*Dear Ralph,*

*Thanks for your invitation for a major revision. Here is our point-to-point reply to the editor and peer reviewers.*

*We have edited our manuscript extensively according to the peer review comments. We have also made it shorter by nearly 2000 words, e.g. by deleting a section on informed consent in the Discussion. Three paragraphs were moved one page up from Results to Methods, and three paragraphs under 3.2.7 were moved half a page down.*

*We can see how Reviewer 1 thought we were biased. This happened as a result of 2 authors pulling slightly different ways and has now been settled.*

*We hope our manuscript is now acceptable.*

Reviewer #1: Manuscript Number: JRS-210034 Full Title: Restoring the two pivotal fluoxetine trials in children and adolescents with depression Reviewer number: 1 Reviewer's comments Reviewer's opinion: Reject

Summary: The authors have attempted to review two trials which they consider to be misreported. The overarching aim remains to discredit the use of fluoxetine as an antidepressant

*We have no such aim. We wrote: “Neither of the peer-reviewed publications, which had Graham Emslie as first author, [7,8] described the suicidal events that were mentioned in the clinical study reports Eli Lilly submitted to the drug regulators for marketing approval [9,10]. We therefore set out to review and restore the public record of the trials.” We have deleted: “which had Graham Emslie as first author, [7,8]”and have added, before the last sentence: “Furthermore, Lilly concluded that fluoxetine was effective, even though the FDA found that both studies were negative on their primary endpoint.”*

Major points:

1. It is worrisome to see the biased approach the authors pursue here and is unfortunate. In the abstract itself, they seem to lose their scientific focus and by the time they reach their conclusion, they have resorted to sensationalization. Scientific literature, as the authors will themselves know the best, is not a vindication.

*We cannot respond to a vague accusation of having been biased, except to agree our tone in the original could have been better. Our conclusion is not sensational, but evidence-based, as it follows from our findings. When drug companies have done placebo-controlled trials, it is standard that they conclude that their drug is safe and effective, no matter what the results were. In this case, Eli Lilly concluded that fluoxetine was safe and effective to minors, but even though the FDA approved the drug, it did not agree with this. We write: “The lack of efficacy in primary endpoints and significant safety hazards were noted by the FDA reviewer of the fluoxetine license application in 2002” and “FDA requested that Lilly conducted a one-year study of the effect of fluoxetine on growth, which the company declined to do.”*

*We found that fluoxetine was neither effective, nor safe, which is why we suggest turning the standard mantra on its head and write as our conclusion: “Fluoxetine is unsafe and ineffective.” We are aware that this can be seen as provocative, but the provocation is really that drug companies virtually always say that their drugs are safe and effective even when the results contradict this. We hope the editor agrees that our conclusion is appropriate. If not, we can change it of course.*

*It is clear from the following that this peer reviewer is very hostile. Furthermore, most comments lack substance we can respond to. We have nonetheless tried to respond, hoping the editor will find it useful. We are convinced that it will not be possible to persuade the reviewer that we did a good job with our paper. A back and forth between us and this reviewer would therefore be pointless. If the editor decides to submit our revision for peer review, we therefore kindly ask the editor not to use this person again.*

*One option could be to publish this peer review in full along with the paper and our comments. Another option is for the editor to add a commentary to the paper, making perhaps such points from this review that the editor believes have some merit and readers may wish to consider.*

2. Eli Lilly and Graham Emslie did not respond as to the authors' expectation, which seems to have further bolstered their misconception of evil hidden behind the trials and their statistics. This has made their approach bitter, reflecting the same in this manuscript.

*We demonstrate what was wrong with the trials, including the statistics. This is how researchers should operate. As noted above, we have changed the wording to make the tone neutral.*

3. The authors have used their skills and judgements to restore and review the trial data, and focus their wrath towards fluoxetine. Their narrative in this manuscript would not have changed if "fluoxetine" was replaced by an evil character from history or fiction. Fluoxetine is the villain, hiding many deaths. The authors have both misplaced and displaced their dissent. If at all anything is to be blamed, it should be how RCTs are conducted and how people can find fault in almost any RCT. We should accept that the data need not be perfect for the drug to "work". The ultimate aim is to help those with mental illnesses and psychological difficulties. This seems to be replaced by statistical analysis and data digging. This is a shame.

*We have dissected the evidence and have concluded accordingly.*

4. Fluoxetine has helped numerous individuals, and the black box warning by FDA on increased suicidal risk may itself be questioned. Again, this should be to understand more, and not to discredit. The authors are knowledgeable but seem to have missed literature like:

\* Fornaro M, Anastasia A, Valchera A, Carano A, Orsolini L, Vellante F, Rapini G, Olivieri L, Di Natale S, Perna G, Martinotti G, Di Giannantonio M, De Berardis D. The FDA "Black Box" Warning on Antidepressant Suicide Risk in Young Adults: More Harm Than Benefits? Front Psychiatry. 2019 May 3;10:294. doi: 10.3389/fpsyt.2019.00294. PMID: 31130881; PMCID: PMC6510161.

\* Beasley CM Jr, Dornseif BE, Bosomworth JC, Sayler ME, Rampey AH Jr, Heiligenstein JH, Thompson VL, Murphy DJ, Masica DN. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. BMJ. 1991 Sep 21;303(6804):685-92. doi: 10.1136/bmj.303.6804.685. Erratum in: BMJ 1991 Oct 19;303(6808):968. PMID: 1833012; PMCID: PMC1670974.

\* Gibbons RD, Brown CH, Hur K, Davis JM, Mann JJ. Suicidal Thoughts and Behavior With Antidepressant Treatment: Reanalysis of the Randomized Placebo-Controlled Studies of Fluoxetine and Venlafaxine. Arch Gen Psychiatry. 2012;69(6):580-587. doi:10.1001/archgenpsychiatry.2011.2048

\* Pompili M, Serafini G, Innamorati M, Ambrosi E, Giordano G, Girardi P, Tatarelli R, Lester D. Antidepressants and Suicide Risk: A Comprehensive Overview. Pharmaceuticals (Basel). 2010 Aug 30;3(9):2861-2883. doi: 10.3390/ph3092861. PMID: 27713380; PMCID: PMC4034101.

*The reviewer claims that, “Fluoxetine has helped numerous individuals.” This is not a scientific but a subjective statement. We found that fluoxetine does not work for children and adolescents, which agreed with their own perceptions of the effect.*

*The reviewer writes that, “the black box warning by FDA on increased suicidal risk may itself be questioned.” There is a huge amount of literature that questions the warning and claims that antidepressants protect minors against suicide. This literature is based on flawed observational research. The randomised trials found, indisputably, that the drugs double the risk of suicide. The warning is therefore highly appropriate. The four references the reviewer mentions are not relevant for us, as we prefer to keep focus on the science and to discuss the most reliable evidence in our paper.*

*Fornaro et al. did not do a systematic review but merely discussed the issues and mentioned two patients in their abstract: “… within the past decade, an increasing number of reports have questioned the actual validity of the FDA warning, especially considering a decline in the prescription of the antidepressant drugs associated with an increase in the rate of suicidal events among people with severe depression. The present report provides an overview of the FDA black box warning, also documenting two Major Depressive Disorder patients whose refusal to undergo a pharmacological antidepressant treatment possibly led to an increased risk for suicidal behaviors. The concerns raised by the FDA black box warning need to be considered in real-world clinical practice, stating the associated clinical and public health implications.”*

*Beasley was a Lilly employee who published a flawed meta-analysis in BMJ in 1991, which was before any studies of fluoxetine in children had been carried out. It is therefore irrelevant for our paper. The abstract states: “The pooled incidence of emergence of substantial suicidal ideation was 1.2% for fluoxetine, 2.6% for placebo, and 3.6% for tricyclic antidepressants.” Thus, placebo doubled the occurrence of suicidal ideation compared to fluoxetine. The fact is that these drugs double suicidality also in adults, as shown, for example, in a review based on FDA data on results from the placebo-controlled trials: Hengartner MP, Plöderl M. Newer-generation antidepressants and suicide risk in randomized controlled trials: a re-analysis of the FDA database. Psychother Psychosom 2019;88:247-8 and Hengartner MP, Plöderl M. Reply to the Letter to the Editor: “Newer-Generation Antidepressants and Suicide Risk: Thoughts on Hengartner and Plöderl’s ReAnalysis.” Psychother Psychosom 2019;88:373-4.*

*Gibbons have published multiple, highly flawed observational studies of suicidality. One of us (PG) wrote about this particular study in his 2015 psychiatry book: “Robert Gibbons and colleagues used individual patient data obtained from Eli Lilly and they claimed that fluoxetine didn’t increase suicide risk in children and adolescents with major depressive disorder,87 but this only illustrates that it is misleading to use too small samples. The FDA researchers found that fluoxetine increased the risk of suicide, suicide attempt or preparation for suicide in people under 25, relative risk 2.32 (95% CI 0.78 to 6.87) and when all drugs were included, this increased risk was the same but it was now statistically significant, relative risk 2.35 (95% CI 1.35 to 4.09).17”*

*The abstract for the 2010 review by Pompili et al. does not have any numerical information. It states that, “Most ecological studies and large clinical studies have found that a general reduction in suicide rates is significantly correlated with higher rates of prescribing modern antidepressants. However, ecological, cohort and case-control studies and data from brief, randomized, controlled trials in patients with acute affective disorders have found increases, particularly in young patients.” There are much better reviews than this one and we quote the most comprehensive one, which did not include published papers (which underreport suicidal events) but only clinical study reports. This is our own 2016 paper: Sharma T, Guski LS, Freund N, Gøtzsche PC. Suicidality and aggression during antidepressant treatment: systematic review and meta-analyses based on clinical study reports. BMJ 2016;352:i65. As we write in our paper for IJRSM, we found that, “the increase in the risk of suicide and violence is a class effect that is difficult to detect in two small trials, and we documented it. The odds ratios for children and adolescents were 2.39 (95% confidence interval 1.31 to 4.33) for suicidality.”*

5. I have absolutely no doubt that the authors will find missing data and statistical fault in these studies too, but I will have only regret if they do so.

*See just above.*

6. It is equally surprising and sad to see such learned authors write such a manuscript for an educational journal, which specifically evaluates the safety of drugs. There is always a balance that needs to be understood, or else, data digging and fault-finding can be done better by computers and artificially intelligent beings. I sincerely request the authors to understand what mental illness is, what psychiatric pharmacology is and what it claims to be. It is not an antibiotic with a narrow and large spectrum. Neither is mental illness an infection. Suicidality fluctuates in populations across the world, we have had suicidal thoughts ourselves too. It is not an all or none phenomenon, unlike fever.

*Our findings speak for themselves, and our many publications, also many in high-impact journals, and our books, illustrate that we are highly knowledgeable in the area we deal with in our paper.*

7. I understand that I have not used educational narrative in this review myself, and I do this deliberately, to reflect what the authors have done to seemingly every sentence that was published in these two trials. Looking back, I am glad that the authors were not the reviewers of these trials, because that would have resulted in humongous tragedy to depression and maybe a miniscule achievement in "how to find fault".

*There are no concrete issues we can respond to.*

8. As for those who criticise others' references, the authors' have done themselves poorly. For instance, Rosenbaum et al. (ref number 22) do not indicate that "The company already knew from a prior study that abrupt withdrawal of antidepressants could cause abstinence depressions in many patients".

*The reviewer misrepresents what we wrote: “The company already knew from a prior study that abrupt withdrawal of antidepressants could cause abstinence depressions in many patients [22].” We cite this prior Lilly study to show that Lilly knew about abstinence depressions from their own prior study. There is nothing wrong with our statement.*

9. The manuscript can only be published outside academic journals to continue the farce of what fluoxetine can do and cannot. The authors may find more acceptance in that, and further help in building resistance towards mental illness and acceptance of these by the society. I hope the authors find trials that show that antidepressants work. I hope they find peace. Maybe they need to understand they have not "restored" much, other than mistrust in those who read their publications.

*There are no concrete issues we can respond to.*

10. I shall give just one piece of data for them to ponder over, a self-citation rate of about 30% in this manuscript. They should reflect on why is this high level of self-citation required. Is their work benefiting others, or only themselves?

*What we cite is relevant for our paper, and some of our own work is unique, which means no one else have done the same. We cite a few of our books to avoid making the reference list excessive.*

Reviewer #2: This manuscript presents very important work regarding the safety of fluoxetine in children and adolescents and its findings should be disseminated to the scientific/ medical community. However, as it stands, this manuscript needs significant revisions in order to communicate its findings more effectively in an unbiased, systematic and thorough manner. It may be useful to engage an external party to critically review your writing, I tried to write some detailed comments for consideration.

*We agree the tone of the original was not as neutral as it might have been, and we have improved on this in our revision. We have tried to be as unbiased, systematic, and thorough as we could, and our paper is very detailed, which allows the readers to judge for themselves whether we succeeded.*

This manuscript is not strictly a trial restoration. The authors have conducted a review of contrasting reports between the CSR and publications. It remains unclear what parts are the CSR/authors/ new authors. This may be perceived as a bias. The parts of the manuscript that are presented as a critical appraisal should be kept separate to the analysis.

*We agree and have clarified in the revision where each element has come from.*

While I have prepared a number of comments that should be addressed, I think this work is critical to the field of child psychiatry and this paper should certainly be published when it is refined. I hope that these comments are helpful rather than discouraging to the authors.

\*General comments\*

1. A thorough review of the manuscript's language in general needs to be conducted, including grammar, the use of emotive language, imprecise statements and unscientific language.

*We have done that in our revision.*

2. The format of reporting does not follow appropriate conventions and is hard to follow (particularly in the results section). I would suggest the use of CONSORT/RIAT guidelines for reporting trial reanalyses.

*We carefully considered this when we did our work, and one of us (PG) has co-authored the CONSORT guidelines right from the beginning, back to the first meeting in 1993 in Ottawa, and other guidelines such as PRISMA for systematic reviews and meta-analyses and STROBE for observational studies. We found it advantageous, also for readers, not to be bound by guidelines. CONSORT is about reporting randomised trials, but we did not report a trial we had conducted. We provided a forensic analysis of two clinical study reports and compared them to what has been published.*

\*Specific comments\* (these are not exhaustive, but may be useful as examples to illustrate points made by general comments above)

3. Abstract, objective- specifics need to be added. What framework did you use for reanalysis? State your specific aims, e.g. benefit/ harms/ change in depression score… to reassure/ reassess safety? Mentions of trial misreporting should occur in the background and they should be specified. For the results, was information missing from both trial CSRs? The conclusion needs more detail here.

*The word limit for abstracts in the journal prevented us from going into so much detail.*

4. Introduction: first paragraph, explain the context of the trials/ where reports came from so that people can follow up; second paragraph, second sentence about adults needs a reference; second paragraph, last sentence CI reporting needs to be fixed.

*We have made this clearer, as suggested. We have moved the mentioning of the two clinical study reports to the third paragraph. We wrote: “An increase in suicidal events is also seen in adults. A meta-analysis of placebo-controlled trials in healthy volunteers using precursor events defined by the FDA found that SSRIs and SNRIs double the risk of harms related to suicidality and violence, and the number needed to treat to harm one healthy person was only 16 (95% confidence interval 8 to 100) [2].” The second paragraph already had a reference. The reference to adults comes in the next sentence. We have made that clear by inserting “adult” before volunteers. We cannot see anything needs to be fixed in relation to the confidence interval we give, which is correct.*

5. Methods: was Emslie PI for both trials?

*Yes, and first author.*

6. Methods Second paragraph, last sentence 'paid attention to' is not scientific language/ not specific, same with 'dropped a plan'.

*“Paid attention to” changed to “focused on.” We cannot see anything wrong with writing: “We dropped a plan to perform meta-analyses based on the quality of the efficacy data” but have improved on the wording: “We did not perform meta-analyses because the efficacy data were biased.”*

7. Methods Third paragraph last sentence why was it not necessary to present terms to a third party?

*We wrote: “We had planned that PCG should resolve any ambiguities with a third investigator before presenting the terms to DH, but this was not necessary.” Terms used by doctors can be imprecise and ambiguous, but we did not run into such difficulties. We have therefore deleted this sentence.*

8. What did you use fisher's tests for?

*It can only be used for proportions, and this is what we did. We write: “We used Fisher’s exact test for proportions.”*

9. Overall, methods need to frame other aspects of your reanalysis like framework/ grounds for critical review and how your results are reported/ structured seeing as it's not conventional (also see feedback below), for example: trial results are presented separately regarding each CSR. Our critical appraisal of each study's protocol and whether it was conducted by the trial is summarised… etc

*We have tried to be more concise.*

\*Results\*

10. The results section needs to be restructured and refined significantly. It is hard to follow which parts are critiques of the protocol/ CSR/ literature versus your actual findings. I suggest keeping these separate (although it is tempting to present them together as they relate to the same issue, this should be saved for the discussion).

*We have rewritten our paper numerous times and what we did cannot fit one of the usual templates for reporting. We have focused on being pedagogic, so that the readers can follow our findings and arguments, some of which need to be in the Results section; otherwise, we would lose the readers.*

11. S3.1 reference for the fact that Lilly's analysis plan was written a posteriori

*We wrote: “This was a single-centre investigator-initiated trial conducted in Texas from 10 April 1991 to 28 February 1995 and published in 1997 [7]. The study report from 2000 contains Emslie’s original protocol and a protocol revised by Lilly [9]. Lilly’s statistical analysis plan was written a posteriori.” Lilly had nothing to do with the trial until after it had been published, apart from delivering blinded trial drugs mid-trial.*

12. 3.1.2 third paragraph regarding origin of trial drugs- what were the different origins? The pharmacy's Prozac/ lactose vs lilly's pills?

*We explained this on the previous page: “The pharmacy initially blinded the drugs by emptying Prozac capsules for the placebo group and refilling them with lactose powder [9]. From August 1993, Lilly supplied blinded drugs in white capsules.”*

13. 3.1.4 first sentence unclear, not sure how this favours fluoxetine? Could you explain this? Is it based on their analysis

*Good point. We have changed the wording into: “All efficacy analyses were biased in favour of fluoxetine.”*

14. Page 6 second sentence Use of unscientific language- 'larger'

*It is not unscientific but factual to write: “After 8 weeks, the difference from baseline in CDRS-R was 9.7 larger on fluoxetine than on placebo using the LOCF method.” This is the easiest way to describe a difference between two differences.*

15. Page 6 paragraph 2 first sentence needs reference

*We already gave the reference, in the previous sentence (ref. 9; it is clear that we talk about the clinical study report here).*

16. Table 1 it's unclear which are yours/ lilly's p-values in the table

*This was clear from the table header: “Table 1. Discrepancies between results reported in tables in the published report [7] and in the study report [9]. P-values for published data calculated by us based on final scores; other P-values calculated by Lilly based on changes from baseline.” But we have improved on the table header.*

17. Page 6 text under table 1, 'lumped' is unscientific

*In meta-analysis, it is common to talk about lumping or splitting data. But we have changed the word into “combined.”*

18. Page 7, second paragraph- no significant differences in what?

*This comes in the following sentences.*

19. 3.1.5 discontinuations should be a separate section

*The section is this one: “3.1.5. Serious adverse events and discontinuations.” These are best discussed together because they are closely related and because Lilly combined them in a most confusing way. We write, for example: “The information on the patient with aggressive behaviour appeared in a table, which mixed nonrandomised patients with patients with serious adverse events and patients withdrawn due to nonserious adverse events.”*

20. Table 2 why are discontinuations presented weekly? What does this add? Maybe just present an overall total

*This is because these data are highly important. We wrote: “The information on discontinuations and suicide attempts was inconsistent. In the published report, [7] 36 patients discontinued, in the study report, 38 [9]. The numbers, reasons, and timing were the same in only 5 of 28 cells with information (table 2). There was no explanation for these discrepancies [9].” This discrepancy is huge and therefore important for readers to know about.*

21. Page 7, text under table 2. Are the adverse events predisposing a suicide attempt your conclusions/ descriptions or the CSRs or Emslie's? There needs to be more clear delineation between your interpretation of their findings and your findings from the CSR.

*We wrote: “Adverse events definitely or possibly predisposing to suicide in this patient were manic reaction, insomnia, and nervousness (likely akathisia).” It is only us who use “Adverse events definitely or possibly predisposing to suicide,” so these are clearly our judgments. But we have made it clearer by adding “according to our criteria.”*

22. Descriptions of adverse events appear as a list. These would be better presented as a summary and/ or tabulated or added to the appendix (if you are including lots of detail).

*We provide lists of adverse events in our supplements to show to the readers how we classified them as definitely or possibly predisposing to suicide. These lists are very important. We do not have any list in the main text.*

23. In some cases, the use of emotive language can appear biased/ accusatory- e.g. Emslie downplaying suicide attempts. It makes more of an impact to state what they said and explain why it was wrong or at least qualify the language

*We agree and have modified the text.*

24. There are parts of the results that either belong in a critical appraisal section of the results or should be moved to the discussion for extrapolation, e.g. terms like 'which suggests Lilly's analysis… page 8 paragraph 3.

*See our reply to the reviewer under item 10.*

25. 3.1.6 is this your p-value? What test did you use? Is it appropriate for the sample size.

*There are many P-values in this section, so, we don’t know which one the reviewer refers to. But we made it clear that we used Fisher’s exact test for our own values, and it is also clear which P-values came from us.*

26. As far as you can ascertain, did trial HCJE use an active or inactive placebo?

*They are all inactive. We have not stated this because companies very rarely use active placebos, and if they did, they would surely mention it.*

27. Did either study report on compliance? How was it measured?

*Not consistently, no.*

28. The reporting of each trial should be consistent. There is more detail about HCJE's statistics. I understand the level of detail may be different between the reports but please clarify. Perhaps this will become more clear once the manuscript is refined/ restructured

*The HCJE is bigger than X065; it ran for 19 weeks as compared to 8 weeks, and the report is much longer, with a lot of relevant detail, which is why we provided more detail.*

I have refrained from collating additional detailed comments about the results thus far, but am happy to provide more detailed comment -IF the major issues from (10) are addressed.

*See our reply to item 10.*

\*Discussion\*

I am hesitant to make detailed comment on conclusions made by the discussion, pending review of the results being presented- but have made some general points:

29. The section on comments from the results could accommodate concerns I stated in points 10, 21,24

*See our reply to item 10.*

30. Comments on the results- important arguments are made here by the authors. However, the flow is disjointed, they might consider grouping common themes of their reflections and/ or adding linking sentences. As it stands, this reads like a series of dot points critiquing the CSRs/ Lilly. This section should focus on the authors' findings, with other criticisms moved to a separate/ new section?

*We have transformed the Discussion now with the removal of certain sections. We had already divided the Discussion into “4.1. Comments on the results” and “4.1. Comments on the context.”*

31. Comments on the context- fix the subheading numbering. Presents a good summary of the history of fluoxetine in children which further highlights the importance of this work to the reader.

*We do not have any subheadings in this section.*

32. Again, the use of emotional/ alarmist language should be minimised. As a scientific paper, use of terms like 'public health emergency' are extreme. The use of such language suggests a certain level of bias from authors/ readers may lose confidence in the quality of their findings.

*We agree and have changed the language.*

\*Supplementary tables\*

Current format is difficult to follow. There are no comprehensive legends, not all have table numbers, for some tables 'having trouble with' could probably be omitted (the format of the checklist can instead be described in the figure legend), a description of 'non-solicited' should be provided. How does table 'fluoxetine side effects checklist' differ from the first table (I can see that it is based on the CSR formatting but dividing it like this is not helpful). The authors should consider readability guidelines for presenting tables and figures.

*These tables will not be seen by many people. If they read our paper first, which of course they should, they should have no problem with understanding them. The table headers are: “Supplementary table 1. Number of adverse events predisposing to violence against self or others” and “Supplementary table 2. Number of adverse events that were or could be extrapyramidal drug harms and might predispose to violence against self or others.”*

Reviewer #3: A window with the view on the details of how drug approval proceeds. Very informative and helpful for understanding mechanisms of drugs (and vaccines!!) launching on the market.

There is a lot of detailed information on drugs trials, which makes it a rather long article. Please consider shortening some section.

*We have done that, thank you.*