Author's response to reviews

Title: Systematic review and meta-analysis: influence of smoking cessation on incidence of pneumonia in HIV

Authors:

Preeti De (preeti_de_licious@hotmail.co.uk)
Amanda Farley (a.c.farley@bham.ac.uk)
Nicola Lindson (n.l.lindson@bham.ac.uk)
Paul Aveyard (p.n.aveyard@bham.ac.uk)

Version: 4 Date: 24 September 2012

Author's response to reviews: see over
Nicola Petrosillo
The authors deal with an important issue: influence of smoking cessation on pneumonia in HIV infected individuals. They face the issue by using a systematic review and a meta-analysis. The result is that smoking cessation can prevent on third of pneumonia in HIV infected individuals, but not PJP. Methods and data are relevant. Of course, findings derive from the published data, and some of them are not enough sound (i.e. case-control studies). However, the authors mention all the limitations above described.

Some concerns:
1. page 6, last 3 lines. Criteria for including studies in the analysis. Are there any studies without radiographic confirmation of pneumonia? If the reply is yes, how acceptable is to include them in the analysis? Table 2 describes how studies categorised the outcome variable of pneumonia. Most studies did this well, but 3 were suboptimal. Crothers, Conley, and Burns relied on medical records to define the occurrence of pneumonia. It is highly likely that these US patients (all three studies) did have a CXR on examination but this is not reported in the paper. All three papers were in the lower quality group and excluding them did not change the results substantially, as we reported in the paper.

2. page 9: in the studies included in the analysis, was the antiretroviral treatment considered as yes or not, or also its duration was considered? This is a meta-analysis with data only at the study level. Thus studies reported patients were on anti-retroviral treatment or not. It was not possible to take duration into account. Only an individual patient meta-analysis could do that. However, many studies did report adjusting for the use of HAART in supplying adjusted estimates, though only for use/non-use and not for duration of use.

3. Did the authors find any difference according to the geographical area of the studies? We did not look for this as we did not have a strong belief that the benefits of smoking cessation would be greater in particular regions of the world. All but one small study was conducted on European and US populations so we think that there is limited scope for detecting effect modification.

4. In the reference list, some articles are mentioned more times. Thanks. We have corrected this.

Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Reviewer: Christopher Carroll
Reviewer's report:
A simple but generally good review. It's nice to read something so straightforward but potentially valuable! Very strong case for the question, and good design with control for key confounders. However, there are some major issues with the search and quality assessment, and the results could be more systematic, sequential in presentation:
current vs former smokers
Former vs never
current vs never
then, by study design, by quality etc.

Major essential revisions:
Methods:
Data sources and searches:
1. It is convention for inclusion criteria (here listed under "study selection") to precede details of the search because the actual study selection process (described under "Data sources and searches") applies these criteria. We have reordered the methods.

2. Only 2 databases were searched - this is certainly not a "strength of this study" as the authors later claim. There may be a strong rationale for only searching these 2, but this needs to be given - at least 4-5 databases would be the norm - please see any guide on systematic review methods. Why was a wider search deemed unnecessary? No grey literature/unpublished/research registers were searched - conference abstracts might have offered further evidence.
We did not phrase this very well and we have rephrased it. What we mean is that we understood many articles would not report data on the association between smoking and pneumonia in the titles or the abstracts so we were comprehensive in searching the full text of all cohort studies on the association between an exposure and pneumonia for data that incidentally reported on the association between smoking and pneumonia. In conducting this review, we had to decide to spend time doing this or searching databases more widely. In fact half of the included studies came from our search of cohort studies on the association between an exposure unrelated to smoking and pneumonia so we believe this strategy was appropriate. We have included this in the methods now and made it clear in the Discussion.

Quality assessment:
3. What if the study was unclear about definition of smoking or whether loss to follow-up was unequal between groups. This option is not acknowledged. If it was unclear, did the reviewers simply score it as "0"? If a study did not define what was meant by a smoker then it scored 0 points. We have amended the description to make this clear. If loss to follow up was unequal it scored 0 points as described in the scoring system "0 points – loss to follow up unequal in comparison arms".

4. I am uncertain about the blinding criterion. How is this applied in the case/control? Someone either has pneumonia (cases) or they do not, surely? What are the assessors assessing? Can you realistically blind something like smoking in a prospective cohort? I think this question is flawed.
In case-control studies the outcome has occurred and interviewers are assessing smoking status and it is ideal that they are blinded to outcome status (i.e. pneumonia or not) [http://ebn.bmj.com/content/7/2/36.full](http://ebn.bmj.com/content/7/2/36.full). Obviously, in severely ill people, pneumonia would be apparent to anyone, but it may be less clear in people in recovery, who presumably are more likely to be recruited into studies. The controls were ill people with HIV in hospital so blinding may have been possible. In cohort studies, the exposure has occurred but it is ideal that the people adjudicating on whether the outcome has occurred or not are blind to exposure status. The use of adjudication committees assessing outcomes of cardiovascular trials is an example of this.

5. I also think a case needs to be clearly made for the "scoring" system as this approach is now often (rightly) criticised; principally because it assumes equal weight for each item. For example, the presence of confounders: is educational attainment really an equal of drug use or viral load or CD4 in this analysis? The discussion suggests not, yet they are treated equally in this assessment. Is HAART a more important confounder than socioeconomic status? Its frequent use as a confounder in analyses suggests it is.

The degree to which a variable confounds depends upon the degree of imbalance between exposed and non-exposed cohorts and the degree to which the potential confounder is associated with the outcome. Dr Carroll picks on two variables that probably vary in the strength of association with the outcome (pneumonia). However, the imbalance between smokers and former or never smokers may well be stronger for SES than for use of HAART. If one were to construct a weighted system of scoring for confounder adjustment we would need investigators to report the association between smoking and pneumonia adjusted for one confounder at a time to see how important a confounder it was. No studies did so. (It is important to remember that 7 of the 14 studies did not primarily aim to report on this association and we extracted data from cohort studies examining other associations). We did, however, try to address this issue in a paragraph subheading 'Assessing the effect of confounding' (p10), where we examined the strength of the association between smoking and pneumonia in studies that reported unadjusted and adjusted HRs or ORs. In two studies, multiple adjustment changed the HR/OR downwards and in two upward. Thus, even though studies varied in the confounders that they adjusted for we felt comfortable that the association between cessation or continued smoking and pneumonia was not easily explained by confounding. This may suggest that pursuing the currently unknown degree to which these variables individually confound the association is not that critical to our findings.

We agree that scoring systems regard all faults as equal when some may be more important. For example, in our study, all included studies were poor at measuring the exposure variable, but this fault is likely to underestimate the strength of the associations between smoking and pneumonia or make cessation appear less beneficial. However, the use of scoring systems is common. We recently published a review in the BMJ where we were criticised by the hanging committee and a referee for not using a scoring system [http://www.bmj.com/content/345/bmj.e4439](http://www.bmj.com/content/345/bmj.e4439). We argued against using Newcastle-Ottawa scale in this particular study. However, in this current review we found little variation between studies in the two key variables of how they assessed exposure and outcome and there were relatively few studies.
included. We therefore do not see the scope for finer grained sensitivity analysis. We have added a justification for the use of a scoring system. There were only 14 studies in the review, of which only four were directly relevant to the association between cessation and pneumonia and the studies have several methodological differences. Finer sub-grouping of limited studies is not possible so the scoring system gives the reader some sense of the importance of methodological differences. As discussed above, we also explored this through comparing unadjusted and adjusted estimates, as is standard in observational epidemiology. We hope the reviewer finds this satisfactory.

Results:
6. "We identified not RCTs ...". But you were not looking for RCTS, were you? You were only looking for cohort and case controls? Unless this was a post hoc decision in the absence of RCTs? You need to be careful that your question and criteria fit what you describe in your results. We should have been more careful in the way we described this. An RCT of cessation support versus no such support that showed a lower incidence of pneumonia in those randomised to cessation support would provide definitive causal evidence. We looked separately for such an RCT and, as we did not find it, proceeded with this review. We have now added this to the introduction and removed this sentence from the results.

7. Some basic study data required here: how many cohort studies; how many case/controls; mean ages; gender; location; total number of participants etc. also, number of cohort (9) and case/control (5) studies. I appreciate that this information is available in the Tables, but, for the sake of 2-3 sentences, the life of the reader could be much easier. such preliminary narrative synthesis is a key part of any review. We had reported average age, gender, total person-years of follow up (which is probably more informative than total persons) for cohort studies. We have included more data on the number of cases and controls in this opening paragraph and geographical location of the studies.

8. Quality assessment: only one defined "stable former smoker" - yet there are 5 references here. This clearly also creates a problem for analysis of this "group". We have checked this and what we said about the one study was wrong (and the five references were a mistake arising from Reference Manager). Although the method of this study reported assessing smoking by various categories including stable former smokers, in practice the analysis compared current smokers and current non-smokers and the fine-grained analysis by types of smoking status was not reported in the results section of this paper. We have thus amended this. In any case, this study was not in the analysis of current versus former smokers, so did not affect the main result.

9. What is the threshold for "higher quality" studies; what does the score have to be, and why? The score chosen was 6 and this was an arbitrary choice that split the studies into 6 higher quality studies and 8 lower quality studies i.e. roughly at the median. We have made this clear. We think, given the limited number of studies, that this is the most reasonable way of trying to explore the issue of study weakness on the results.
10. There really is a problem of reporting. What characterised the first 4 studies analysed? pp.8-9. Please - use sub-headings if possible to specify analyses, eg. current vs former smokers (4 studies - all cohort?) (where's the figure for the first analysis?) current vs never (4 studies - all cohort?) - Appendix 2 (should be 3) Figure 1 Current vs non (8 cohorts) - is this a combination of these 2 sets? The following is taken from Dr Carroll's opening remarks pasted here as the issue is the same “results could be more systematic, sequential in presentation: current vs former smokers, Former vs never, current vs never, then, by study design, by quality etc.”

We are assessing the effects of smoking cessation on incidence of pneumonia so we have presented the data in what we believe is the most logical order for this. Thus, we have estimated the association between former smoking versus current smoking and pneumonia first, then between former smoking versus never smoking second. The data on current smoking versus current non-smoking are rather less directly relevant and they are presented third. We have added subheadings to make this clearer. The figures for the 'main analysis' of former smokers versus current smokers are in a separate Figures file (as requested by the journal). It's Figure 2. I am not sure if its absence was our fault or BMCs fault but we have definitely submitted this now.

We could have presented the data split by study design then address each question within study design, or by question and split by study design within question. We chose to present the results orientated by question and split within that into higher/lower quality and present cohort and case-control studies separately within each question. This seemed and still seems the most logical presentation to us and no other referees remarked on this.

The first four studies presented in the results examine the difference in risk between former smokers and current smokers for the occurrence of pneumonia. The 8 studies examine the association between current smokers and current non-smokers incorporate the data from the four studies that present data on former smokers versus current smokers but also incorporate data from never smokers to match the other four studies where only data on current smokers and current non-smokers were presented. We have added titles to help make navigation through this section easier.

11. Why have ORs rather than HR been used for case/control (I appreciate there is a reason, but it is not explained)
The reason is that HRs require people to be followed and time-to-event to be calculated. Two sentences explaining this has been added to the methods.

12. No case/controls rated "highly". Is this inevitable, given they are judged by the same criteria as a different study design?
It is not inevitable. The same criteria apply except in two respects. The category ‘Addressing major biases’ is loss to follow up in cohort studies and is selection of controls in case-control studies. The other respect is that in a cohort study it is ideal if the assessors are blind to exposure status when assessing the outcome and in a case-control study that assessors are blind to outcome status when assessing the exposure. We had not translated this second variation to the scoring system as we
presented it but have corrected this now. We therefore see no reason to vary the scoring system by study type except for these two variations with each study type able to achieve the same score.

Minor essential revisions:

Introduction:
1. Please define CD4
CD stands for cluster of differentiation 4, which probably does not mean much in itself as it is based on a method of identification of lymphocytes. We have defined this as a type of lymphocyte, which may be more meaningful.

Methods:
2. Text claims 2 reviewers screened titles/abstracts; flow diagram suggests 6 (which would be excessive!)
The text says two people extracted data, which is true. Six people did screen the titles and the abstracts - 3 sets of 2 people for various reasons that would take too long to explain. The flow diagram does not need to include numbers of people doing the task so we have removed it.

3. Appendix 1: The search: Need line numbers if the combining of numbers is to make sense. Why was no date limit applied? There cannot have been any relevant materials from before, eg. 1980.
We have added line numbers. There was no need for date limits because searching as described did not include old articles that were not about HIV.

4. Appendix 2 also has forest plots and additional figures. Presumably this should be Appendix 3.
It should be Appendix 3. Sorry about the mislabelling.

Results:
5. Figure 1: PRISMA: The number need specifying as citations or references. The final number=14 is therefore studies (if all are separate studies)
We have made it clear that there were 14 separate studies.

6. The risk "difference" between post-HAART groups based on study design is not explored (and is questionable whether it is valid - given that it is so small).
We think that Dr Carroll is referring to data on current smokers versus current non-smokers on the incidence of bacterial pneumonia in the cohort studies and case-control studies where participants were treated with HAART. This shows an RR of 2.15 (1.49 to 3.10) for cohort studies and an OR of 1.64 (1.09 to 2.47) for case-control studies. If this is the case we do not think that it is worth discussing this difference. The risk estimates are not very different and the confidence intervals overlap markedly.

7. Reduced risk for smokers of PJP? (less than 1, 0.94 and 0.97)? Surely worth a comment?
We are not convinced that this finding is worth further discussion. The two estimates were 0.94 (0.79, 1.12) and 0.97 (0.81, 1.16), indicating no evidence of a difference in risk between current and non-smokers. It's somewhat implausible that smoking may reduce risk of PJP and it would have no direct practical implications if it did (because
smoking has so many other harms). We have summarised these findings as smoking has no apparent risk for causing PCP. We do not know why this is and so do not know enough of the biology of this disease to speculate as to the reason.

Discussion:
8. It would be good to see the results of the funnel plot to assess publication bias. We have included these in the Appendix. It is worth noting that 7 of the 14 studies were studies about something else but it was possible to extract data presented incidentally on the association between smoking and pneumonia. The decision to publish these studies was unrelated to any considerations about the strength of association between smoking and pneumonia, which is some guarantee against publication bias.

9. Some key within-study limitations are acknowledged (though surely it might be better to write that a proportion of former smokers "might" relapse, or are likely to, and vice versa for smokers quitting; writing that they "would" relapse seems a little absolute and is not demonstrated here). We have adjusted the language to might.

10. What are the cost benefits of reducing the incidence of pneumonia in this population, must be good? Compared with paying-for NRT too? Just acknowledge as a limitation. We cannot assess the cost-effectiveness directly and have added this as a limitation. All decisions about additional services must be based on cost-effectiveness and all that has been demonstrated here is a difference in risk. Given what we know about the efficacy of smoking cessation treatment and its cost-effectiveness we can speculate that such interventions would be highly cost-effective by NICE standards for example. We have added comments to this effect in the Discussion.

11. What are the interventions that have been found to work for this population? Compliance issues with smoking cessation programmes etc., even more so in this population? We searched for and could not find any definitive trials of smoking cessation in the population with HIV so the effectiveness of services is unknown. This meta-analysis should help create the perception that this would be a useful undertaking. There are some pilot randomised trials showing difficulties with adherence to the programme and we have now included reference to these. There is no reason to assume that people with HIV who smoke will respond to treatment differently from people without HIV so normal treatment protocols may apply.

**Quality of written English:** Acceptable  
**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.
Reviewer: Ronan Breen

Reviewer's report:

Major compulsory revisions:
1) the phrase "delaying the eventual decline" implies that CD4 ultimately declines even with HAART which is not true. Please correct.
We gave done so.

2) Although I agree that the evidence does suggest that smoking cessation will reduce bacterial pneumonia I am not sure that it is "good evidence". I think what the manuscript does well is to describe the evidence available but shows how better data would be welcome. I think there should be some mention of how better data from existing large cohorts (of which there are plenty) is needed. I also wonder whether the authors might like to comment upon cost-effectiveness of smoking cessation as an intervention and whether it is more or less so than vaccination in this patient group.
We agree that better data would be welcome. The major weakness of the data is the poor assessment of smoking status which is the exposure variable here and therefore inaccurate measurement has a potentially major influence on the findings. We have expanded our discussion of this and included more discussion of the particular features of future studies that would help.

It is always hard to decide on terminology as the scale is essentially arbitrary. We believe there is good data that smoking is a causal factor for pneumonia. (A relatively strong association with evidence of dose-response relationships). It stands to reason that cessation would ameliorate this and this is what we found some evidence of, but we agree the data are weaker. We have therefore moderated the language to reasonable evidence.

Minor essential revisions:
Although PJP is a correct abbreviation of the organism most people would still use PCP (pneumocystis pneumonia). I think this would be better.
Thanks. We have changed this.

2) I assume that page 14, first para, second last line should say "for" after pneumonia?
Yes thanks- corrected.

3) Page 15 first para, I think the term "gay smoker" is a porr one and should be changed please
We have changed to gay people who smoke

Discretionary revisions:
1) the focus is on pneumonia but there is evidence of an elevated risk of TB with smoking which is relevant to HIV cohorts and also of COPD and lung cancer. I would have thought that this would be worth mentioning at some point as part of the rationale for smoking cessation in HIV.
We have expanded our discussion of the implications of this research in the Discussion and incorporated this.
2) The lack of association with PCP is no surprise as we known that CD4 is crucial unlike with bacterial pneumonia, PCP is never seen even in the heaviest smokers unless CD4 is low and HAART appears completely protective despite smoking or drug use judging by the data showing that prophylaxis can be safely stopped. Perhaps this could be reflected in the discussion. We have added data to this effect.

3) ref 50 is from 2010. Is this now in press?
Yes now published. We have updated this.

**Quality of written English:** Acceptable
**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.
Reviewer: Lawrence Mbuagbaw
Reviewer's report:
The statistical techniques and analyses seem to be appropriate. The reporting can be improved. See below:

Major compulsory revisions:
1. Consider reporting the p values for the odds ratios and hazard ratios alongside the confidence intervals, all in the same brackets and using square brackets if necessary e.g. (risk ratio [RR] 1.06, 95% CI 0.89, 1.29; p= 0.542).
   We have done so.

2. Publication bias is not described completely. The mentioned funnel plot can be added to the paper.
   We have done so. It is worth noting that 7 of the 14 studies were studies about something else but it was possible to extract data presented incidentally on the association between smoking and pneumonia. The decision to publish these studies was unrelated to any considerations about the strength of association between smoking and pneumonia, which is some guarantee against publication bias.

3. Measures of heterogeneity are not described. What do you make of the values for I2 reported? Are they high, moderate or low?
   We have commented on the heterogeneity where relevant and put a more general statement about it in the methods to save repetition.

4. Figure 2 should not be the main or only Forest plot shown. It is more useful to present one for each major comparison described in the abstract. Also there is no statistical heterogeneity in this plot, so there is no reason to present the data as if there were. I2=0% in the subtotals and in the total.
   All Forest plots for every comparison were displayed in the Appendix but we have moved them to the main body of the report.

We believe Dr Mbuagbaw is questioning the subgroup analysis by pre- and post-HAART studies. We defined in the Method that we would do this. This is because HAART has made such a difference to the prognosis that we felt it was important to see whether smoking continued to be an important risk factor for pneumonia and hence cessation provide a benefit. Displaying the subgroups and showing there is no heterogeneity between them is therefore an important result and not a post-hoc decision in the face of heterogeneity, which we agree is not shown by the analysis.

Minor essential revisions:
1. The odds ratios reported are quite close to one. It would be interesting to perform more extensive sensitivity analyses. Apart from study quality and design you can also look at the results based on the definition of smoking (as an example).
   We agree that the key variables that would be interesting to explore are how the exposure was assessed and how the outcome was assessed as these are likely to have the biggest influence on the strength of the association. However, there was near unanimity among studies in these respects (Table 2). The key aspect of assessing exposure is that smoking status varies over time- people who were smoking stop doing so and people who were former smokers relapse. No studies took this into account. We do not believe that minor differences in whether smoking status was assessed by questionnaire or some other means are likely to be
influential— they are all versions of self-reported baseline smoking status only. Likewise, only three studies did not use optimal methods of assessing the outcome. As reported in the response to Petrosillo, by excluding lower quality studies from the analysis we excluded all three such studies and the results were not substantially different. As a result, we do not think there is much scope for additional sensitivity analyses.

2. Study size might be interesting to investigate. Do the smaller studies consistently provide contrary findings?
   We had no hypothesis that smaller studies would produce contrary results and so did not investigate this. As noted, there was no substantial heterogeneity in most analyses. Where this was noted, exclusion of the lower quality studies removed the heterogeneity. We therefore feel that exploratory analysis by post-hoc groupings such as study size will add complications and not help the reader.

3. The fourth column in table 1 is not well presented and it is difficult to distinguish between means and medians and SDs.
   We would have preferred to present only means and SDs but not every study reported this, so we have had to report the one that is presented. We have included the word Mean xx, xx where the number is a mean and the word Median xx where it is a median. SDs are presented in brackets and we have indicated this at the head of the column. We are not clear what else we could do to make this clearer.

Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Editorial requests
1. Please ensure that your manuscript fully adheres to the PRISMA reporting guidelines (http://www.prisma-statement.org/2.1.2%20-%20PRISMA%202009%20Checklist.pdf)
We reported this according to the MOOSE guidelines. PRISMA is mainly used for studies assessing the effectiveness of interventions. However, we have ensured that our results also comply with PRISMA.

2. Please include an Authors' contributions section after Competing interests. More information can be found here: www.biomedcentral.com/bmcmed/authors/instructions/researcharticle#formatting-contributions
We have added this.