PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How do authors of systematic reviews deal with research malpractice and misconduct in original studies? A cross-sectional analysis of systematic reviews and survey of their authors
AUTHORS	ELIA, Nadia; von Elm, Erik; Chatagner, Alexandra; Pöpping, Daniel; Tramer, Martin

VERSION 1 - REVIEW

REVIEWER	Serina Stretton ProScribe - Envision Pharma Group, Sydney, Australia I am currently paid as an employee and work on a voluntary basis for commercial and non-commercial organisations that provide education and services related to the ethical and effective preparation of publications.
REVIEW RETURNED	30-Nov-2015

GENERAL COMMENTS	I congratulate the authors on tackling a topical subject that is likely to be of great interest to readers. The paper presents predominantly descriptive data on whether authors of systematic reviews report on factors that may be considered malpractice or misconduct. This is a very interesting study that highlights an important facet of systematic reviews – what should systematic review investigators do when misconduct is suspected and how or should this be handled in a public forum? For this review I have taken the previous peer-reviewers' comments into account and focussed primarily on methodology. The research question and study objective is clearly defined and the overall study design is appropriate for the research question. In the methods, the authors state that the study is a cross-sectional analysis of all systematic reviews published in the four journals selected. However, the search strategy applied may not retrieve all systematic reviews. Although identification of reviews as systematic in the title is a PRISMA requirement, not all systematic reviews may comply with this guideline and be identified as such in the title (eg, Morgan et al 2013, Ann Intern Med, 158:329 - Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies). Inclusion of the abstract field in this search string may have identified additional studies. The authors do not describe how randomization was achieved in the identification of the studies from the Cochrane Library. This is an important point for the reader and is needed to provide support for the veracity of the comparison between Cochrane and the four journals. Although there is a word count limit of 4000 words, inclusion of "(computer-generated sequence)" would only require

	two words with minimal editing to meet the required word count limit. The authors rightly state that the lack of an accepted definition of research misconduct is a limitation of this study. However, the lack of a definition in the manuscript is confusing. For example, plagiarism as misconduct is not consistently described throughout the manuscript. In the introduction, the authors infer that plagiarism is considered as “malpractice” (last line of paragraph 2) whereas plagiarism is grouped with misconduct in the results and discussion sections. Given that plagiarism is considered by the US Office of Integrity as research misconduct (http://ori.hhs.gov/definition-misconduct), and can range from a few sentences in the methods to egregious copy/paste or copying of others’ data or ideas, this distinction should be made more clear. A major finding is the identification of duplicate publication as a key concern in systematic reviews. However, duplicate publication does not seem to be considered misconduct by the authors. As noted by the US Office of Integrity, duplicate publication is one of the most serious forms of plagiarism and is considered as research misconduct (http://ori.hhs.gov/plagiarism-14). Although the authors note in the results that their definition of duplicate publication changed during the study, a clear definition should be provided in the methods. The authors should also take into consideration how they handled legitimate publications that may report a secondary objective and, therefore, report new insights or an extension of the primary analysis. In the second paragraph the authors suggest that reviewers who do not report having searched for duplicate publications do not consider duplicate publication worthwhile to disclose. Systematic reviews with robust search strategies would uncover most duplicate publications during the retrieval phase and perhaps it is more likely that reviewers are not sure on how best to address duplicate publications that are suspected cases of misconduct. The last two paragraphs of the review are an opportunity for the authors to recommend what action authors and editors can take, based on the findings presented in this study. While I am in agreement that systematic review investigators are in a unique position to identify potential cases of malpractice or misconduct, a clear pathway for how these cases should be handled needs to be developed – trial by publication may not be the most robust method for “whistle-blowing”. It may be more appropriate present the information on previous systematic reviews, then the information on PRISMA, and then merge the second half of the first paragraph in this section with the second half of second paragraph to present the authors’ concluding points and recommendations for authors and editors more clearly.

REVIEWER	Xavier Bosch Department of Internal Medicine, Hospital Clínic, University of Barcelona, Spain
REVIEW RETURNED	04-Dec-2015

GENERAL COMMENTS	I agree in general with that the article is worth publishing: it is scientifically sound and original. I also see that the authors have responded satisfactorily to the concerns raised by the BMJ reviewers. However, I think the authors should briefly elaborate on a critical point raised by one of the BMJ reviewers. In particular: "The part that interested me the most was where the authors of the systematic reviews had suspected intentional fraud but done little or

	nothing..." What interests to me, and fully agree with, is the proposal of the reviewer on how authors should act in front of this: "But it also means that the reviewers run the risk of libel, which they probably didn't appreciate. The right thing to do would be to contact the head of the institution where the research was done in that employers have the legal legitimacy (locus standi) to investigate and can ensure due process. As far as I can see, none of the reviewers took this step, and ironically I could argue that to suspect serious misconduct and to take no action is itself misconduct." The authors should build on this fruitful comment as a possible way to move forward in case of blatant misconduct, perhaps meaning eventual retraction and correction of the literature. I would add that, if the systematic reviewer is not willing or simply does not have the time to contact the institution of the relevant author, or the institution keeps silent, he/she can try to contact the journal where the paper was originally published. Plus I would also report the suspicion to the editor of the journal where the systematic reviewer is going to submit his/her systematic review.

REVIEWER	Doug Altman University of Oxford UK
REVIEW RETURNED	14-Dec-2015

GENERAL COMMENTS	General points An important and interesting topic. This study has strengths but also some weaknesses. I welcome the study despite the reservations on details of the methods. The findings add to a still quite slim knowledge base about the actual conduct of systematic reviews. Clearly there is a need for clearer guidance for future reviews. The text is rather cavalier with language – I give several examples of (presumably) unintentional confusion of key concepts. The text would benefit from careful editing. Specific points Intro and methods 1. P7/8: I'm surprised that the 6 listed procedures don't include searching for a protocol, especially as this is an item in the PRISMA Statement. Also it's directly relevant to the issue of selective outcome reporting. On p8 we rad that protocols were available for 17 non-Cochrane reviews but not how they were obtained. 2. Conversely I'm unclear why the authors sought information about whether the reviewers sought information on the ethical approval for the primary studies. I don't recall this being highlighted before or is the rationale for this item mentioned in the opening remarks. Finally the issue is discussed on p14! While of course I believe that ethics approval and informed consent are essential for RCTs I don't see their direct relevance to the question being investigated. While I know that there is a high profile fraud case in anaesthesiology reporting that ethics approval or informed consent were obtained doesn't mean they actually were. 3. 7/25: how do the authors define misconduct? There is a sort of definition under "Sample size" but this mentions only serious misconduct. The issue should be addressed in the main text. The

	brief comments on p15 are not really adequate. 4. I don't generally expect a statistical sample size calculation for such a study but once is provided. However I don't find it especially convincing. One reason is that the calculation relates only to falsification or fabrication (i.e. serious misconduct), whereas the study has a much broader remit. Is Fanelli's very wide suggested range of prevalence of estimated misconduct (2 to 14%) based mainly on RCTs? If not how relevant is it? Which value was actually used in the calculation? 5. Why those 4 journals? I don't understand how sampling the whole of a calendar year chimes with a prespecified sample size target. 6. The Methods section says nothing about statistical analysis. Specifically it doesn't mention performing significance tests, nor which methods were used. In particular, I'm concerned about tests to compare journals. I don't see a good justification for such analyses, as they relate neither to the study objectives nor do they have good power. Also the actual numbers in many of the cells are too small for reliable analysis. 7. I have reservations about the ad hoc composite score shown in table 1. Such composites are known to be problematic in many areas e.g. quality scores. Results and Discussion 8. It's strange to take a random sample of 25 Cochrane reviews and yet end up with only 19 of these. Why not ensure you had 25 eligible reviews? 9. How many reviews were reviews of trials, and how many of other study types? 10. The probability of suspecting something will be related to the number of studies in the review. Yet this information isn't provided. (By contrast we're told how many authors there were which seems far less relevant.) How large were the reviews in which some inappropriate practices were suspected? 11. P11: I can't match the numbers in the paragraph on publication bias to Table 3. Minor points 12. Abstract: Objective is ambiguous – "To study whether authors of systematic reviews apply procedures to counter-balance research malpractice such as not publishing completed research, duplicate publications, or selective reporting of outcomes, and whether they identify and report misconduct." Need to distinguish bad practices from counterbalancing procedures 13. Publication bias is not the same thing as unpublished studies. Nor is identifying bias the same as reducing it (p14). 14. Nor is possible misconduct the same as misconduct (15/3). Nor is misconduct the same as fraud (p15). 15. It is stated that PRISMA-P "recommends reporting of publication bias and selective outcome reporting" – i.e. in a review protocol. This doesn't make sense and it's not true. What PRISMA-P says (item 16) is: "specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies)" which is very different indeed.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 (Serina Stretton)

I congratulate the authors on tackling a topical subject that is likely to be of great interest to readers.

The paper presents predominantly descriptive data on whether authors of systematic reviews report on factors that may be considered malpractice or misconduct. This is a very interesting study that highlights an important facet of systematic reviews – what should systematic review investigators do when misconduct is suspected and how or should this be handled in a public forum?
For this review I have taken the previous peer-reviewers' comments into account and focussed primarily on methodology.

The research question and study objective is clearly defined and the overall study design is appropriate for the research question.

1)In the methods, the authors state that the study is a cross-sectional analysis of all systematic reviews published in the four journals selected. However, the search strategy applied may not retrieve all systematic reviews. Although identification of reviews as systematic in the title is a PRISMA requirement, not all systematic reviews may comply with this guideline and be identified as such in the title (eg, Morgan et al 2013, Ann Intern Med, 158:329 - Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies). Inclusion of the abstract field in this search string may have identified additional studies.

Answer: We agree with this comment. Our search strategy may have missed a couple of reviews that were not reported as "systematic reviews" in their title. We do not think that including these additional reviews would have drastically changed our message. However, for purpose of clarity, we have removed the term "all" from the description of the study design, which now reads (page 6, "study design"): "We conducted a cross-sectional analysis of all systematic reviews published in 2013 in four general medical journals (Annals of Internal Medicine (Ann Int Med), The British Medical Journal (BMJ), Journal of the American Medical Association (JAMA), and The Lancet (Lancet))." Furthermore, we discuss this issue as limitation (page 15, last paragraph). "Systematic reviews that were not identified as such in their titles were not included. However including these reviews would not have changed our findings".

2)The authors do not describe how randomization was achieved in the identification of the studies from the Cochrane Library. This is an important point for the reader and is needed to provide support for the veracity of the comparison between Cochrane and the four journals. Although there is a word count limit of 4000 words, inclusion of "(computer-generated sequence)" would only require two words with minimal editing to meet the required word count limit.

Answer: We agree. We have added "computer generated sequence" in the sentence, which now reads (page 6, "selection of systematic reviews"): "A computer-generated random sequence was used to select 25 reviews published in 2013 in the Cochrane Library"
For the interested reader, the detailed procedure was as follow: a random sample was drawn by 1) taking the list of the Cochrane reviews newly published in 2013, 2) assigning a computer generated random number to each of the reviews 3) re-arranging the reviews according to increasing random numbers 4) selecting the 25 smaller ones. In order to remain below 4000 words as requested, we did not add this detailed information in the manuscript.

3)The authors rightly state that the lack of an accepted definition of research misconduct is a limitation of this study. However, the lack of a definition in the manuscript is confusing. For example, plagiarism as misconduct is not consistently described throughout the manuscript. In the introduction, the authors infer that plagiarism is considered as "malpractice" (last line of paragraph 2) whereas plagiarism is grouped with misconduct in the results and discussion sections. Given that plagiarism is considered by the US Office of Integrity as research misconduct (http://ori.hhs.gov/definition-misconduct), and can range from a few sentences in the methods to egregious copy/paste or copying of others' data or ideas, this distinction should be made more clear.

Answer: We agree. We have clarified our definition of malpractice and of misconducts in the introduction, which now reads:
(page 5, paragraph 1): “Some of these practices, however, are very common and may not be regarded by all as “misconducts”. Therefore, this research defines “malpractice” as relatively common and minor misconducts, while the term “misconduct” is used for data fabrication, falsification, plagiarism, or any other intentional malpractices.”
(page 5, paragraph 2): “The impact of other malpractices, such as gift or ghost authorship or plagiarism,, is less clear.”

4) A major finding is the identification of duplicate publication as a key concern in systematic reviews. However, duplicate publication does not seem to be considered misconduct by the authors. As noted by the US Office of Integrity, duplicate publication is one of the most serious forms of plagiarism and is considered as research misconduct (http://ori.hhs.gov/plagiarism-14).
Although the authors note in the results that their definition of duplicate publication changed during the study, a clear definition should be provided in the methods.

Answer: We agree that duplicate publication is misconduct. Our definition of duplicate publication has not changed during the study. However, we realized that the authors of systematic reviews are using very diverse terms to describe “true” duplicate publications. Also, authors of systematic reviews are often using the term “duplicate” or “duplicate publication” to desribe a reference that they have found several times in different databases. To answer the reviewer’s concern, we have added a clear definition of what we considered duplicate publication in the method section which now reads (page 7, “variables”, 2nd paragraph): “3. Search for duplicate publications (defined as a redundant re-publication of an already published trial, with or without a cross-reference to the original article).”

4.1) The authors should also take into consideration how they handled legitimate publications that may report a secondary objective and, therefore, report new insights or an extension of the primary analysis.

Answer: This remained unclearly described in most systematic reviews. Therefore, we have included a question regarding duplicate publications when contacting the reviewers. We asked them, for example, whether the reported number of “companion articles” were really duplicate publications in the sense of redundant re-publication of an already published study or trial, with or without cross-reference to the initial publication. The problem of the different terms used to report duplicate publication is highlighted in the discussion (page 14, “comparison with similar analyses”, end of 1rst paragraph): “Finally, there is no widely accepted definition of the term “duplicate”, which in turns, adds to the confusion. For example, a number of reviewers used the term “duplicate” to describe references identified more than once through the search process.”

5) In the second paragraph the authors suggest that reviewers who do not report having searched for duplicate publications do not consider duplicate publication worthwhile to disclose. Systematic reviews with robust search strategies would uncover most duplicate publications during the retrieval phase and perhaps it is more likely that reviewers are not sure on how best to address duplicate publications that are suspected cases of misconduct.

Answer: We agree and have changed the sentence accordingly (page 14, “comparison with similar analyses”, 1rst paragraph): “It remains unclear whether reviewers do not consider duplicate publication worth disclosing or whether they are unsure on how to address the issue.”

6) The last two paragraphs of the review are an opportunity for the authors to recommend what action

authors and editors can take, based on the findings presented in this study. While I am in agreement that systematic review investigators are in a unique position to identify potential cases of malpractice or misconduct, a clear pathway for how these cases should be handled needs to be developed – trial by publication may not be the most robust method for “whistle-blowing”. It may be more appropriate present the information on previous systematic reviews, then the information on PRISMA, and then merge the second half of the first paragraph in this section with the second half of second paragraph to present the authors’ concluding points and recommendations for authors and editors more clearly.

Answer: Thank you for this suggestion. We have re-ordered the last two paragraphs.

Reviewer: 2 (Xavier Bosch)

I agree in general with that the article is worth publishing: it is scientifically sound and original. I also see that the authors have responded satisfactorily to the concerns raised by the BMJ reviewers. However, I think the authors should briefly elaborate on a critical point raised by one of the BMJ reviewers. In particular: "The part that interested me the most was where the authors of the systematic reviews had suspected intentional fraud but done little or nothing..."

What interests to me, and fully agree with, is the proposal of the reviewer on how authors should act in front of this: "But it also means that the reviewers run the risk of libel, which they probably didn’t appreciate. The right thing to do would be to contact the head of the institution where the research was done in that employers have the legal legitimacy (locus standi) to investigate and can ensure due process. As far as I can see, none of the reviewers took this step, and ironically I could argue that to suspect serious misconduct and to take no action is itself misconduct."

The authors should build on this fruitful comment as a possible way to move forward in case of blatant misconduct, perhaps meaning eventual retraction and correction of the literature. I would add that, if the systematic reviewer is not willing or simply does not have the time to contact the institution of the relevant author, or the institution keeps silent, he/she can try to contact the journal where the paper was originally published. Plus I would also report the suspicion to the editor of the journal where the systematic reviewer is going to submit his/her systematic review.

Answer: We agree. We have added the following statement to our conclusion (page 17, last paragraph): “The need for explicit guidelines on what reviewers should do once misconduct has been suspected or identified has already been highlighted.18 These guidelines remain to be defined and implemented. The proper procedure requires the reviewer to request the institution where the research was conducted to investigate on the suspected misconducts, as the institution holds the legal legitimacy. Whether alternative procedures could be applied, such as, for example contact the editor in chief of the journal where the suspected paper was originally published, or inform the editor in chief where the systematic review will eventually be published, should be discussed. “

Reviewer: 3 (Doug Altman)

General points
An important and interesting topic. This study has strengths but also some weaknesses.
I welcome the study despite the reservations on details of the methods. The findings add to a still quite slim knowledge base about the actual conduct of systematic reviews. Clearly there is a need for clearer guidance for future reviews.
The text is rather cavalier with language – I give several examples of (presumably) unintentional confusion of key concepts. The text would benefit from careful editing.

Answer: We would like to apologize for this. The Manuscript has been completely revised with the help of an English-native person and we do hope that it is now much improved.

Specific points
Intro and methods
1. P7/8: I'm surprised that the 6 listed procedures don't include searching for a protocol, especially as this is an item in the PRISMA Statement. Also it's directly relevant to the issue of selective outcome reporting.

Answer: The PRISMA STATEMENT requires that an "ideal systematic review" should have a protocol registered, but, to our knowledge, it does not clearly require that the protocols of all included studies should be searched. However, we understand that our wording may have brought some confusion and have therefore have deleted the term "ideal systematic reviews" and re-phrased the sentence which now reads (page 7, "Variables", 2nd paragraph): "Furthermore, we examined whether each of the selected reviews applied the following six procedures:"

1.1 On p8 we rad that protocols were available for 17 non-Cochrane reviews but not how they were obtained.

Answer: For each systematic review, we checked whether a protocol for the review had been registered, and we used the registration number reported in the review to access the protocol.

2. Conversely I'm unclear why the authors sought information about whether the reviewers sought information on the ethical approval for the primary studies. I don't recall this being highlighted before or is the rationale for this item mentioned in the opening remarks. Finally the issue is discussed on p14!
While of course I believe that ethics approval and informed consent are essential for RCTs I don't see their direct relevance to the question being investigated. While I know that there is a high profile fraud case in anaesthesiology reporting that ethics approval or informed consent were obtained doesn't mean they actually were.

Answer: Weingarten (2004) and Vergnes (2010) have both highlighted the role that systematic reviewers could play by identifying "unethical research". Both of these references are given in the introduction.
The question regarding which study design requires approval by an ethics committee has no universal answer, and depends on the cultures and countries where the research is performed. In some countries, only experimental designs require ethical approval, while in some others, even retrospective analyses of observational data need to obtain ethical approval (or waiver of).
We have now added a description of the study designs included in each reviews in table 2, and in the Appendix table 1A and 2A. There were only three reviews, which included exclusively study designs that obviously did not require approval by an ethics committee. We have therefore corrected the scoring for these reviews in Appendix Table 1A.

3. 7/25: how do the authors define misconduct? There is a sort of definition under "Sample size" but this mentions only serious misconduct. The issue should be addressed in the main text. The brief comments on p15 are not really adequate.

Answer: We have improved our definition of malpractice and of misconduct in the introduction (page 5, 1rst paragraph): "While a common definition of research misconduct is still lacking, there is an urgent need to come up with strategies to prevent it. Fifteen years ago, Richard Smith proposed "A preliminary taxonomy of research misconduct" describing 15 practices ranging from "minor" to "major"

misconducts.6 Some of these practices, however, are very common and may not be regarded by all as “misconducts”. Therefore, this research defines “malpractice” as relatively common and minor misconducts, while the term “misconduct” is used for data fabrication, falsification, plagiarism, or any other intentional malpractices.”

4. I don’t generally expect a statistical sample size calculation for such a study but once is provided. However I don’t find it especially convincing. One reason is that the calculation relates only to falsification or fabrication (i.e. serious misconduct), whereas the study has a much broader remit. Is Fanelli’s very wide suggested range of prevalence of estimated misconduct (2 to 14%) based mainly on RCTs? If not how relevant is it?

Answer: We agree that Fanelli’s article was not the correct reference here and have deleted it. We have rephrased the paragraph, which now reads (page 8, “sample size”): “The capacity of systematic reviewers to identify misconduct is unknown. It has been suggested that serious misconduct (data falsification or fabrication) affects 2 to 14% of original articles.20 Our hypothesis was that 5% of systematic reviewers would identify misconduct. Therefore, we needed a minimum of 110 systematic reviews to allow us to detect a prevalence of 5%, if it existed, with a margin of error of 4% assuming an alpha-error of 0.05.”

4.1 Which value was actually used in the calculation?

Answer: We used an estimated proportion of 5% of reviews that identified misconduct, with a confidence level of 95% and a desired precision of 4% for an infinite population using the formula: N=(Z2 X P(P1-P))/E2.

5. Why those 4 journals? I don’t understand how sampling the whole of a calendar year chimes with a prespecified sample size target.

Answer: Our sample size calculation suggested that approximately 115 reviews would enable us to identify 5% of reviews suspecting misconduct. We aimed to include systematic reviews published in high impact journals in the hope that they would be of high quality and adequately reported. Since we aimed to include about 130 reviews in our initial sample to allow for “non-inclusion” (systematic reviews not fulfilling our inclusion criteria), a rapid Pubmed search suggested that focusing on these 4 journals and the Cochrane Library during a one year period would be sufficient. If it had not been the case, we would have included yet an additional journal.

6. The Methods section says nothing about statistical analysis. Specifically it doesn’t mention performing significance tests, nor which methods were used. In particular, I’m concerned about tests to compare journals. I don’t see a good justification for such analyses, as they relate neither to the study objectives nor do they have good power.

Answer: The method section/statistical analyses now reads (page 8): “Descriptive results are reported as numbers (proportions) and median (inter-quartile range (IQR)) as required. In order to check whether systematic reviews were different from one journal to the other, we performed all descriptive analyses separately according to title of the journal. Chi2 or Kruskal-Wallis tests were applied to test the null hypothesis of homogeneous distribution of characteristics and outcomes. We compared reviews from reviewers who answered our inquiry with reviews from those who did not, and across journals. Since Cochrane reviews were expected to be different from those published in the journals, we performed separate analyses with and without Cochrane reviews. We did not expect missing data. Statistical significance was defined as an alpha-error of 0.05 or less in two-sided tests. Analyses were performed using STATA version 13.”

6.1 Also the actual numbers in many of the cells are too small for reliable analysis.

Answer: We agree. Therefore, in order not to over-interpret the results of statistical tests comparing journals, we deleted most of the description in the secondary outcome section (page 13) and only kept this information in an Appendix table 2A and 3A.

7. I have reservations about the ad hoc composite score shown in table 1. Such composites are known to be problematic in many areas e.g. quality scores.
Answer: The composite score was created with the aim to standardize data extraction regarding the 6 procedures examined. We agree that composite scores are problematic and difficult to interpret per se. However rating each of the procedures allowed us to describe the application of the 6 procedures, for each review, in Appendix table 1A. Therefore, we kept our rating system for descriptive reasons in Appendix Table 1A but we deleted the paragraph: “Composite score” from the text (page 11), and from Table 3. Also, we deleted the column “total score” from Appendix Table 1.

Results and Discussion
8. It’s strange to take a random sample of 25 Cochrane reviews and yet end up with only 19 of these. Why not ensure you had 25 eligible reviews?

Answer: This illustrates that Cochrane reviews are the only published reviews that sometimes report on… 0 trials included! It is unlikely that including 25 eligible Cochrane reviews (instead of 19) would have changed the main message of our paper.

9. How many reviews were reviews of trials, and how many of other study types?

Answer: We have now included this information on the study designs included in each review in table 2, appendix table 1A and 2A.

10. The probability of suspecting something will be related to the number of studies in the review. Yet this information isn’t provided. (By contrast we’re told how many authors there were which seems far less relevant.)

Answer: The information regarding the number of studies included in each review was provided in appendix table 1A. We have now added this information in Table 2 and appendix table 2A and in the text (page 9): “The median number of articles included in each review was 28. Half of the systematic reviews included a mix of various study designs, while 39% included only RCTs (Table 2).”

10.1 How large were the reviews in which some inappropriate practices were suspected?

Answer: We thank you for pointing this out. We have added a sentence at the end of the results section, which reads (page 13): “The median number of studies included in the reviews that detected misconduct was 56 (IQR, 11 to 97) and was 28 (IQR, 12 to 57) in the reviews that did not detect misconduct. The difference did not reach statistical significance.”
11. P11: I can’t match the numbers in the paragraph on publication bias to Table 3.

Answer: We have clarified this, which now reads (page 10, “search for unpublished trials and test for publication bias”: “Seven reviews (6%) only discussed the risk of publication bias, and 32 reviews (27%) did not mention it at all (Table 3)..”

Minor points
12. Abstract: Objective is ambiguous – “To study whether authors of systematic reviews apply procedures to counter-balance research malpractice such as not publishing completed research,

duplicate publications, or selective reporting of outcomes, and whether they identify and report misconduct." Need to distinguish bad practices from counterbalancing procedures

Answer: the objective now reads:" To study whether systematic reviewers apply procedures to counter-balance some common research malpractice such as not publishing completed research, duplicate publications, or selective reporting of outcomes, and to see whether they identify and report misconduct"

13. Publication bias is not the same thing as unpublished studies. Nor is identifying bias the same as reducing it (p14).

Answer: We have changed the results headings "publication bias" to "search of unpublished trials and test for publication bias" and of "selective reporting of outcomes" to "contact with authors" in the text and the tables. Also, we hope that the English editing has now solved these confusions.

14. Nor is possible misconduct the same as misconduct (15/3). Nor is misconduct the same as fraud (p15).

Answer: We have deleted the sentence mentioning the term "fraud".

15. It is stated that PRISMA-P "recommends reporting of publication bias and selective outcome reporting" – i.e. in a review protocol. This doesn't make sense and it's not true. What PRISMA-P says (item 16) is: "specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies)" which is very different indeed.

Answer: We have rephrased this sentence which now reads: "The 17-items list mentions the assessment of meta-bias(es) such as publication bias across studies and selective outcome reporting within studies."

VERSION 2 – REVIEW

REVIEWER	Xavier Bosch Department of Internal Medicine, Hospital Clinic, Barcelona, Spain
REVIEW RETURNED	08-Feb-2016

GENERAL COMMENTS	All concerns and suggestions have been addressed

REVIEWER	Doug Altman University of Oxford, UK
REVIEW RETURNED	10-Feb-2016

GENERAL COMMENTS	The authors have given very satisfactory responses to all the revewis and i have no additioanl concerns.