Education, and Welfare; 1976. pp. 218-22.
[18] Young RC, Biggs JT, Ziegler VE, et al. A rating scale for mania: Reliability, validity, and sensitivity. Brit J Psychiatry. 1978;133:429-35.
[19] US Code of Federal Regulations, Consolidated Good Clinical Practice (GCP) Guidelines, US DHHS, Food & Drug Administration Center for Drug Evaluation and Research (CDER), Washington, DC; April 1, 2002.
[20] American Psychiatric Association: Practice guidelines for 1994;151(Suppl):1-36.
[21] American Psychiatric Association Practice guidelines for the treatment of patients with bipolar disorder (revision). Am J Psychiatry. 2002;159(Suppl):1-50.
[22] http://blogs.nature.com/news/files/2012/06/Jay-Amsterdam-today-filed-a-24-page-complaint-with-the-Office-of-Research-Integrity-at-the-US-National-Institutes-of-Health.pdf (Accessed August 2012).
[23] DeAngelis CD, Fontanarosa PB. Impugning the integrity of medical science: The adverse effects of industry influence. JAMA. 2008;299/15:1833-5.
[24] http://www.icmje.org. (Accessed June 2012).
[25] Keller MB, Ryan ND, Strober M, Klein RG, Kutcher SP, Birmaher B, et al. Efficacy of paroxetine in the treatment of adolescent major depression: A randomized, controlled trial. J Am Acad Child Adolesc Psychiatry. 2001;40:762-72.
[26] McHenry L, Jureidini J. Industry-sponsored ghostwriting in clinical trial reporting: A case study, Account Res. 2008;15:152-67.
[27] Steinman MA, Bero LA, Chren MM, Landefeld CS. Narrative review: The promotion of gabapentin: An analysis of internal industry documents. Ann Inter Med. 2006;145:284-93.
[28] Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: A case study of industry documents from rofecoxib litigation. JAMA. 2008;299/15:1800-12.
[29] Jureidini J, McHenry L, Mansfield P. Clinical trials and drug promotion: selective reporting of study 329. Int J Risk Saf Med. 2008;20(1-2):73-81.
[30] Fugh-Berman AJ. The haunting of medical journals: How ghostwriting sold "HRT". PLoS Med. 2010;7(9):e1000335. doi:10.1371/journal.pmed.1000335.
[31] Newman M. The rules of retraction. BMJ. 2010;341:1246-8.