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)	Electronic nicotine delivery systems and/or electronic non-nicotine delivery systems for tobacco smoking cessation or reduction: a systematic review and meta-analysis
AUTHORS	El Dib, Regina; Suzumura, Erica; Akl, Elie; Goma, Huda; Agarwal, Arnav; Chang, Yaping; Prasad, Manya; Ashoorion, Vahid; Heels-Ansdell, Diane; Maziak, Wasim; Guyatt, Gordon

VERSION 1 - REVIEW

REVIEWER	Konstantinos Farsalinos Onassis Cardiac Surgery Center Department of Pharmacy, University of Patras Greece 2 unpublished studies performed using unrestricted funds provided to the institution (Onassis Cardiac Surgery Center) by e-cigarette companies in 2013 (more than 3 years ago). Two studies funded by the non profit association AEMSA in 2013 (less than 3 years ago)and one study funded by the non-profit association TSFA in 2015.
REVIEW RETURNED	19-Jul-2016

GENERAL COMMENTS	This is one of the several (4 cited by the author) reviews on the effects of ENDS on smoking cessation. The authors have done an extensive and very detailed analysis, and they should be praised for that. I am not sure if this review adds anything new to the current knowledge. Despite that, there is a potential to add the message: the studies performed until now are totally inappropriate to make any conclusions about the effects of ENDS on smoking cessation. The studies included to the analysis are, in most cases totally inappropriate. The authors mention in the discussion section: “One could argue that these limitations make the pooling of results we have undertaken inadvisable. On the other hand, the pooling does highlight the possibility of an adverse effect of e-cigarettes on quit rates, a possibility that until definitively refuted by randomized trials needs consideration in policy debates regarding e-cigarettes”. In my opinion, analyzing these studies is not only uninformative but can be highly misleading. Thus, they should be probably ignored rather than considered in any policy debate. I do not see a purpose of trying to generate data on ENDS efficacy in smoking cessation while such data do not really exist.

There are several very important problems in the studies included, some of which are highlighted by the authors.

Selection of studies. There is an issue with the selection of studies that were analyzed. The authors do present the problem of the outcome being present at the start of the study. For example, reference 44 (Vickerman et al.) was a study of a quitline service. The study did not assess whether ENDS were used before the subjects called the quitline services (7 months before the survey was performed). Of course, this is mentioned herein, but it represents a very serious problem because it is likely that some subjects were by definition “failures to quit” at baseline. This does not really adherent to the principles of a cohort study evaluating the impact of ENDS on smoking cessation because the outcome may have existed since the baseline. In fact, the authors released a press statement saying that: “The recently published article by Dr Katrina Vickerman and colleagues has been misinterpreted by many who have written about it. It was never intended to assess the effectiveness of the e-cig as a mechanism to quit.” (unfortunately, the press statement has been removed from their website, but it was available until June 2015 – See Farsalinos & Le Houezec, Risk Manag Healthc Policy. 2015; 8: 157–167).

The same issue is present in the study by Manzoli et al. (reference 42), although in that case there is absolute certainty that the outcome existed from the beginning. Dual users were by definition “failures to quit” at baseline.

The study by Al Delaimy et al. (reference 40) is even worse, with several methodological issues. First, the comparison was between subjects having used ENDS and subjects who will never use ENDS. The methodology section mentions: “The main predictor was the use of e-cigarettes, which, for the purposes of the current study, we categorized as (1) a report of “will never use e-cigarettes” at baseline and follow-up (reference group in regression models) and (2) a report of “have used e-cigarettes” at baseline and follow-up”. So, technically, the study was not a comparison between ENDS users vs. non-users. Moreover, you mention that there were 628 participants, but Al-Delaimy et al. present the analysis of 368 participants (table 4, 191 who have used ENDS vs. 177 who will never use). Moreover, you mention that there was no bias of the outcome being present at the start of the study, but Al-Delaimy et al. mention that the group was subjects with “a report of “have used e-cigarettes” at baseline and follow-

up”. The title is characteristic of this: “E-Cigarette Use in the Past and Quitting Behavior in the Future”. I think the assessment of bias and other information needs to be corrected for this study.

Finally, to the best of my knowledge, the study by Harrington et

	al. (reference 45) is available only as a conference abstract. If I am correct that this is just an abstract, I do not think it is appropriate to include this in the analysis because I am unable to evaluate the findings. How were the ORs calculated? There is nothing mentioned in the abstract, and based on the numbers reported I have serious doubts about the ORs reported in the analysis. This study should be removed. In case I am wrong, please provide a link to the full text of the study. I am wondering if the results would change if these studies were excluded from the analysis, since it is obvious from Figure 4 that they are the worse in terms of the bias. It would be better to remove them rather than use largely irrelevant studies. Of course, in that case the number of studies included would be very small and probably inadequate for any form of analysis. In any case, it seems to me that the scientific community (referring not only to the present analysis but to all previous similar reviews) is trying to generate data and information when such data do not exist. These studies should not even be classified as cohort studies evaluating ENDS efficacy as smoking substitutes. Any analysis has only academic interest but very limited (and potentially misleading) information about regulatory or policy decisions. Therefore, I do not suggest rejecting the manuscript, for 2 reasons: 1. The authors have done a substantial and detailed work analyzing all these studies 2. It is a good opportunity to present in more depth the major problems related to the studies which, in my opinion, are only supposed to evaluate the effects of ENDS on smoking cessation but fail to do so. There are other potential sources of bias that are not mentioned or discussed in the manuscript. The authors mention that they assessed the impact of ENDS vs. no smoking cessation aid or alternative smoking cessation aid, regardless of whether the users were using them as part of a quit attempt. The problem is that efforts to quit “cold turkey” (no aid) or with other aids (NRTs, oral medications, psychological support) are by definition intended and conscious quit attempts. With ENDS there is a lot of experimentation for curiosity, without any real intention to quit. Thus, there is a potential for selection bias, the populations compared are not very similar in terms of intention to quit. An added problem is that experimentation with any smoking-cessation aid is unlikely to result in smoking cessation. Thus, this is further extending to the above mentioned bias. Another type of selection bias is related to the dependence on smoking of the populations compared. Usually the FTCD is used, however, studies of population samples have shown that

	other tests (like Strength of Urges To Smoke) are stronger predictors of successful cessation (see: Fidler et al. Addiction 2011, Fidler & West, Drug Alcohol Depend 2011, Kozlowski et al., Drug Alcohol Depend 1994). Did the authors assess if the dependence scales used in different studies (if available) were appropriate? Considering that smoking cessation is more a behavioral change than the therapy of a disease (despite the WHO classification of smoking as a disease), I consider the above limitations extremely important. Added to that, the problem with RCTs is the inability to select products based on self-preference. There is a lot of discussion in the literature showing that ENDS are chosen based on self-preference and satisfaction. This cannot be addressed in a conventional RCT which follows the principles of RCTs for pharmaceuticals. Another limitation not mentioned in the study is the use of outdated products, especially in the RCTs by Bullen et al. and Caponnetto et al. The products were outdated and off the market by the time the study were published, and that was due to the development of better products. The studies were published more than 3 years ago, and were probably initiated 5 years ago. Thus, how can someone support the relevance of the findings for today, considering the fast development (I would say, complete transformation) of the ENDS products? In fact, considering the dates of the studies, it is highly likely that almost all studies have evaluated what are now considered outdated products. My suggestion for the manuscript is to perform major revisions. These should include the more extensive discussion of the limitations of the studies included in the analysis (based on my comments above), the exclusion of some studies which are particularly misleading, and a clear presentation of the conclusion that current evidence is inappropriate to assess the impact of ENDS on smoking cessation or reduction. I think the correct term is “inappropriate”, not “insufficient”. These limitations and the conclusion should also be presented in the abstract. Minor comments Page 4 to 5. The presentation of 3rd and 4th generation devices is not entirely correct. 3rd generation devices (which include variable wattage devices; variable voltage is already considered outdated and is in reality wrong since volts cannot determine the power of the device without considering the resistance of the atomizers) are used only with refillable tank systems (which are presented as 4th generation). Page 5, line 41. Please substitute the word “although” with the word “however”.

	Table 4, risk of bias. An RCT evaluating ENDS vs. any other intervention cannot be participant-blinded, unless you compare zero-nicotine with nicotine-containing ENDS (as it happened with the ECLAT study). Thus, by definition all RCTs evaluating ENDS will suffer from this bias. Is it appropriate to judge the quality of the studies based on the inherent inability to blind participants? Perhaps the authors should discuss this issue. Appendix Figure 4. Why is the study by Hajek et al. missing from the analysis? It is 1 of the 2 cohort studies which were not rated as high risk of bias for limitations in matching exposed and unexposed groups or adjusting analysis for prognosis variables. I think it should be included, despite the low sample size. Page 11, lines 18-23. Reference 46 (Hajek et al.) is mentioned as both supplying and not supplying the requested data. Appendix table 3. The study by Adriaens et al. used an eGo-type battery which is not cigalike and included a tank-type atomizer (one of the first tank atomizers developed). Although an outdated product, it is not a cigalike (the latter is a combination of a small, cigarette-like battery and a prefilled cartomizer).

REVIEWER	Kristian Filion McGill University, Canada
REVIEW RETURNED	20-Jul-2016

GENERAL COMMENTS	GENERAL COMMENTS: In this systematic review and meta-analysis, El Dib and colleagues examine the effect of electronic nicotine delivery systems (ENDS) and electronic non-nicotine delivery systems (ENNDS) on smoking abstinence and cigarette smoking reduction. Overall, this study has many strengths, including an a priori protocol, the inclusion of RCT and cohort data, and the use of the GRADE criteria. However, there are also some important limitations and some issues that require clarification; these are discussed in the Specific Comments below. SPECIFIC COMMENTS: 1. The systematic review and meta-analysis included “current or former cigarette smokers”. The rationale for including former cigarette smokers is unclear. Are they at risk of the outcomes (smoking cessation or a 50% reduction in cigarette use)? Similarly, the authors also included users of other combustible

	tobacco products. Were they at risk of the study outcomes? If not, they should be excluded. 2. P-values should be removed from the text. The estimation of the amount of heterogeneity that is present via the I2 statistic is preferred to hypothesis testing, particularly given the small number of studies and how underpowered such tests of heterogeneity are. 3. Along those lines, greater caution is needed when interpreting the results when no association is present. The authors are currently concluding that no difference is present in several places where confidence intervals are wide and include clinically important treatment effects. It would be more accurate to describe such results as inconclusive. 4. Quality assessment of cohort studies was performed using the Newcastle-Ottawa scale. This approach is now outdated, having been replaced by the A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI) and the forthcoming ROBINS-I tool (Risk Of Bias In Non-randomized Studies - of Interventions; the tool is available but the detailed description of the ROBINS-I is forthcoming – see https://sites.google.com/site/riskofbiastool/). 5. The flow diagram and text describes 3 included RCTs but the quality assessment figures and forest plot only include 2 RCTs. Please clarify in the figure legends. 6. On page 26, the authors state that the minimal criteria for pooling was five available studies (when discussing subgroup analyses). The rationale for pooling data across the three RCTs is therefore unclear. 7. There is substantial clinical heterogeneity among included studies. Consequently, the appropriateness of pooling these data is unclear. For this reason, the systematic review component of the manuscript is particularly informative. 8. How were patients who were lost-to-follow-up handled among the included studies? Were they assumed to have returned to smoking? If so, to what amount of smoking? Some discussion of this issue is warranted. 9. I have some concerns about relying on the previous Cochrane review for the identification of all relevant studies published prior to 2014. While I agree that it is a high quality review, given the differences in inclusion criteria, could some relevant studies have been missed? 10. Risk differences were calculated for 6 to 12 months of follow-up. Doing so assumes that the absolute treatment effect is the same over the duration of this period, which is unlikely. Some discussion of the assumptions involved in this approach is needed.

	11. Minor points: a. Throughout the manuscript, there are several typographical and some formatting issues. For example: i. Page 2, line 6: ENDS and ENNDS are not defined in the abstract. ii. Page 2, line 46: “data from RCTs are of low and observational studies of very low certainty” is missing a word. iii. Page 2, line 19: “…while ENNDS devices ENNDS are…”. iv. The references are not in the order they appear, with references 47, 48, and 49 appearing after reference 11 and before reference 12. v. Page 6, line 32: The abbreviations ENDS and ENNDS were already defined in the Introduction. b. Page 4, line 10: Bupropion and varenicline should also be mentioned as pharmacological interventions. c. Page 7, line 37: please provide the date of the submission of the manuscript. d. Several tables appear in the middle of the text rather than at the end of the manuscript. e. Was the protocol registered? If so, please provide the protocol registration number and/or website?

REVIEWER	Hayden McRobbie Wolfson Institute of Preventive Medicine, Queen Mary University of London, UK Hayden McRobbie has received investigator led research funding and honoraria for speaking at educational meetings from Pfizer Inc. He has also received honoraria from Johnson and Johnson for speaking at educational meetings and an advisory board meeting. Hayden McRobbie is an author of the Cochrane Systematic Review of Electronic cigarettes for smoking cessation and reduction.
REVIEW RETURNED	21-Jul-2016

GENERAL COMMENTS	This manuscript presents the findings of a systematic review and meta-analysis of electronic cigarettes for smoking cessation. I can appreciate the large amount of work that has gone into undertaking this systematic review and meta-analysis. I'll start with my more major comments and have listed a couple of more minor comments at the end. Regarding the data from randomised controlled trials: The way that missing data was handled in this meta-analysis is unusual for this field. The convention in smoking cessation studies is to assume that those lost to follow-up (LTFU) are smoking. I understand that this approach is not perfect, but I think that an intention-to-treat analysis that imputes those LTFU as smokers should be the primary outcome, with a complete-case analysis as a sensitivity analysis. This would align with what the two RCTs reported. The sensitivity analysis used in this meta-

	analysis (i.e. all participants with missing data from the arm of the study with the lower quit rates were assumed to have 3 times the quit rate as those with complete data, and those with missing data from the other arm were assumed to have the same quit rate as participants with complete data) does not appear to be a plausible assumption to me, even as a worst case scenario. Part of the problem with loss-to-follow-up in smoking cessation studies is that those who relapse do not want to be followed-up. So, whilst there may be limitations in assuming all those LTFU are smoking, there would be more plausible to expect that the quit rates in this group would be lower than those who were followed-up. What was the rationale for using a random effects model? The Cochrane Group prefers to use a fixed effects model and this would have been appropriate in this case as there was no significant heterogeneity. Regarding the prospective cohort data: The longitudinal surveys (i.e. Al-Delaimy 2015, Bordeud 2014, Brose 2015, Manzoli 2015, and Prochaska 2014) all share the same serious limitation. That is they recruited only people who were currently smoking. Any people from the same population who had used e-cigarettes and stopped smoking were excluded, meaning that you are left with people who were not helped by e-cigarettes (treatment failures in other words). This would be true of other smoking cessation aids, as they do not help everyone that uses them. To determine the effectiveness of e-cigarettes for smoking cessation smokers need to be recruited before starting e-cigarettes. I acknowledge that you do highlight this limitation in the discussion as well as other methodological problems with these studies including the definition of e-cigarette use, whether smokers were using these to make a quit attempt and recall bias). However, given the significance of tis problem I think that it would be more appropriate to not pool these data. Another reason for not pooling data with these studies is the significant heterogeneity that exists. The Cochrane Handbook gives a number of options for dealing with this. One is to perform a random-effects meta-analysis, but notes that 'this is not a substitute for a thorough investigation of heterogeneity.' Given my comments above, as well as your own statements concerning the certainty of evidence from these observational studies (noted in the last paragraph on page 20), which is rated as low to very low, the most appropriate course of action that would be to not undertake meta-analysis with data from these studies. Regarding study selection The RCT, Adriaens 2014 does not technically meet the inclusion criteria. Whilst people were followed up for 8 months, a comparison group (no e-cigarettes) is only provided for 3 months and then this group was provided with e-cigarettes. I don't think that Vickerman 2013 can be considered as a prospective cohort study. This study surveyed people seven months after they enrolled into a smoking cessation service (Quitline), asking them about their retrospective use of e-

	cigarettes. It is in my view a cross-sectional survey and does not provide any useful information about the effects of e-cigarette use on smoking cessation. Abstract ENDS and ENNDS abbreviations need to be written fully before used first. The abstract should be updated to reflect the above suggested changes Methods The type of abstinence is not defined (e.g. point-prevalence, continuous abstinence). Where studies reported multiple outcomes, which was preferred? Discussion Whilst the discussion reflects the current results, it needs to be updated to reflect the changes suggested above.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name

Konstantinos Farsalinos

Institution and Country

Onassis Cardiac Surgery Center
Department of Pharmacy, University of Patras
Greece

Please state any competing interests or state 'None declared':
2 unpublished studies performed using unrestricted funds provided to the institution (Onassis Cardiac Surgery Center) by e-cigarette companies in 2013 (more than 3 years ago). Two studies funded by the non profit association AEMSA in 2013 (less than 3 years ago)and one study funded by the non-profit association TSFA in 2015.

This is one of the several (4 cited by the author) reviews on the effects of ENDS on smoking cessation. The authors have done an extensive and very detailed analysis, and they should be praised for that. I am not sure if this review adds anything new to the current knowledge. Despite that, there is a potential to add the message: the studies performed until now are totally inappropriate to make any conclusions about the effects of ENDS on smoking cessation. The studies included to the analysis are, in most cases totally inappropriate.
The authors mention in the discussion section: "One could argue that these limitations make the pooling of results we have undertaken inadvisable. On the other hand, the pooling does highlight the possibility of an adverse effect of e-cigarettes on quit rates, a possibility that until definitively refuted by randomized trials needs consideration in policy debates regarding e-cigarettes". In my opinion, analyzing these studies is not only uninformative but can be highly misleading. Thus,

they should be probably ignored rather than considered in any policy debate. I do not see a purpose of trying to generate data on ENDS efficacy in smoking cessation while such data do not really exist.

Response: We agree with the reviewer that the studies are very biased and have highlighted in the manuscript. That it's what we found out there in the literature, and we are trying to show to the scientific community that nothing can be concluded so far related to the effectiveness of ENDS and/or ENNDS. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter.

There are several very important problems in the studies included, some of which are highlighted by the authors.
Selection of studies. There is an issue with the selection of studies that were analyzed. The authors do present the problem of the outcome being present at the start of the study. For example, reference 44 (Vickerman et al.) was a study of a quitline service. The study did not assess whether ENDS were used before the subjects called the quitline services (7 months before the survey was performed). Of course, this is mentioned herein, but it represents a very serious problem because it is likely that some subjects were by definition “failures to quit” at baseline. This does not really adherent to the principles of a cohort study evaluating the impact of ENDS on smoking cessation because the outcome may have existed since the baseline. In fact, the authors released a press statement saying that: “The recently published article by Dr Katrina Vickerman and colleagues has been misinterpreted by many who have written about it. It was never intended to assess the effectiveness of the e-cig as a mechanism to quit.” (unfortunately, the press statement has been removed from their website, but it was available until June 2015 – See Farsalinos & Le Houezec, Risk Manag Healthc Policy. 2015; 8: 157–167).

Response: We agree with the reviewer regarding the limitations and have highlighted in the manuscript. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter.

The same issue is present in the study by Manzoli et al. (reference 42), although in that case there is absolute certainty that the outcome existed from the beginning. Dual users were by definition “failures to quit” at baseline.

Response: Our view is, once again, rather than dismissing altogether, the study should be considered in light of its limitations.

The study by Al Delaimy et al. (reference 40) is even worse, with several methodological issues. First, the comparison was between subjects having used ENDS and subjects who will never use ENDS. The methodology section mentions: “The main predictor was the use of e-cigarettes, which, for the purposes of the current study, we categorized as (1) a report of “will never use e-cigarettes” at baseline and follow-up (reference group in regression models) and (2) a report of “have used e-cigarettes” at baseline and follow-up”. So, technically, the study was not a comparison between ENDS users vs. non-users. Moreover, you mention that there were 628 participants, but Al-Delaimy et al. present the analysis of 368 participants (table 4, 191 who have

used ENDS vs. 177 who will never use). Moreover, you mention that there was no bias of the outcome being present at the start of the study, but Al-Delaimy et al. mention that the group was subjects with “a report of “have used e-cigarettes” at baseline and follow-up”. The title is characteristic of this: “E-Cigarette Use in the Past and Quitting Behavior in the Future”. I think the assessment of bias and other information needs to be corrected for this study.

Response: 628 were the number at baseline. Table 1. “Study characteristics related to design of study, setting, number of participants, mean age, gender, inclusion and exclusion criteria, and follow-up.” presents the number of baseline, not the number of patients in the final analysis. With regards the risk assessment, we have corrected it in Figure 4 showing high risk of bias for the domain “can we be confident that the outcome of interest was not present at start of study?”

Finally, to the best of my knowledge, the study by Harrington et al. (reference 45) is available only as a conference abstract. If I am correct that this is just an abstract, I do not think it is appropriate to include this in the analysis because I am unable to evaluate the findings. How were the ORs calculated? There is nothing mentioned in the abstract, and based on the numbers reported I have serious doubts about the ORs reported in the analysis. This study should be removed. In case I am wrong, please provide a link to the full text of the study.

Response: Harrington study is only available as abstract and, after contact with the author by e-mail they provided us with a poster presentation. Although it is abstract we did not state in our manuscript that we would exclude conference abstracts. Usually for systematic reviews we considered all type of published and unpublished data. We do not exclude studies due to be an abstract. Please find attached the poster the authors sent to us.

I am wondering if the results would change if these studies were excluded from the analysis, since it is obvious from Figure 4 that they are the worse in terms of the bias. It would be better to remove them rather than use largely irrelevant studies. Of course, in that case the number of studies included would be very small and probably inadequate for any form of analysis.

Response: We have conducted a subgroup analysis comparing the four studies with which the reviewer has a particular concern (Al-Delaimy; Harrington; Manzoli; Vickerman) and we also have included Brose study (as per the risk of bias assessment figure shows that this study is also one of the worst studies) to the three other studies (Biener; Borderud; and Prochaska). Please find the new meta-analysis attached as Appendix Figure 4 and the text below:

A second sensitivity analysis from the same eight cohort studies [26-29, 40-45], examined whether low and high risk of bias limited to “two or fewer domains rated as low risk of bias” versus “three or more domains rated as low risk of bias” differed substantially. There were substantial differences in the point estimates between the “two or fewer domains rated as low risk of bias” group (adjusted OR 0.61, 95% CI 0.49, 0.75; $p < 0.001$; I2=0%) and the “three or more domains rated as low risk of bias” (adjusted OR 1.26, 95% CI 0.68, 2.33; p=0.46; I2=51%), with an interaction p-value of 0.03 (Appendix Figure 4).

In any case, it seems to me that the scientific community (referring not only to the present analysis but to all previous similar reviews) is trying to generate data and information when such

data do not exist. These studies should not even be classified as cohort studies evaluating ENDS efficacy as smoking substitutes. Any analysis has only academic interest but very limited (and potentially misleading) information about regulatory or policy decisions. Therefore, I do not suggest rejecting the manuscript, for 2 reasons:
1. The authors have done a substantial and detailed work analyzing all these studies
2. It is a good opportunity to present in more depth the major problems related to the studies which, in my opinion, are only supposed to evaluate the effects of ENDS on smoking cessation but fail to do so.

Response: Our disagreement with the reviewer is not fundamental – the issue is a matter of degree. We agree with the major limitations of these studies. The reviewer would have us simply point out that they are so flawed they do not bear on the issue at hand. We agree they are highly flawed, and have pointed out the limitations, but do not think they should be dismissed altogether.

There are other potential sources of bias that are not mentioned or discussed in the manuscript. The authors mention that they assessed the impact of ENDS vs. no smoking cessation aid or alternative smoking cessation aid, regardless of whether the users were using them as part of a quit attempt. The problem is that efforts to quit “cold turkey” (no aid) or with other aids (NRTs, oral medications, psychological support) are by definition intended and conscious quit attempts. With ENDS there is a lot of experimentation for curiosity, without any real intention to quit. Thus, there is a potential for selection bias, the populations compared are not very similar in terms of intention to quit. An added problem is that experimentation with any smoking-cessation aid is unlikely to result in smoking cessation. Thus, this is further extending to the above mentioned bias. Another type of selection bias is related to the dependence on smoking of the populations compared. Usually the FTCD is used, however, studies of population samples have shown that other tests (like Strength of Urges To Smoke) are stronger predictors of successful cessation (see: Fidler et al. Addiction 2011, Fidler & West, Drug Alcohol Depend 2011, Kozlowski et al., Drug Alcohol Depend 1994). Did the authors assess if the dependence scales used in different studies (if available) were appropriate?

Response: We added the issue about potential for selection bias under “Strengths and limitations” of the review under discussion. Please, find it below:
“Finally, another limitation of the observational studies in this review is the potential for selection bias as the populations compared differ in terms of intention to quit.”
We did not assess if the dependence scales used in different studies were appropriate because there was no information provided by the included studies about them.

Considering that smoking cessation is more a behavioral change than the therapy of a disease (despite the WHO classification of smoking as a disease), I consider the above limitations extremely important. Added to that, the problem with RCTs is the inability to select products based on self-preference. There is a lot of discussion in the literature showing that ENDS are chosen based on self-preference and satisfaction. This cannot be addressed in a conventional RCT which follows the principles of RCTs for pharmaceuticals. Another limitation not mentioned in the study is the use of outdated products, especially in the RCTs by Bullen et al. and Caponnetto et al. The products were outdated and off the market by the time the study were published, and that was due to the development of better products. The studies were published more than 3 years ago, and were probably initiated 5 years ago. Thus, how can someone support the relevance of the findings for today, considering the fast development (I would say, complete

transformation) of the ENDS products? In fact, considering the dates of the studies, it is highly likely that almost all studies have evaluated what are now considered outdated products.

Response: Thanks for the valuable suggestions, we have added the following:

Regarding the issue of self-selection, the reviewer has compellingly made the point regarding selection bias. If one compares those who choose to use e-cigarettes versus those who don't, it is very likely that any differences will be due to the nature of the populations rather than e-cigarettes (i.e. prognostic differences). This issue can be dealt with only in the context of an RCT, which, ideally, would enroll individuals open to using e-cigarettes. This openness would deal, at least to some extent, with the reviewer's concern.

Regarding the earlier versions of e-cigarettes tested in the RCTs, we have highlighted this issue in the revised manuscript as follows: "Furthermore, in all these RCTs, the ENDS tested were earlier generation; it is possible that later generation of e-cigarettes would have greater benefit."

My suggestion for the manuscript is to perform major revisions. These should include the more extensive discussion of the limitations of the studies included in the analysis (based on my comments above), the exclusion of some studies which are particularly misleading, and a clear presentation of the conclusion that current evidence is inappropriate to assess the impact of ENDS on smoking cessation or reduction. I think the correct term is "inappropriate", not "insufficient". These limitations and the conclusion should also be presented in the abstract.

Response: We have re-worded the text to reflect that the results failed to show a difference between the comparison groups.

Minor comments
Page 4 to 5.
The presentation of 3rd and 4th generation devices is not entirely correct. 3rd generation devices (which include variable wattage devices; variable voltage is already considered outdated and is in reality wrong since volts cannot determine the power of the device without considering the resistance of the atomizers) are used only with refillable tank systems (which are presented as 4th generation).

Response: Thanks. We re-phrased it to:

"The third generation of ENDS include variable wattage devices are used only with refillable tank systems."

Page 5, line 41.
Please substitute the word "although" with the word "however".

Response: Thanks; done.

Table 4, risk of bias.
An RCT evaluating ENDS vs. any other intervention cannot be participant-blinded, unless you compare zero-nicotine with nicotine-containing ENDS (as it happened with the ECLAT study). Thus, by definition all RCTs evaluating ENDS will suffer from this bias. Is it appropriate to judge the quality of the studies based on the inherent inability to blind participants? Perhaps the authors should discuss this issue.

Response: We agree that why we rated the domain “was there blinding of participant?” as high risk of bias to Adriaens and Bullen studies (Figure 3).

Appendix Figure 4. Why is the study by Hajek et al. missing from the analysis? It is 1 of the 2 cohort studies which were not rated as high risk of bias for limitations in matching exposed and unexposed groups or adjusting analysis for prognosis variables. I think it should be included, despite the low sample size.

Response: The Hajek study was not included in the meta-analysis because there was no raw data provided by the study to enable us to include it in the meta-analysis.

Page 11, lines 18-23.
Reference 46 (Hajek et al.) is mentioned as both supplying and not supplying the requested data.

Response: The Hajek study provided some but not all of the missing data.

Appendix table 3.
The study by Adriaens et al. used an eGo-type battery which is not cigalike and included a tank-type atomizer (one of the first tank atomizers developed). Although an outdated product, it is not a cigalike (the latter is a combination of a small, cigarette-like battery and a prefilled cartomizer).

Response: Many thanks, we changed it to:

“Not a cigalike (tank-type atomizer) (second generation ENDS devices).”

We have acknowledged your great inputs under the acknowldgements section. Thank you very much.

Reviewer: 2
Reviewer Name

Kristian Filion

Institution and Country

McGill University, Canada

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below
GENERAL COMMENTS:
In this systematic review and meta-analysis, El Dib and colleagues examine the effect of electronic nicotine delivery systems (ENDS) and electronic non-nicotine delivery systems (ENNDS) on smoking abstinence and cigarette smoking reduction. Overall, this study has many strengths, including an a priori protocol, the inclusion of RCT and cohort data, and the use of the GRADE criteria. However, there are also some important limitations and some issues that require clarification; these are discussed in the Specific Comments below.

SPECIFIC COMMENTS:
1. The systematic review and meta-analysis included “current or former cigarette smokers”. The rationale for including former cigarette smokers is unclear. Are they at risk of the outcomes (smoking cessation or a 50% reduction in cigarette use)? Similarly, the authors also included users of other combustible tobacco products. Were they at risk of the study outcomes? If not, they should be excluded.

Response: This review was initially conducted as part of a contract with the World Health Organization. The WHO specified that they wanted former smokers included, presumably addressing the issue of whether e-cigarette use prevented the resumption of smoking in former smokers, and they wanted those using other tobacco products included, presumably because e-cigarettes might reduce the use of these products. We did not, however, find any studies addressing either of these populations. We have therefore removed references to these populations from the manuscript.

2. P-values should be removed from the text. The estimation of the amount of heterogeneity that is present via the I2 statistic is preferred to hypothesis testing, particularly given the small number of studies and how underpowered such tests of heterogeneity are.

Response: The figures are generated automatically along with the I2 and their p-values.

3. Along those lines, greater caution is needed when interpreting the results when no association is present. The authors are currently concluding that no difference is present in several places where confidence intervals are wide and include clinically important treatment effects. It would be more accurate to describe such results as inconclusive.

Response: The reviewer is quite right. We have changed wording: Each time previously we said “showed no difference” or “suggested no difference” we have substituted “failed to show a difference.”

4. Quality assessment of cohort studies was performed using the Newcastle-Ottawa scale. This approach is now outdated, having been replaced by the A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI) and the forthcoming

ROBINS-I tool (Risk Of Bias In Non-randomized Studies - of Interventions; the tool is available but the detailed description of the ROBINS-I is forthcoming – see https://sites.google.com/site/riskofbiastool/).

Response: The ROBINS-I tool is extremely complicated and labor intensive. We used a modified version of the Ottawa-Newcastle instrument (reference below) that includes confidence in assessment of exposure and outcome, adjusted analysis for differences between groups in prognostic characteristics, and missing data, to assess the quality assessment of cohort studies. We continue to think that this instrument is a reasonable alternative to ROBINS-I.

Reference
Guyatt GH, Busse JW. Modification of Ottawa-Newcastle to assess risk of bias in nonrandomized trials. http://distillercer.com/resources/.

5. The flow diagram and text describes 3 included RCTs but the quality assessment figures and forest plot only include 2 RCTs. Please clarify in the figure legends.

Response: The forest plot includes only two RCTs because there was no available raw data for the third one. The figures for risk of bias assessment indeed included the three RCTs: Figure 3 includes Adriaens and Bullen while Figure 2 includes Bullen and Caponnetto, the latter one comparing ENDS versus ENNDS and, the former one comparing ENDS versus other cessation aids.

6. On page 26, the authors state that the minimal criteria for pooling was five available studies (when discussing subgroup analyses). The rationale for pooling data across the three RCTs is therefore unclear.

Response: This criterion was related to subgroup analysis in which we stated “at least two in each sub-group”.

7. There is substantial clinical heterogeneity among included studies. Consequently, the appropriateness of pooling these data is unclear. For this reason, the systematic review component of the manuscript is particularly informative.

Response: Adhering to the GRADE approach, in the presence of substantial heterogeneity we still presented pooled estimates, but rated down the quality of the evidence for inconsistency.

8. How were patients who were lost-to-follow-up handled among the included studies? Were they assumed to have returned to smoking? If so, to what amount of smoking? Some discussion of this issue is warranted.

Response: For dealing with missing data, we used complete case as our primary analysis; that is, we excluded participants with missing data. If results of the primary analysis achieved or approached statistical significance, we conducted sensitivity analyses to test the robustness of those results. Specifically, we conducted a plausible worst-case sensitivity analysis in which all

participants with missing data from the arm of the study with the lower quit rates were assumed to have 3 times the quit rate as those with complete data, and those with missing data from the other arm were assumed to have the same quit rate as participants with complete data.

9. I have some concerns about relying on the previous Cochrane review for the identification of all relevant studies published prior to 2014. While I agree that it is a high quality review, given the differences in inclusion criteria, could some relevant studies have been missed?

Response: We would argue that the answer is no. The previous Cochrane review presented similar eligibility criteria, and was comprehensive in their search strategies and databases scrutinized.

10. Risk differences were calculated for 6 to 12 months of follow-up. Doing so assumes that the absolute treatment effect is the same over the duration of this period, which is unlikely. Some discussion of the assumptions involved in this approach is needed.

Response: We have acknowledged this in our limitations section as follows: “A final statistical limitation is that we calculated differences from 6 to 12 months of follow-up. Absolute differences may differ across this time frame and constitute a source of variability.”

11. Minor points:
a. Throughout the manuscript, there are several typographical and some formatting issues. For example:
i. Page 2, line 6: ENDS and ENNDS are not defined in the abstract.

Response: Many thanks; we have added it.

ii. Page 2, line 46: “data from RCTs are of low and observational studies of very low certainty” is missing a word.

Response: Many thanks; we rephrased to “data from RCTs are of low certainty and observational studies of very low certainty.”

iii. Page 2, line 19: “…while ENNDS devices ENNDS are…”.

Response: Many thanks; we have deleted the second “ENNDS”.

iv. The references are not in the order they appear, with references 47, 48, and 49 appearing after reference 11 and before reference 12.

Response: Many thanks; we fixed it.

v. Page 6, line 32: The abbreviations ENDS and ENNDS were already defined in the Introduction.

Response: Many thanks; we deleted the full text for ENDS and ENNDS as this was already defined in the Introduction.

b. Page 4, line 10: Bupropion and varenicline should also be mentioned as pharmacological interventions.

Response: Many thanks; added it.

c. Page 7, line 37: please provide the date of the submission of the manuscript.
Response: Not sure, we believe this is due to the journal style request.

d. Several tables appear in the middle of the text rather than at the end of the manuscript.

Response: This is due to the journal style request.

e. Was the protocol registered? If so, please provide the protocol registration number and/or website?

Response: No protocol register exists.

Reviewer: 3

Reviewer Name

Hayden McRobbie

Institution and Country

Wolfson Institute of Preventive Medicine, Queen Mary University of London, UK

Please state any competing interests or state 'None declared':
Hayden McRobbie has received investigator led research funding and honoraria for speaking at educational meetings from Pfizer Inc. He has also received honoraria from Johnson and Johnson for speaking at educational meetings and an advisory board meeting.

Hayden McRobbie is an author of the Cochrane Systematic Review of Electronic cigarettes for smoking cessation and reduction.

Please leave your comments for the authors below

This manuscript presents the findings of a systematic review and meta-analysis of electronic cigarettes for smoking cessation. I can appreciate the large amount of work that has gone into undertaking this systematic review and meta-analysis.

I'll start with my more major comments and have listed a couple of more minor comments at the end.

Regarding the data from randomised controlled trials:
The way that missing data was handled in this meta-analysis is unusual for this field. The convention in smoking cessation studies is to assume that those lost to follow-up (LTFU) are smo king. I understand that this approach is not perfect, but I think that an intention-to-treat analysis that imputes those LTFU as smokers should be the primary outcome, with a complete-case analysis as a sensitivity analysis. This would align with what the two RCTs reported. The sensitivity analysis used in this meta-analysis (i.e. all participants with missing data from the arm of the study with the lower quit rates were assumed to have 3 times the quit rate as those with complete data, and those with missing data from the other arm were assumed to have the same quit rate as participants with complete data) does not appear to be a plausible assumption to me, even as a worst case scenario. Part of the problem with loss-to-follow-up in smoking cessation studies is that those who relapse do not want to be followed-up. So, whilst there may be limitations in assuming all those LTFU are smoking, there would be more plausible to expect that the quit rates in this group would be lower than those who were followed-up.

Response: As the reviewers pointed out this approach is not perfect. For dealing with missing data, we usually use complete case as primary analysis. If results of the primary analysis achieve or approach statistical significance, we conduct sensitivity analyses to test the robustness of those results. Specifically, in this review we conducted a plausible worst-case sensitivity analysis in which all participants with missing data from the arm of the study with the lower quit rates were assumed to have 3 times the quit rate as those with complete data, and those with missing data from the other arm were assumed to have the same quit rate as participants with complete data. Therefore, we believe that using 3 times the quit rate as those with complete data we used the worst case scenario and, a sufficient approach in this review.

What was the rationale for using a random effects model? The Cochrane Group prefers to use a fixed effects model and this would have been appropriate in this case as there was no significant heterogeneity.

Response: There are three schools of thought with respect to use of fixed and random effect models: those who prefer always to use fixed effects, those who prefer (almost) always random effects, and those who would choose fixed and random depending on the degree of heterogeneity. Each argument has its proponents within the statistical community. The argument in favor of the second rather than the third is a) there is always some heterogeneity, so any threshold of switching models is arbitrary and b) when there is little heterogeneity, fixed and random yield similar or identical results, so one might as well commit oneself to random from the start. We find these two arguments compelling; thus, our choice.

Regarding the prospective cohort data:
The longitudinal surveys (i.e. Al-Delaimy 2015, Bordeud 2014, Brose 2015, Manzoli 2015, and Prochaska 2014) all share the same serious limitation. That is they recruited only people who

were currently smoking. Any people from the same population who had used e-cigarettes and stopped smoking were excluded, meaning that you are left with people who were not helped by e-cigarettes (treatment failures in other words). This would be true of other smoking cessation aids, as they do not help everyone that uses them. To determine the effectiveness of e-cigarettes for smoking cessation smokers need to be recruited before starting e-cigarettes. I acknowledge that you do highlight this limitation in the discussion as well as other methodological problems with these studies including the definition of e-cigarette use, whether smokers were using these to make a quit attempt and recall bias). However, given the significance of tis problem I think that it would be more appropriate to not pool these data. Another reason for not pooling data with these studies is the significant heterogeneity that exists. The Cochrane Handbook gives a number of options for dealing with this. One is to perform a random-effects meta-analysis, but notes that 'this is not a substitute for a thorough investigation of heterogeneity.' Given my comments above, as well as your own statements concerning the certainty of evidence from these observational studies (noted in the last paragraph on page 20), which is rated as low to very low, the most appropriate course of action that would be to not undertake meta-analysis with data from these studies.

Response: We agree with the reviewer regarding the limitations and have highlighted in the manuscript. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter. On top of it, the second sensitivity analysis suggested by one of the reviewers suggests very different results and it worth to keep the analysis in the manuscript. Please find it below:

A second sensitivity analysis from the same eight cohort studies [26-29, 40-45], examined whether low and high risk of bias limited to “two or fewer domains rated as low risk of bias” versus “three or more domains rated as low risk of bias” differed substantially. There were substantial differences in the point estimates between the “two or fewer domains rated as low risk of bias” group (adjusted OR 0.61, 95% CI 0.49, 0.75; $p < 0.001$; $I2=0\%$) and the “three or more domains rated as low risk of bias” (adjusted OR 1.26, 95% CI 0.68, 2.33; $p=0.46$; $I2=51\%$), with an interaction p-value of 0.03 (Appendix Figure 4).

Regarding study selection
The RCT, Adriaens 2014 does not technically meet the inclusion criteria. Whilst people were followed up for 8 months, a comparison group (no e-cigarettes) is only provided for 3 months and then this group was provided with e-cigarettes.

Response: Adriaens provided us with data for 8 months throughout contact by e-mail.

I don’t think that Vickerman 2013 can be considered as a prospective cohort study. This study surveyed people seven months after they enrolled into a smoking cessation service (Quitline), asking them about their retrospective use of e-cigarettes. It is in my view a cross-sectional survey and does not provide any useful information about the effects of e-cigarette use on smoking cessation.

Response: The reviewer completely discounts participants’ report of their e-cigarette use. We think this is too harsh an assessment of individuals’ ability to remember whether or not they were using e-cigarettes.

Abstract
ENDS and ENNDS abbreviations need to be written fully before used first.
The abstract should be updated to reflect the above suggested changes

Response: Many thanks; we've written fully the abbreviations at first used.

Methods
The type of abstinence is not defined (e.g. point-prevalence, continuous abstinence). Where studies reported multiple outcomes, which was preferred?

Response: No included study reported both types of abstinence.

Discussion
Whilst the discussion reflects the current results, it needs to be updated to reflect the changes suggested above.

Response: We added on the discussion the following:

"Although this review presents several limitations, the issue is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter. "

We also added under discussion our preference about random effect model:

"A final statistical limitation is that we calculated differences from 6 to 12 months of follow-up. Absolute differences may differ across this time frame and constitute a source of variability. Moreover, there are three schools of thought with respect to use of fixed and random effect models: those who prefer always to use fixed effects, those who prefer (almost) always random effects, and those who would choose fixed and random depending on the degree of heterogeneity. Each argument has its proponents within the statistical community. The argument in favor of the second rather than the third is a) there is always some heterogeneity, so any threshold of switching models is arbitrary and b) when there is little heterogeneity, fixed and random yield similar or identical results, so one might as well commit oneself to random from the start. We find these two arguments compelling; thus, our choice."

VERSION 2 – REVIEW

REVIEWER	Konstantinos Farsalinos Onassis Cardiac Surgery Center University of Patras 2 unpublished studies performed using unrestricted funds provided to the institution (Onassis Cardiac Surgery Center) by e-cigarette companies in 2013 (more than 3 years ago). Two studies funded by the non profit association AEMSA in 2013 (less than 3 years ago)and one study funded by the non-profit

	association TSFA in 2015.
REVIEW RETURNED	18-Sep-2016

GENERAL COMMENTS	General comments. I am satisfied with the response of the authors and with the revisions and addition of a secondary sensitivity analysis. My main concern (and reason to ask for revision) is related to the Harrington study. The numbers mentioned in the cited abstract do not match with the numbers in the present analysis. See all my comments in detail below. To facilitate reading, I have included my 1st review comments followed by the authors response and then my 2nd review comments. 1. Review#1: This is one of the several (4 cited by the author) reviews on the effects of ENDS on smoking cessation. The authors have done an extensive and very detailed analysis, and they should be praised for that. I am not sure if this review adds anything new to the current knowledge. Despite that, there is a potential to add the message: the studies performed until now are totally inappropriate to make any conclusions about the effects of ENDS on smoking cessation. The studies included to the analysis are, in most cases totally inappropriate. The authors mention in the discussion section: “One could argue that these limitations make the pooling of results we have undertaken inadvisable. On the other hand, the pooling does highlight the possibility of an adverse effect of e-cigarettes on quit rates, a possibility that until definitively refuted by randomized trials needs consideration in policy debates regarding e-cigarettes”. In my opinion, analyzing these studies is not only uninformative but can be highly misleading. Thus, they should be probably ignored rather than considered in any policy debate. I do not see a purpose of trying to generate data on ENDS efficacy in smoking cessation while such data do not really exist. Authors’ Response: We agree with the reviewer that the studies are very biased and have highlighted in the manuscript. That it's what we found out there in the literature, and we are trying to show to the scientific community that nothing can be concluded so far related to the effectiveness of ENDS and/or ENNDS. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter. Review#2: It seems obvious that there is no disagreement, and of course the authors of this review cannot be criticized for the studies they have included. I do not suggest dismissing them but I do suggest making clear statements about their usefulness (or

	lack of usefulness) in addressing the question of e-cigarettes' efficacy on smoking reduction and cessation. I think the revised manuscript is much clearer in this aspect. It is also a good opportunity for the authors to discuss some recommendations on how future studies should appropriately address the question. 2. Review#1: There are several very important problems in the studies included, some of which are highlighted by the authors. Selection of studies. There is an issue with the selection of studies that were analyzed. The authors do present the problem of the outcome being present at the start of the study. For example, reference 44 (Vickerman et al.) was a study of a quitline service. The study did not assess whether ENDS were used before the subjects called the quitline services (7 months before the survey was performed). Of course, this is mentioned herein, but it represents a very serious problem because it is likely that some subjects were by definition "failures to quit" at baseline. This does not really adherent to the principles of a cohort study evaluating the impact of ENDS on smoking cessation because the outcome may have existed since the baseline. In fact, the authors released a press statement saying that: "The recently published article by Dr Katrina Vickerman and colleagues has been misinterpreted by many who have written about it. It was never intended to assess the effectiveness of the e-cig as a mechanism to quit." (unfortunately, the press statement has been removed from their website, but it was available until June 2015 – See Farsalinos & Le Houezec, Risk Manag Healthc Policy. 2015; 8: 157–167). Response: We agree with the reviewer regarding the limitations and have highlighted in the manuscript. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter. Review#1: The same issue is present in the study by Manzoli et al. (reference 42), although in that case there is absolute certainty that the outcome existed from the beginning. Dual users were by definition "failures to quit" at baseline.

	Response: Our view is, once again, rather than dismissing altogether, the study should be considered in light of its limitations. Review#1: The study by Al Delaimy et al. (reference 40) is even worse, with several methodological issues. First, the comparison was between subjects having used ENDS and subjects who will never use ENDS. The methodology section mentions: “The main predictor was the use of e-cigarettes, which, for the purposes of the current study, we categorized as (1) a report of “will never use e-cigarettes” at baseline and follow-up (reference group in regression models) and (2) a report of “have used e-cigarettes” at baseline and follow-up”. So, technically, the study was not a comparison between ENDS users vs. non-users. Moreover, you mention that there were 628 participants, but Al-Delaimy et al. present the analysis of 368 participants (table 4, 191 who have used ENDS vs. 177 who will never use). Moreover, you mention that there was no bias of the outcome being present at the start of the study, but Al-Delaimy et al. mention that the group was subjects with “a report of “have used e-cigarettes” at baseline and follow-up”. The title is characteristic of this: “E-Cigarette Use in the Past and Quitting Behavior in the Future”. I think the assessment of bias and other information needs to be corrected for this study. Response: 628 were the number at baseline. Table 1. “Study characteristics related to design of study, setting, number of participants, mean age, gender, inclusion and exclusion criteria, and follow-up.” presents the number of baseline, not the number of patients in the final analysis. With regards the risk assessment, we have corrected it in Figure 4 showing high risk of bias for the domain “can we be confident that the outcome of interest was not present at start of study?” Review#2: To all the above comments, I would like to clarify that I do not suggest dismissing the studies but presenting clearly their major limitations, which in my opinion make these studies inappropriate to provide any reliable response to the review purpose. Concerning the study by Al-Delaimy et al., i have some comments. You define outcome bias as smoking cessation being present at the start of the study. I have presented the problem that failure to quit may have been already present at baseline (the study may have recruited subjects which have already failed for the outcome assessed). How do you present this bias? Is this also included in the outcome bias? In figure 4 you now present an outcome bias for Al-Delaimy et al., but in appendix figure 3 and in table 5 it is presented as having no such bias. Please correct for consistency, and also clarify how you considered the bias of “failed outcome” at baseline. Finally, I should note that I

	consider inappropriate in a cohort study to define an exclusion criterion at follow up (those reporting “might use e-cigarettes” at follow up). 3. Review#1: Finally, to the best of my knowledge, the study by Harrington et al. (reference 45) is available only as a conference abstract. If I am correct that this is just an abstract, I do not think it is appropriate to include this in the analysis because I am unable to evaluate the findings. How were the ORs calculated? There is nothing mentioned in the abstract, and based on the numbers reported I have serious doubts about the ORs reported in the analysis. This study should be removed. In case I am wrong, please provide a link to the full text of the study. Response: Harrington study is only available as abstract and, after contact with the author by e-mail they provided us with a poster presentation. Although it is abstract we did not state in our manuscript that we would exclude conference abstracts. Usually for systematic reviews we considered all type of published and unpublished data. We do not exclude studies due to be an abstract. Please find attached the poster the authors sent to us. Review#2: I could not find the attached file of the poster of the Harrington study. I contacted the editorial office and they responded that there was no attachment in your response. In any case, I accept that you want to include a conference abstract in the analysis. However, there are some problems in the presentation of the results. In table 2, you mention the study included 171 ENDS users and 759 non users. The total is 930 subjects. But in the abstract it is mentioned that “979 smokers hospitalized at a tertiary care medical center were recruited… 823 (84.1%) participants provided data at both time points”. In table 3, you mention that self-reported quitters were 21/171 (12.3%) ENDS users and 62/464 (13.4%) non users (total of 635). The numbers presented here do not match with the numbers mentioned in the abstract cited (although the percentages match). Moreover, I cannot understand how an OR of 0.49 can be derived from these numbers. The OR should be very close to 1, since quitting was similarly prevalent in both groups. Finally, what happened to the additional 295 non users? Were they lost to follow up? Isn’t it strange that no ENDS users but so many non-users were lost to follow up? How were lost to follow-up subjects treated in this study? Should they be treated as failures? The abstract by Harrington et al. mentions: “Current e-cig use

	was reported by 171 (20.7%) at baseline and 246 (29.9%) at 6-month follow-up, with 98 (11.9%) reporting current e-cig use at both time points.". How were the subjects who were using e-cigs only at follow-up treated in the analysis (it seems there were 148 such cases)? As non-users, based on baseline status, or as users, based on follow-up status? Please clarify. Unless all these issues are clarified, I do not think this study should be included in the analysis. The most important issue is that the OR reported in the review is not in line with the abstract cited. If the poster contains different numbers from the abstract, then there is a question of reliability of the reported data. 4. Review#1: I am wondering if the results would change if these studies were excluded from the analysis, since it is obvious from Figure 4 that they are the worse in terms of the bias. It would be better to remove them rather than use largely irrelevant studies. Of course, in that case the number of studies included would be very small and probably inadequate for any form of analysis. Response: We have conducted a subgroup analysis comparing the four studies with which the reviewer has a particular concern (Al-Delaimy; Harrington; Manzoli; Vickerman) and we also have included Brose study (as per the risk of bias assessment figure shows that this study is also one of the worst studies) to the three other studies (Biener; Borderud; and Prochaska). Please find the new meta-analysis attached as Appendix Figure 4 and the text below: A second sensitivity analysis from the same eight cohort studies [26-29, 40-45], examined whether low and high risk of bias limited to "two or fewer domains rated as low risk of bias" versus "three or more domains rated as low risk of bias" differed substantially. There were substantial differences in the point estimates between the "two or fewer domains rated as low risk of bias" group (adjusted OR 0.61, 95% CI 0.49, 0.75; $p < 0.001$; I2=0%) and the "three or more domains rated as low risk of bias" (adjusted OR 1.26, 95% CI 0.68, 2.33; p=0.46; I2=51%), with an interaction p-value of 0.03 (Appendix Figure 4). Review#2: That was very helpful. Thank you.

	5. Review#1: In any case, it seems to me that the scientific community (referring not only to the present analysis but to all previous similar reviews) is trying to generate data and information when such data do not exist. These studies should not even be classified as cohort studies evaluating ENDS efficacy as smoking substitutes. Any analysis has only academic interest but very limited (and potentially misleading) information about regulatory or policy decisions. Therefore, I do not suggest rejecting the manuscript, for 2 reasons: 1. The authors have done a substantial and detailed work analyzing all these studies 2. It is a good opportunity to present in more depth the major problems related to the studies which, in my opinion, are only supposed to evaluate the effects of ENDS on smoking cessation but fail to do so. Response: Our disagreement with the reviewer is not fundamental – the issue is a matter of degree. We agree with the major limitations of these studies. The reviewer would have us simply point out that they are so flawed they do not bear on the issue at hand. We agree they are highly flawed, and have pointed out the limitations, but do not think they should be dismissed altogether. Review#2: If by the word “dismiss” you mean not include them in the analysis, I agree. But the point is if this metanalysis provide any valuable conclusions? The review cannot conclude if e-cigarettes are helping or are preventing smoking cessation or reduction, not because of the statistical outcome but because of the inappropriate studies which were not really designed to assess the impact of e-cigarettes on smoking cessation and reduction. In that respect, I agree with the conclusion presented in the abstract. 6. Review#1: There are other potential sources of bias that are not mentioned or discussed in the manuscript. The authors mention that they assessed the impact of ENDS vs. no smoking cessation aid or alternative smoking cessation aid, regardless of whether the users were using them as part of a quit

	attempt. The problem is that efforts to quit “cold turkey” (no aid) or with other aids (NRTs, oral medications, psychological support) are by definition intended and conscious quit attempts. With ENDS there is a lot of experimentation for curiosity, without any real intention to quit. Thus, there is a potential for selection bias, the populations compared are not very similar in terms of intention to quit. An added problem is that experimentation with any smoking-cessation aid is unlikely to result in smoking cessation. Thus, this is further extending to the above mentioned bias. Another type of selection bias is related to the dependence on smoking of the populations compared. Usually the FTCD is used, however, studies of population samples have shown that other tests (like Strength of Urges To Smoke) are stronger predictors of successful cessation (see: Fidler et al. Addiction 2011, Fidler & West, Drug Alcohol Depend 2011, Kozlowski et al., Drug Alcohol Depend 1994). Did the authors assess if the dependence scales used in different studies (if available) were appropriate? Response: We added the issue about potential for selection bias under “Strengths and limitations” of the review under discussion. Please, find it below: “Finally, another limitation of the observational studies in this review is the potential for selection bias as the populations compared differ in terms of intention to quit.” We did not assess if the dependence scales used in different studies were appropriate because there was no information provided by the included studies about them. Review#2: Thank you. 7. Review#1: Considering that smoking cessation is more a behavioral change than the therapy of a disease (despite the WHO classification of smoking as a disease), I consider the above limitations extremely important. Added to that, the problem with RCTs is the inability to select products based on self-preference. There is a lot of discussion in the literature showing that ENDS are chosen based on self-preference and satisfaction. This cannot be addressed in a conventional RCT which follows the principles of RCTs for pharmaceuticals. Another limitation not mentioned in the study is the use of outdated products, especially in the RCTs by Bullen et al. and Caponnetto et al. The products were outdated and off the market by the time the study were published, and that was due to the development of better products. The studies were published more than 3 years ago, and were probably initiated 5 years ago. Thus, how can someone

	support the relevance of the findings for today, considering the fast development (I would say, complete transformation) of the ENDS products? In fact, considering the dates of the studies, it is highly likely that almost all studies have evaluated what are now considered outdated products. Response: Thanks for the valuable suggestions, we have added the following: Regarding the issue of self-selection, the reviewer has compellingly made the point regarding selection bias. If one compares those who choose to use e-cigarettes versus those who don't, it is very likely that any differences will be due to the nature of the populations rather than e-cigarettes (i.e. prognostic differences). This issue can be dealt with only in the context of an RCT, which, ideally, would enroll individuals open to using e-cigarettes. This openness would deal, at least to some extent, with the reviewer's concern. Regarding the earlier versions of e-cigarettes tested in the RCTs, we have highlighted this issue in the revised manuscript as follows: "Furthermore, in all these RCTs, the ENDS tested were earlier generation; it is possible that later generation of e-cigarettes would have greater benefit." Review#2: Thank you. I consider these comments appropriate and valuable. I would use the following sentence as: "it is unknown whether providing later generation of e-cigarettes or a realistic scenario of allowing users to choose e-cigarettes based on self-preference would have greater benefit." 8. Review#1: My suggestion for the manuscript is to perform major revisions. These should include the more extensive discussion of the limitations of the studies included in the analysis (based on my comments above), the exclusion of some studies which are particularly misleading, and a clear presentation of the conclusion that current evidence is inappropriate to assess the impact of ENDS on smoking cessation or reduction. I think the correct term is "inappropriate", not "insufficient". These limitations and the conclusion should also be presented in the abstract. Response: We have re-worded the text to reflect that the results failed to show a difference between the comparison groups.

	Review#2: In my opinion, a more appropriate statement as a conclusion (or in the discussion section) would be that due to the limitations of the studies included in this analysis it is impossible to make any reliable conclusions on whether e-cigarette use promotes, has no effect or hinders smoking cessation. “Failed to show a difference” implies that the effects are similar; I think the data are so inconclusive (mainly due to the limitations of the studies analyzed) that we cannot be sure of that. In the results section were you present separately the results of each study, the statement “failed to show a difference” is OK. Minor comments 9. Review#1: Appendix Figure 4. Why is the study by Hajek et al. missing from the analysis? It is 1 of the 2 cohort studies which were not rated as high risk of bias for limitations in matching exposed and unexposed groups or adjusting analysis for prognosis variables. I think it should be included, despite the low sample size. Response: The Hajek study was not included in the meta-analysis because there was no raw data provided by the study to enable us to include it in the meta-analysis. Review#2: Thank you for the clarification 10. Review#1: Page 11, lines 18-23. Reference 46 (Hajek et al.) is mentioned as both supplying and not supplying the requested data. Response: The Hajek study provided some but not all of the missing data. Review#2: Thank you for the clarification Further comments

	Since the outcome analyzed included smoking reduction and adverse effects (besides cessation), I think it would be appropriate to add a sentence about these in the abstract (unless there is a word limitation). Page 25, lines 16-25. I think another limitation of the cohort studies was the non-assessment of experimentation vs. regular use. I agree with the statement in lines 39-45 that regular use may indicate more motivation. However, it should be noted that motivation to quit smoking is a major determinant of success in all cases and irrespective of the aid used. Page 25, lines 49-52. “Such individuals may cohort studies may already be failing in their attempts to stop smoking.” Please erase the “may cohort studies”. Page 27, lines 47-54. “This is an important finding, and raises serious questions regarding the importance of thee behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine.” This is a strong statement. I think the sensory and behavioral aspects of smoking dependence are well established. Moreover, the nicotine delivery of e-cigarettes is directly related to nicotine concentration and aerosol yield. Both these factors influence the sensory aspects (more nicotine concentration produces more throat hit, more aerosol yield produces more throat hit and more flavor), so it will be difficult to separate the nicotine effects from the sensory aspects. Please correct the word “thee” Also, the statement “The available evidence, of low or very low quality, provides no support for the hypothesis that….” (lines 41-43) could be revised as: “The available evidence, of low or very low quality, can neither verify nor exclude the hypothesis that ….”.

REVIEWER	Kristian Filion McGill University, Canada
REVIEW RETURNED	09-Sep-2016

GENERAL COMMENTS	This systematic review and meta-analysis of electronic nicotine and non-nicotine delivery systems is a resubmission. The authors

	have addressed many comments raised in previous reviews, resulting in an improved manuscript. However, some issues remain. The most important issue is that all three reviewers raised concerns about the appropriateness of pooling these data. The authors have argued that doing so is appropriate “bearing in mind the limitations” of these data. They have also added a sensitivity analysis that demonstrates that the estimates vary with study quality. Given the clinical heterogeneity present in these studies and the results of their sensitivity analysis, the interpretability of the pooled estimates is unclear. For that reason, I believe that it is more appropriate to systematically review this literature without meta-analysis, explicitly discussing the sources of heterogeneity and limitations of these data. The results of the systematic review represent a far more important contribution to the literature than the pooled estimates. A second issue is the handling of patients who were lost-to-follow-up. As discussed in previous reviews, convention in this area is to assume that such patients have returned to smoking. While the authors have included a ‘worst case’ scenario sensitivity analysis when the results of the primary analysis achieved or approached statistical significance, the use of a complete case scenario is still an important limitation. The potential bias associated with this approach make it difficult to assess both the results of the individual studies in the systematic review of these data as well as the results of the meta-analysis. At a minimum, a sensitivity analysis in which patients who are lost-to-follow-up are considered to have returned to smoking should be included.

REVIEWER	Hayden McRobbie Queen Mary University of London, UK
REVIEW RETURNED	04-Sep-2016

GENERAL COMMENTS	Unfortunately, I do not feel that the major points I raised were adequately addressed. The authors justify their approach to missing data by arguing that there is no perfect approach and that their sensitivity analysis was sufficient. I remain skeptical about the approach used. As stated by West et al 20051 “Experience from past trials shows that smokers who claim abstinence but do not attend for biochemical verification or who cannot be contacted through the various means at the investigator’s disposal almost always turn out to be smoking. Therefore, classifying them as non-smokers, or leaving them out of the analysis, for instance, would lead to greater bias.” If the authors want to include complete data only in the meta-analysis, I would have thought that conducting a sensitivity analysis that considers those LTFU as smokers should have been included. Regardless of my views here, the sensitivity analysis on page 105 shows quit rates in the Bullen study of 25/289 in the intervention group (calculated by 21 + 3.5 [48 LTFU x 7.3%]) and 6/73 in the control group. When I calculate the number I seem to get 5 (calculated by 3 + 1.96 [16 LTFU x (3 x

	4.1%)]. Can the authors please check this? Both myself and another reviewer question that appropriateness of pooling the data from the cohort studies. The authors agree with the limitations and although they highlight these in the paper they still pool the data in a meta-analysis. Regarding the Adriaens study, I think that the authors misunderstood the point I was making. It wasn't the fact that there were not 8-month follow-up data, it was the point that all people in the control group were given e-cigarettes at 3-months, so comparative data was only available up to 3 months. It was the study design that was problematic, not the final follow-up period. The authors responded to my question about the definition of abstinence by stating that no included study reported both types of abstinence. This is not the case. Bullen et al 2013 provides both continuous and 7-day point prevalence. Regarding the abstract, despite the fact that this will be an open access paper, the abstract may be all that some people read. Although you included three RCTs, the results come from just two of these. This should be made clear. As the authors know from my initial comments I think that there are serious problems of including the cohort studies in a meta-analysis. In the strengths and limitations section you state that "The limitations of the cohort studies… from which no credible inferences can be drawn". A statement to this effect needs to be included in the abstract. I also just want to comment on your implications (apologies for missing this last time) where the authors state "The available evidence, of low or very low quality, provides no support for the hypothesis that, because they address not only nicotine addiction but also potentially deal with behavioural and sensory aspects of cigarette use, ENDS may be more effective than other nicotine replacement strategies. This is an important finding, and raises serious questions regarding the importance of thee behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine." I don't understand the point the authors are making here. So far, there are data from only one RCT that examined the efficacy of ENDS versus a nicotine patch. Both nicotine containing EC and non-nicotine EC provide behavioural and sensorimortor aspects. The comments made in this section are out of line with their findings. There are now over 10 systematic reviews on electronic cigarettes. Whilst I appreciate the work that has gone into undertaking this review and meta-analysis, my view is that its still contains a number of methodological problems that lead to erroneous conclusions. Finally, I do not think that it add significantly to what is already published.

	1. West, R., Hajek, P., Stead, L. & Stapleton, J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction 100, 299–303 (2005).

VERSION 2 – AUTHOR RESPONSE

All comments by reviewers have been addressed, with corresponding changes made directly to the manuscript in yellow color. We also copied and pasted the text below each comment. Accompanying this letter, please find a revised version of our manuscript.

We also acknowledge the peer reviewers for their great inputs during revisions in the manuscript.

Detailed responses to the reviewers are included below in blue color follows peer reviewers review comments.

General comments.

I am satisfied with the response of the authors and with the revisions and addition of a secondary sensitivity analysis. My main concern (and reason to ask for revision) is related to the Harrington study. The numbers mentioned in the cited abstract do not match with the numbers in the present analysis. See all my comments in detail below.

To facilitate reading, I have included my 1st review comments followed by the authors response and then my 2nd review comments.

Authors' Response#2: We appreciate the opportunity to resubmit our article "Electronic nicotine delivery systems and/or electronic non-nicotine delivery systems for tobacco smoking cessation or reduction: a systematic review and meta-analysis" (bmjopen-2016-012680).

All comments by reviewers have been addressed, with corresponding changes made directly to the manuscript in yellow color. We also copied and pasted the text below each comment. Accompanying this letter, please find a revised version of our manuscript.

We also acknowledge the peer reviewers for their great inputs during revisions in the manuscript.

Detailed responses to the reviewers are included below in blue color follows peer reviewers 2nd review comments.

Best regards,

Regina El Dib

1. Review#1: This is one of the several (4 cited by the author) reviews on the effects of ENDS on smoking cessation. The authors have done an extensive and very detailed analysis, and they should be praised for that. I am not sure if this review adds anything new to the current knowledge. Despite that, there is a potential to add the message: the studies performed until now are totally inappropriate to make any conclusions about the effects of ENDS on smoking cessation. The studies included to the analysis are, in most cases totally inappropriate.

The authors mention in the discussion section: “One could argue that these limitations make the pooling of results we have undertaken inadvisable. On the other hand, the pooling does highlight the possibility of an adverse effect of e-cigarettes on quit rates, a possibility that until definitively refuted by randomized trials needs consideration in policy debates regarding e-cigarettes”. In my opinion, analyzing these studies is not only uninformative but can be highly misleading. Thus, they should be probably ignored rather than considered in any policy debate. I do not see a purpose of trying to generate data on ENDS efficacy in smoking cessation while such data do not really exist.

Authors’ Response: We agree with the reviewer that the studies are very biased and have highlighted in the manuscript. That it's what we found out there in the literature, and we are trying to show to the scientific community that nothing can be concluded so far related to the effectiveness of ENDS and/or ENNDS. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter.

Review#2: It seems obvious that there is no disagreement, and of course the authors of this review cannot be criticized for the studies they have included. I do not suggest dismissing them but I do suggest making clear statements about their usefulness (or lack of usefulness) in addressing the question of e- cigarettes’ efficacy on smoking reduction and cessation. I think the revised manuscript is much clearer in this aspect. It is also a good opportunity for the authors to discuss some recommendations on how future studies should appropriately address the question.

Authors’ Response#2: We agree with the reviewer that the studies are limited and have highlighted the limitations in the manuscript. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. We believe that the community needs to know these results are there and what inferences might be possible. We think omitting these results altogether is excessively presumptuous on the part of the reviewers. We note that Reviewer 2 considers it reasonable to include all studies while noting their limitations. We added recommendations on how future studies should appropriately address the question in both abstract and discussion. Please find it highlighted in yellow color below:

Abstract

Data synthesis: Three randomized trials including 1,007 participants and nine cohort including 13,115 participants proved eligible. Results provided by the RCTs suggest a possible increase in tobacco smoking cessation with ENDS in comparison to ENNDS (RR 2.03, 95% CI 0.94, 4.38; p = 0.07; I^2=0%, risk difference (RD) 64/1,000 over 6 to 12 months, low certainty evidence). Results from cohort studies suggested a possible reduction in quit rates with use of ENDS compared to no use of ENDS (OR 0.74, 95% CI 0.55, 1.00; p = 0.051; I^2=56%, very low certainty).

Conclusions: There is very limited evidence regarding the impact of ENDS or ENNDS on tobacco smoking cessation or reduction: data from RCTs are of low certainty and observational studies of very low certainty. This review underlines the need to conduct well-designed randomized trials measuring biochemically validated outcomes and adverse effects.

Discussion

Implications

Existing smoking reduction aids such as nicotine replacement therapy are effective, but their impact is limited: the proportion of those who quit when using these aids remains small. The available evidence, of low or very low quality, provides no support for the hypothesis that, because they address not only nicotine addiction but also potentially deal with behavioural and sensory aspects of cigarette use, ENDS may be more effective than other nicotine replacement strategies. This is an important finding, and raises serious questions regarding the importance of thee behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine. It is possible that type of ENDS or dose of exposure may influence quit rates, and that newer models may be more effective, but there is insufficient data to provide insight into these issues. This review underlines the need to conduct well-designed randomized trials measuring biochemically validated outcomes and adverse effects and accompanied by complete and transparent reporting of methods and results.

2. Review#1: There are several very important problems in the studies included, some of which are highlighted by the authors.

Selection of studies. There is an issue with the selection of studies that were analyzed. The authors do present the problem of the outcome being present at the start of the study. For example, reference 44 (Vickerman et al.) was a study of a quitline service. The study did not assess whether ENDS were used before the subjects called the quitline services (7 months before the survey was performed). Of course, this is mentioned herein, but it represents a very serious problem because it is likely that some subjects were by definition “failures to quit” at baseline. This does not really adherent to the principles of a cohort study evaluating the impact of ENDS on smoking cessation because the outcome may have existed since the baseline. In fact, the authors released a press statement saying that: “The recently published article by Dr Katrina Vickerman and colleagues has been misinterpreted by many who have written about it. It was never intended to assess the effectiveness of the e-cig as a mechanism to quit.” (unfortunately, the press statement has been removed from their website, but it was available until June 2015 – See Farsalinos & Le Houezec, Risk Manag Healthc Policy. 2015; 8: 157–167).

Response: We agree with the reviewer regarding the limitations and have highlighted in the manuscript. The issue of possible contention is whether one should dismiss these results entirely, or consider them bearing in mind the limitations. The latter represent our view of the matter.

Review#1: The same issue is present in the study by Manzoli et al. (reference 42), although in that case there is absolute certainty that the outcome existed from the beginning. Dual users were by definition “failures to quit” at baseline.

Response: Our view is, once again, rather than dismissing altogether, the study should be considered in light of its limitations.

Review#1: The study by Al Delaimy et al. (reference 40) is even worse, with several methodological issues. First, the comparison was between subjects having used ENDS and subjects who will never use ENDS. The methodology section mentions: “The main predictor was the use of e-cigarettes, which, for the purposes of the current study, we categorized as (1) a report of “will never use e-cigarettes” at baseline and follow-up (reference group in regression models) and (2) a report of “have used e- cigarettes” at baseline and follow-up”. So, technically, the study was not a comparison between ENDS users vs. non-users. Moreover, you mention that there were 628 participants, but Al-Delaimy et al. present the analysis of 368 participants (table 4, 191 who have used ENDS vs. 177 who will never use). Moreover, you mention that there was no bias of the outcome being present at the start of the study, but Al-Delaimy et al. mention that the

group was subjects with “a report of “have used e-cigarettes” at baseline and follow-up”. The title is characteristic of this: “E-Cigarette Use in the Past and Quitting Behavior in the Future”. I think the assessment of bias and other information needs to be corrected for this study.

Response: 628 were the number at baseline. Table 1. “Study characteristics related to design of study, setting, number of participants, mean age, gender, inclusion and exclusion criteria, and follow-up.” presents the number of baseline, not the number of patients in the final analysis. With regards the risk assessment, we have corrected it in Figure 4 showing high risk of bias for the domain “can we be confident that the outcome of interest was not present at start of study?”

Review#2: To all the above comments, I would like to clarify that I do not suggest dismissing the studies but presenting clearly their major limitations, which in my opinion make these studies inappropriate to provide any reliable response to the review purpose.

Concerning the study by Al-Delaimy et al., i have some comments. You define outcome bias as smoking cessation being present at the start of the study. I have presented the problem that failure to quit may have been already present at baseline (the study may have recruited subjects which have already failed for the outcome assessed). How do you present this bias? Is this also included in the outcome bias? In figure 4 you now present an outcome bias for Al-Delaimy et al., but in appendix figure 3 and in table 5 it is presented as having no such bias. Please correct for consistency, and also clarify how you considered the bias of “failed outcome” at baseline. Finally, I should note that I consider inappropriate in a cohort study to define an exclusion criterion at follow up (those reporting “might use e-cigarettes” at follow up).

Authors’ Response#2: We corrected figure 4 from high risk of bias to low risk of bias in the domain “Can we be confident that the outcome of interest was not present at start of study?” matching appendix figure 3 and in table 5. Related to the bias of “failed outcome” at baseline, this is not addressed in the risk of bias instrument we used. Therefore, we have added this as an additional consideration in the discussion as follows:

“These studies had a number of limitations: an unknown number of these participants were not using ENDS as a cessation device; some were not using ENDS during a quit attempt; many did not have immediate plans to quit smoking; and some may have already failed attempts to stop smoking.”

3. Review#1: Finally, to the best of my knowledge, the study by Harrington et al. (reference 45) is available only as a conference abstract. If I am correct that this is just an abstract, I do not think it is appropriate to include this in the analysis because I am unable to evaluate the findings. How were the ORs calculated? There is nothing mentioned in the abstract, and based on the numbers reported I have serious doubts about the ORs reported in the analysis. This study should be removed. In case I am wrong, please provide a link to the full text of the study.

Response: Harrington study is only available as abstract and, after contact with the author by e-mail they provided us with a poster presentation. Although it is abstract we did not state in our manuscript that we would exclude conference abstracts. Usually for systematic reviews we considered all type of published and unpublished data. We do not exclude studies due to be an abstract. Please find attached the poster the authors sent to us.

Review#2: I could not find the attached file of the poster of the Harrington study. I contacted the editorial office and they responded that there was no attachment in your response. In any case, I accept that you want to include a conference abstract in the analysis. However, there are some problems in the presentation of the results. In table 2, you mention the study included 171 ENDS users and 759 non users. The total is 930 subjects. But in the abstract it is mentioned that “979 smokers hospitalized at a tertiary care medical center were recruited... 823 (84.1%) participants provided data at both time

points”. In table 3, you mention that self-reported quitters were 21/171 (12.3%) ENDS users and 62/464 (13.4%) non users (total of 635). The numbers presented here do not match with the numbers mentioned in the abstract cited (although the percentages match). Moreover, I cannot understand how an OR of 0.49 can be derived from these numbers. The OR should be very close to 1, since quitting was similarly prevalent in both groups. Finally, what happened to the additional 295 non users? Were they lost to follow up? Isn’t it strange that no ENDS users but so many non-users were lost to follow up? How were lost to follow-up subjects treated in this study? Should they be treated as failures?

The abstract by Harrington et al. mentions: “Current e-cig use was reported by 171 (20.7%) at baseline and 246 (29.9%) at 6-month follow-up, with 98 (11.9%) reporting current e-cig use at both time points.”. How were the subjects who were using e-cigs only at follow-up treated in the analysis (it seems there were 148 such cases)? As non-users, based on baseline status, or as users, based on follow-up status? Please clarify.

Unless all these issues are clarified, I do not think this study should be included in the analysis. The most important issue is that the OR reported in the review is not in line with the abstract

cited.

If the poster contains different numbers from the abstract, then there is a question of reliability of the reported data.

Authors' Response#2: Our protocol stated that we would include evidence from abstracts and would attempt to contact authors for additional information. The Harrington study is only available as abstract and, after contact with the authors by e-mail; they provided us with a poster presentation. Please find attached the poster the authors sent to us which we think provides sufficient information to include the study. Not doing so would introduce the issue of publication bias. Please find attached the poster of the Harrington study at the end of this file. One major problem we faced with these studies was the inconsistency of information in the publication itself to build the "puzzle" and, therefore we contacted the authors to gather precise answers. With regard to the lost of follow-up we rated it as high risk of bias for the Harrington study consequent on the additional 295 non users. For incomplete outcome data in individual studies (both RCTs and prospective cohort studies) we stipulated as low risk of bias for loss to follow-up of less than 10% and a difference of less than 5% in missing data between intervention/exposure and control groups. Furthermore, we conducted a plausible worst-case sensitivity analysis in which all participants with missing data from the arm of the study with the lower quit rates were assumed to have 3 times the quit rate as those with complete data, and those with missing data from the other arm were assumed to have the same quit rate as participants with complete data.

With regards the OR of 0.49 from Harrington 2015 study we get it through the following:

12.2% e-cig users (12.2% of 171 = 21) – 21/171

13.4% non-users (13.4% of (823-171) = 108) – 87/652

However, from Regina's contact with the authors, we used 62/464 for non-users.

From Author: "If you want to know if using e-cigs during the 6 months after baseline was predictive of smoking abstinence at 6 months, the rates are: 24/246 and 85/468 (complete cases)".

4. Review#1: I am wondering if the results would change if these studies were excluded from the analysis, since it is obvious from Figure 4 that they are the worse in terms of the bias. It would be better to remove them rather than use largely irrelevant studies. Of course, in that case the number of studies included would be very small and probably inadequate for any form of analysis.

Response: We have conducted a subgroup analysis comparing the four studies with which the reviewer has a particular concern (Al-Delaimy; Harrington; Manzoli; Vickerman) and we also have included Brose study (as per the risk of bias assessment figure shows that this study is also one of the worst studies) to the three other studies (Biener; Borderud; and Prochaska). Please find the new meta-analysis attached as Appendix Figure 4 and the text below:

A second sensitivity analysis from the same eight cohort studies [26-29, 40-45], examined whether low and high risk of bias limited to “two or fewer domains rated as low risk of bias” versus “three or more domains rated as low risk of bias” differed substantially. There were substantial differences in the point estimates between the “two or fewer domains rated as low risk of bias” group (adjusted OR 0.61, 95% CI 0.49, 0.75; $p < 0.001$; $I2=0\%$) and the “three or more domains rated as low risk of bias” (adjusted OR 1.26, 95% CI 0.68, 2.33; $p=0.46$; $I2=51\%$), with an interaction p-value of 0.03 (Appendix Figure 4).

Review#2: That was very helpful. Thank you.

Authors’ Response#2: You are very welcome!

5. Review#1: In any case, it seems to me that the scientific community (referring not only to the present analysis but to all previous similar reviews) is trying to generate data and information when such data do not exist. These studies should not even be classified as cohort studies evaluating ENDS efficacy as smoking substitutes. Any analysis has only academic interest but very limited (and potentially misleading) information about regulatory or policy decisions. Therefore, I do not suggest rejecting the manuscript, for 2 reasons:

1. The authors have done a substantial and detailed work analyzing all these studies 2. It is a good opportunity to present in more depth the major problems related to the studies

which, in my opinion, are only supposed to evaluate the effects of ENDS on smoking cessation but fail to do so.

Response: Our disagreement with the reviewer is not fundamental – the issue is a matter of degree. We agree with the major limitations of these studies. The reviewer would have us simply point out that they are so flawed they do not bear on the issue at hand. We agree they are highly flawed, and have pointed out the limitations, but do not think they should be dismissed altogether.

Review#2: If by the word “dismiss” you mean not include them in the analysis, I agree. But the point is if this metanalysis provide any valuable conclusions? The review cannot conclude if e-cigarettes are helping or are preventing smoking cessation or reduction, not because of the statistical outcome but because of the inappropriate studies which were not really designed to assess the impact of e-cigarettes on smoking cessation and reduction. In that respect, I agree with the conclusion presented in the abstract.

Authors’ Response#2: We agree with the reviewer, thank you.

6. Review#1: There are other potential sources of bias that are not mentioned or discussed in the manuscript.

The authors mention that they assessed the impact of ENDS vs. no smoking cessation aid or alternative smoking cessation aid, regardless of whether the users were using them as part of a quit attempt. The problem is that efforts to quit “cold turkey” (no aid) or with other aids (NRTs, oral medications, psychological support) are by definition intended and conscious quit attempts. With ENDS there is a lot of experimentation for curiosity, without any real intention to quit. Thus, there is a potential for selection bias, the populations compared are not very similar in terms of intention to quit. An added problem is that experimentation with any smoking-cessation aid is unlikely to result in smoking cessation. Thus, this is further extending to the above mentioned bias. Another type of selection bias is

related to the dependence on smoking of the populations compared. Usually the FTCD is used, however, studies of population samples have shown that other tests (like Strength of Urges To Smoke) are stronger predictors of successful cessation (see: Fidler et al. Addiction 2011, Fidler & West, Drug Alcohol Depend 2011, Kozlowski et al., Drug Alcohol Depend 1994). Did the authors assess if the dependence scales used in different studies (if available) were appropriate?

Response: We added the issue about potential for selection bias under “Strengths and limitations” of the review under discussion. Please, find it below:

“Finally, another limitation of the observational studies in this review is the potential for selection bias as the populations compared differ in terms of intention to quit.”

We did not assess if the dependence scales used in different studies were appropriate because there was no information provided by the included studies about them.

Review#2: Thank you.

Authors’ Response#2: You are very welcome.

7. Review#1: Considering that smoking cessation is more a behavioral change than the therapy of a disease (despite the WHO classification of smoking as a disease), I consider the above limitations extremely important. Added to that, the problem with RCTs is the inability to select products based on self-preference. There is a lot of discussion in the literature showing that ENDS are chosen based on self- preference and satisfaction. This cannot be addressed in a conventional RCT which follows the principles of RCTs for pharmaceuticals. Another limitation not mentioned in the study is the use of outdated products, especially in the RCTs by Bullen et al.

and Caponnetto et al. The products were outdated and off the market by the time the study were published, and that was due to the development of better products. The studies were published more than 3 years ago, and were probably initiated 5 years ago. Thus, how can someone support the relevance of the findings for today, considering the fast development (I would say, complete transformation) of the ENDS products? In fact, considering the dates of the studies, it is highly likely that almost all studies have evaluated what are now considered outdated products.

Response: Thanks for the valuable suggestions, we have added the following:

Regarding the issue of self-selection, the reviewer has compellingly made the point regarding selection bias. If one compares those who choose to use e-cigarettes versus those who don't, it is very likely that any differences will be due to the nature of the populations rather than e-cigarettes (i.e. prognostic differences). This issue can be dealt with only in the context of an RCT, which, ideally, would enroll individuals open to using e-cigarettes. This openness would deal, at least to some extent, with the reviewer's concern.

Regarding the earlier versions of e-cigarettes tested in the RCTs, we have highlighted this issue in the revised manuscript as follows: "Furthermore, in all these RCTs, the ENDS tested were earlier generation; it is possible that later generation of e-cigarettes would have greater benefit."

Review#2: Thank you. I consider these comments appropriate and valuable. I would use the following sentence as: "it is unknown whether providing later generation of e-cigarettes or a realistic scenario of allowing users to choose e-cigarettes based on self-preference would have greater benefit."

Authors' Response#2: Thank you, we added the sentence suggested by the reviewer:

Furthermore, in all these RCTs, the ENDS tested were earlier generation; it is unknown whether providing later generation of e-cigarettes or a realistic scenario of allowing users to choose e-cigarettes based on self-preference would have greater benefit.

8. Review#1: My suggestion for the manuscript is to perform major revisions. These should include the more extensive discussion of the limitations of the studies included in the analysis (based on my comments above), the exclusion of some studies which are particularly misleading, and a clear presentation of the conclusion that current evidence is inappropriate to assess the impact of ENDS on smoking cessation or reduction. I think the correct term is “inappropriate”, not “insufficient”. These limitations and the conclusion should also be presented in the abstract.

Response: We have re-worded the text to reflect that the results failed to show a difference between the comparison groups.

Review#2: In my opinion, a more appropriate statement as a conclusion (or in the discussion section) would be that due to the limitations of the studies included in this analysis it is impossible to make any reliable conclusions on whether e-cigarette use promotes, has no effect or hinders smoking cessation. “Failed to show a difference” implies that the effects are similar; I think the data are so inconclusive (mainly due to the limitations of the studies analyzed) that we cannot be sure of that. In the results section were you present separately the results of each study, the statement “failed to show a difference” is OK.

Authors’ Response#2: Thank you very much. We have added the statement the reviewer suggested under implications/conclusions section:

Implications

Existing smoking reduction aids such as nicotine replacement therapy are effective, but their impact is limited: the proportion of those who quit when using these aids remains small. The available evidence, of low or very low quality, can neither verify nor exclude the hypothesis that, because they address not only nicotine addiction but also potentially deal with behavioural and sensory aspects of cigarette use, ENDS may be more effective than other nicotine replacement strategies. This is an important finding, and raises questions regarding the how effective it may be addressing the behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine, though teasing out the nicotine effects from sensory aspects is likely to be challenging. It is possible that type of ENDS or dose of exposure may influence quit rates, and that newer models may be more effective, but there is insufficient data to provide insight into these issues. Lack of

usefulness with regard to address the question of e- cigarettes' efficacy on smoking reduction and cessation was largely due to poor reporting.

Therefore, due to the limitations of the studies included in this analysis it is impossible to make strong inferences regarding whether e-cigarette use promotes, has no effect or hinders smoking cessation.This review underlines the need to conduct well-designed trials in this field measuring biochemically validated outcomes and adverse effects.

Minor comments

9. Review#1: Appendix Figure 4. Why is the study by Hajek et al. missing from the analysis? It is 1 of the 2 cohort studies which were not rated as high risk of bias for limitations in matching exposed and unexposed groups or adjusting analysis for prognosis variables. I think it should be included, despite the low sample size.

Response: The Hajek study was not included in the meta-analysis because there was no raw data provided by the study to enable us to include it in the meta-analysis.

Review#2: Thank you for the clarification

Authors' Response#2: You are very welcome.

10. Review#1: Page 11, lines 18-23. Reference 46 (Hajek et al.) is mentioned as both supplying and not supplying the requested data.

Response: The Hajek study provided some but not all of the missing data.

Review#2: Thank you for the clarification

Authors' Response#2: You are very welcome!

Further comments

Since the outcome analyzed included smoking reduction and adverse effects (besides cessation), I think it would be appropriate to add a sentence about these in the abstract (unless there is a word limitation).

Authors' Response#2: Thank you, we added these words under conclusions in the abstract, but please note that our abstract is now with 300 words:

Conclusions: There is very limited evidence regarding the impact of ENDS or ENNDS on tobacco smoking cessation, reduction or adverse effects: data from RCTs are of low certainty and observational studies of very low certainty. This review underlines the need to conduct well-designed trials in this field measuring biochemically validated outcomes and adverse effects.

Page 25, lines 16-25. I think another limitation of the cohort studies was the non-assessment of experimentation vs. regular use. I agree with the statement in lines 39-45 that regular use may indicate more motivation. However, it should be noted that motivation to quit smoking is a major determinant of success in all cases and irrespective of the aid used.

Authors' Response#2: Thanks. We have added the statement the reviewer suggested under discussion:

An alternative interpretation, however, is that those that used ENDS daily were more motivated to stop smoking, and the increased motivation, rather than daily use of ENDS, was responsible for their degree of success. It is worth to mention that motivation to quit smoking is a major determinant of success regardless of the aid used.

Page 25, lines 49-52. "Such individuals may cohort studies may already be failing in their attempts to stop smoking."

Please erase the "may cohort studies".

Authors' Response#2: Erased, thanks:

Such individuals may already be failing in their attempts to stop smoking.

Page 27, lines 47-54. "This is an important finding, and raises serious questions regarding the importance of thee behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine."

This is a strong statement. I think the sensory and behavioral aspects of smoking dependence are well established. Moreover, the nicotine delivery of e-cigarettes is directly related to nicotine concentration and aerosol yield. Both these factors influence the sensory aspects (more nicotine concentration produces more throat hit, more aerosol yield produces more throat hit and more flavor), so it will be difficult to separate the nicotine effects from the sensory aspects.

Authors' Response#2: We have modified as below:

This is an important finding, and raises questions regarding the how effective it may be addressing the behavioural and sensory aspects of cigarette use in their addictive potential. Thus, the focus of subsequent work should perhaps be on the dose and delivery of nicotine, though teasing out the nicotine effects from sensory aspects is likelhy to be challenging.

Please correct the word “thee”

Authors’ Response#2: Corrected, thanks:

“This is an important finding, and raises serious questions regarding the importance of the behavioural...”

Also, the statement “The available evidence, of low or very low quality, provides no support for the hypothesis that....” (lines 41-43) could be revised as: “The available evidence, of low or very low quality, can neither verify nor exclude the hypothesis that ....”.

Authors’ Response#2: Changed, thanks:

The available evidence, of low or very low quality, can neither verify nor exclude the hypothesis that, because they address not only nicotine addiction but also potentially deal with behavioural and sensory aspects of cigarette use, ENDS may be more effective than other nicotine replacement strategies.

Correction: *electronic nicotine delivery systems and/or electronic non-nicotine delivery systems for tobacco smoking cessation or reduction: a systematic review and meta-analysis*

El Dib R, Suzumura EA, Akl EA, *et al*. Electronic nicotine delivery systems and/or electronic non-nicotine delivery systems for tobacco smoking cessation or reduction: a systematic review and meta-analysis. *BMJ Open* 2017;7:e012680. doi: 10.1136/bmjopen-2016-012680.

The folllowing amendments were considered to the original version of this article.

The following paragraph was added in the 'Strengths and limitations' subheading under 'DISCUSSION' section: 'We usually conduct worst-case sensitivity analysis when there are significant results. However, because we noticed a possible increase in smoking cessation with ENDS (Figure 5) for cessation smoking, we have decided to conduct a worst-case sensitivity analysis to test the robustness of our findings.'

In Table 6,

- The first row should be read as 'Tobacco smoking cessation' instead of 'Mortality' and 'Reduction in cigarette use of at least 50%' instead of 'Renal insufficiency'. Also, Tobacco smoking cessation refers to OR.
- The header of eighth column should read as 'Relative risk and odds ratio (95% CI)' instead of 'Relative risk (95% CI)'.

In table 7, subheading of the seventh column should read 'Odds ratio (95% CI)' instead of 'Relative risk (95% CI)'.

The following footnote is added in both tables 6 and 7:
CI: confidence interval.

In the 'Data synthesis and statistical analysis' section under 'METHODS', the below statement has been added in the 3nd paragraph:

After calculating pooled relative effects, we also calculated absolute effects and 95% CI. For each outcome, we multiplied the pooled RR and its 95% CI by the median probability of that outcome. We obtained the median probability from the control groups of the available randomised trials. **When it is not possible, we obtained the median probability from the cohort studies**. We planned to perform separate analyses for comparisons with interventions consisting of ENDS and/or ENNDS and each type of control interventions with known different effects (no smoking cessation aid; alternative non-electronic smoking cessation aid including NRT and alternative electronic smoking cessation aid (ENDS or ENNDS)). For meta-analyses, we used 6 months data or the nearest follow-up to 6 months available.

The below statement has been added in the Acknowledgements section:

We would also like to thank Dr Aravind Gandhi Periyasamy for bringing these mistakes to our attention in order to issue an erratum rectifying.

BMJ Open 2020;**10**:e012680corr1. doi:10.1136/bmjopen-2016-012680corr1