Author's response to reviews

Title: Meta-analysis of the relation between European and American smokeless tobacco and oral cancer

Authors:

Rolf Weitkunat (Rolf.Weitkunat@pmintl.com)
Edward Sanders (Edward.Sanders@pmintl.com)
Peter N Lee (PeterLee@pnlee.demon.co.uk)

Version: 2 Date: 4 April 2007

Author's response to reviews: see over
Author’s responses to reviewer’s comments on the submitted manuscript

“Metaanalysis of the relation between European and American smokeless tobacco and oral cancer”

Before responding specifically their comments, we want to express our gratitude to the two reviewers, Paolo Boffetta and Julia Critchley. They have made an enormous effort to point out potential weaknesses of the manuscript and thereby to help us to improve it. We very much acknowledge their extremely competent and helpful comments and want to thank them for the time they have invested.

Again before answering the points in turn, it is perhaps helpful to summarize some of the main changes made to the paper:

1. The results section of the abstract has been rewritten,
2. The materials and methods section has undergone major changes,
3. We have ensured that all estimates included in any meta-analysis are independent, where appropriate deriving new estimates for smokeless tobacco from separate data for snuff and for chewing tobacco,
4. Some sensitivity analyses are now included in the results section,
5. The discussion section now includes a new paragraph and Table concerning results in never smokers, and
6. The discussion section relating to confounding has been considerably changed. These and other changes are described further below.

Reviewer: Paolo Boffetta

Major Compulsory Revisions

1. There have been reviews and meta-analyses of oral cancer risk from use of smokeless tobacco. The authors should stress what their review adds to these previous papers.

Our meta-analysis includes 11 more articles than Rodu & Cole[1], and their analysis does not include publications after 2002. Compared to Rodu & Cole, the present analysis evaluates various sources of heterogeneity (cf. table 5) and provides random effects estimates. The other recent review, by Critchley and Unal,[2] is based on a literature review and does not include meta-analyses.

We have added a sentence at the end of the background section [now on page 8] to make it clearer what our review was intended to add, and have pointed out the greater extent of our coverage in the first paragraph of the discussion section. [See new page 21.]

2. The use of quality scores is debatable, and several authors (e.g., Greenland) have argued against them. Although the authors describe in detail the scores they have applied to the studies under review, they do not seem to have use them extensively in the analysis.

The comments of the reviewer are valid. On reconsidering the issue, and noting also the comments given by the second reviewer, we decided to remove the parts related to the quality scores from the manuscript.
3. In general, the authors seem somehow biased in their interpretation of the evidence. Potential confounding is mainly assumed to act in the direction of generating a positive association, even if the evidence for that is not very strong (see for example page 23, first paragraph; page 24, first paragraph).

The role of smoking and alcohol in the etiology of oral cancer is indisputable. The question of uncontrolled confounding from these sources is of potential importance and merits careful discussion. We feel that this is done in a rather balanced way. For example, in the first paragraph on page 25 we point out that in CPS I and II as well as in NHANES little association between the use of smokeless tobacco and alcohol consumption has been found. Furthermore in women, reported alcohol consumption in NHANES I is lower in smokeless tobacco users than in non-users. With respect to the Winn study, and the situation in North Carolina many years ago, we cite Nilsson’s views. It should also be pointed out that in the last sentence of this paragraph we clearly say that we do not know: “Whether smokeless tobacco users are more likely to underreport alcohol consumption is unclear.” Similarly, in the first paragraph of page 26 we point out that some of the observed differences “...may be due to confounding by alcohol...”. The remaining part of this paragraph correctly points out that other sources of confounding are not even mentioned by most authors.

4. Along the same lines, the authors should avoid speculation (e.g., page 18, third full paragraph; page 22, second paragraph; page 23, first paragraph).

Page 18: The relevant paragraph has now been amended to read:

“Sex-specific overall risk estimates were higher for women than men. The fact that the sex difference was larger using unadjusted estimates suggests that at least part of it may be due to confounding.” [Now on page 21.]

Page 22 and 23:
The second and the third examples in the reviewer’s parenthesis seem in fact to be a repetition of the previous revision request, No. 3, which we have already responded to.

5. Several statement do not see fully relevant (e.g., page, bottom of first paragraph; page 27, first paragraph). Given the length of the manuscript, the authors should consider focusing on the most relevant issues.

We felt that the issues we discussed were all relevant. Of the two examples cited by Dr Boffetta, one is incompletely referenced (no page number) so we do not know what it relates to, and the other relates to a discussion which the other reviewer thought we should expand on (see reply to her point 21).

6. The comparison of summary RR with and without adjustment for potential confounders (tables 4 and 5) is not straightforward, since it compares adjusted results of a subset of studies to unadjusted results of the whole set of studies. A more informative (with respect to confounding) comparison would be between meta-analyses of unadjusted and adjusted results of the same subset of studies.

We believe that the reviewer had misunderstood the results we were comparing and we have endeavoured to make this clearer in the methods section. For each study we extracted adjusted effect estimates if available and unadjusted estimates otherwise, the
full set of estimates used being as shown in Table 3. When we compare estimates by degree of confounding we are therefore comparing sets of results from different studies, not estimates for a subset of studies with estimates for the whole set, as the reviewer thought. We agree with the reviewer that comparison of adjusted and unadjusted results from the same set of studies would be desirable, but in practice this would have limited attention to very few studies. Table 5 compares 19 estimates unadjusted for any factor and 15 estimates adjusted for smoking. Had we restricted attention to studies that provided both unadjusted and adjusted estimates, the 16 studies providing the 19 (sex-specific) unadjusted estimates could not have contributed and one would have been left at most with the 14 studies providing the 15 adjusted estimates. When we looked at the detail of these, 4 of the 15 adjusted estimates were adjustment by restriction (to smokers or nonsmokers) so the unadjusted and adjusted estimates were perforce the same. We attempted to extract unadjusted estimates to compare with the adjusted estimates for the other 11 but for study 21 this was impossible, and for the three cohort estimates (from studies 31 and 32), crude estimates seemed inappropriate because of lack of age-standardization. For the other seven estimates (from seven studies) a comparison was possible, see below, and was consistent with adjustment reducing the risk, but used so little of the data that we did not feel it worth including in the paper.

<table>
<thead>
<tr>
<th>Study</th>
<th>Adjusted RR (CI)</th>
<th>Crude RR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8  Keller</td>
<td>3.63 (1.02-12.95)</td>
<td>2.26 (0.71-7.20)</td>
</tr>
<tr>
<td>13 Winn</td>
<td>2.67 (1.83-3.90)</td>
<td>1.97 (1.41-2.76)</td>
</tr>
<tr>
<td>23 Mashberg</td>
<td>0.96 (0.70-1.33)</td>
<td>1.35 (0.98-1.86)</td>
</tr>
<tr>
<td>24 Perry</td>
<td>1.43 (0.64-3.21)</td>
<td>1.70 (0.81-3.55)</td>
</tr>
<tr>
<td>25 Kabat</td>
<td>1.11 (0.81-1.53)</td>
<td>1.21 (0.88-1.66)</td>
</tr>
<tr>
<td>27 Lewin</td>
<td>0.98 (0.63-1.50)</td>
<td>1.07 (0.72-1.60)</td>
</tr>
<tr>
<td>30 Schwartz</td>
<td>1.00 (0.40-2.30)</td>
<td>1.20 (0.55-2.62)</td>
</tr>
<tr>
<td>Fixed-effects</td>
<td>1.28 (1.08-1.51)</td>
<td>1.40 (1.19-1.64)</td>
</tr>
</tbody>
</table>

**Minor Essential Revisions**

1. The argument that smokeless tobacco products used in South and Southeast Asia are too heterogeneous for an assessment of carcinogenicity is weak. In fact, several international panels and health authorities have reviewed the toxicity of these products and have resulted in product-specific evaluations. A stronger argument to exclude them from this review is that they likely differ in toxicity from products used in Europe and America.

We agree with the reviewer. We have amended the paragraph (now starting at the bottom of page 7) so that we refer to the likely differences in toxicity, and we now say that the evidence in parts of the world other than Europe and the US is “difficult to quantify” rather than “impossible to assess.” We have also amended the second sentence of the abstract.

2. The authors fail to refer to the recent IARC monograph: although still unpublished, the summary and evaluations have been reported (Cogliano et al., Lancet Oncol 2004).
It is somewhat difficult to refer to an as yet unpublished manuscript which one does not know and which is, according to http://monographs.iarc.fr/ (retrieved: 16 February 2007) still in preparation. The information given by Cogliano et al. is relatively sparse. From what they write, it seems that IARC will not change their evaluation of smokeless tobacco with respect to oral cancer: “Overall, there is sufficient evidence that smokeless tobacco causes oral cancer and pancreatic cancer in humans…” (page 708 of Cogliano et al.) While this evaluation seems to refer to all smokeless tobacco throughout the world, we do not know whether a more differentiated point of view will be adopted somewhere in the monograph with respect to smokeless tobacco used in different regions of the world. As the reviewer has requested this, we have made some changes in the last but one paragraph of the discussion in order to be able to refer to the forthcoming IARC monograph 89.

The relevant sentences, now on page 28, now say, with suitable references:

“In the mid 1980s smokeless tobacco was assessed as a risk factor for oral cancer by IARC and the US Surgeon General. Although, according to an advance report on a forthcoming monograph, IARC seems to maintain this view, more recent reviews have reached different conclusions.”

Reviewer: Julia Critchley

The authors have carried out a systematic review and meta-analysis of epidemiological studies estimating associations between smokeless tobacco use (in different forms) and oral cancers. I am concerned that the authors may have made a statistical error (‘double counting’ control series for studies which report results for two or more types of smokeless tobacco use, described in more detail below). I’d like to see either details of how the authors adjusted size of control groups, or a re-analysis to correct for this, before making any recommendation.

Major Compulsory Revisions

1) See general comment above. The authors have identified and extracted data from 32 studies, obtaining 50 effect estimates in total. Some studies provide information on more than one form of smokeless tobacco (commonly chewing tobacco and snuff), so each study may provide more than one effect size estimate. On page 12 the authors describe how they have avoided duplication between case series by omitting 2 estimates based on cases already used in effect size estimates for other subgroups in the same study. However, as far as I can tell, the authors have not adjusted for the size of the control group when including estimates from 2 case series from the same study in the same meta-analysis, as Table 4 shows they have combined 48 estimates (top row) from studies of any type of smokeless tobacco use. When two or more case series from the same study are included in the same meta-analysis, the size of the control series must be adjusted appropriately. If this is not taken into account, control patients are effectively ‘double counted’, erroneously improving the precision of the pooled estimate, and potentially altering effect size estimates also. The Cochrane collaboration open learning material provides further details and suggestions on how to deal with this. It is imperative that the authors state clearly how they have dealt with this problem, and correct the analyses presented if need be.

Pairs of estimates in which double-counting of control patients had occurred were studies 1, 11, 14, 17, 21, 22, 23, 24, and 29. Although the corresponding pairs of estimates are indeed strictly not completely independent, they were originally taken to be so for the purposes of meta-analysis regarding any or unspecified smokeless tobacco, any resultant error being thought likely to be minimal. Furthermore, type-specific meta-estimates (i.e. those for chewing tobacco and snuff) were not affected by this issue. However, as the reviewer is technically correct, we have corrected this issue of “double-counting.”
For the two studies 1 Broders 1920 and 17 Spitz 1988 the source population gave counts (or proportions which could be converted to counts) which allowed one to derive directly estimates for smokeless (chewing or snuff).

For the three studies 11 Wynder 1977, 14 Wynder 1983 and 29 Schildt 1998 counts were available for chewing and for snuff but not for joint use, and we derived estimates for smokeless (chewing or snuff) assuming there were no joint users. Given at least one of the types of smokeless use was very uncommon in these populations, this seems a reasonable approximation, the adequacy of which is tested by a sensitivity analysis.

For the other studies, we used a method for estimating combined estimates based on ideas originally put forward by Greenland and Longnecker[3] The principle is to derive a set of pseudo-numbers of cases and controls which can then be combined to allow the required overall estimates to be generated. The method used has been described briefly[4] and a more detailed publication of this method of estimating “effective numbers” (EN), which also describes a SAS macro, is currently submitted for publication.

For studies 21 Sterling 1992 and 24 Perry 1993 the EN method was used to combine adjusted estimates for smokeless tobacco for two different levels of exposure. For study 22 Zahn 1992 the EN method was used to combine adjusted estimates for buccal cavity and for pharynx. For study 23 Mashberg 1993 estimates for chewing and for snuff were combined by the EN method. This also makes the assumption there were no joint chewers and snuff users.

Note that the methods section has been extended to clarify the methods used, and Table 3 attaches superscript footnotes to each OR/RR estimate to indicate the specific source of the estimate.

2) Abstract results -the authors present evidence of funnel plot asymmetry, but there are a number of possible explanations for funnel plot asymmetry, other than publication bias. I therefore feel this sentence should be amended or excluded.

The reviewer is right as indeed there could be selection biases operative in smaller studies (towards including populations with higher risk) or smaller studies may have been of lesser quality. Although these and other explanations might account for funnel plot asymmetry, the method is still being used mainly to assess the possibility of publication bias. In the present findings, this possibility cannot be excluded. The last sentence of the abstract’s Results section has been modified to read:

“The pattern of estimates is suggestive of some publication bias.”

3) Page 8 onwards. Further details of the methods of study identification, selection, and quality assessment are needed. How many authors selected studies, assessed them for inclusion, and study quality? Did at least two authors select and assess studies independently? If so, what was the level of agreement in terms of exclusion and quality assessment between the two authors? How were any disagreements resolved?
We now no longer refer to study quality in the paper. As regards study identification and selection, the methods section now makes it clear that RW carried this out and then compared his results with an earlier unpublished review conducted by PNL. In practice RW missed nothing that PNL (who only used MEDLINE, reference lists and reviews) cited and found a few additional references. Given also that there have been other quite recent reviews - notably by Critchley herself - it seems unlikely that anything (except the most obscure source) has been missed.

4) Page 9. Estimates of study quality. Can the authors provide information as to whether study scores used (0-9) are based on any empirical evidence or have been validated? Is there any independent evidence that scores of 6-9 represent ‘better quality studies’ than lower scores? Many systematic reviewers are skeptical of study scores (it has been shown empirically that different scoring systems used as a filter for inclusion of studies in systematic reviews can result in different conclusions). Study quality is clearly important, but for these reasons I would recommend carrying out sensitivity analyses based on pre-specified quality criteria (such as control of confounding) rather than such numerical estimates.

We agree that quality scoring systems are controversial. The scoring system that we used was specified a-priori. It is bases on quality-related features of case-control and cohort studies, each of which are covered by 9 specific items. For example, with respect to case-control studies, these items include “explicit case definitions”, “disease state validated”, “controls randomly selected” and others. For each criterion that was fulfilled, one score point was added to the study-specific quality score. As the scoring was done only by the first author, we have no empirical basis of assessing the quality of the quality assessment. As quality assessment at the level of stratifying for adjustment of confounding factors has been conducted already (table 4 of the original manuscript), and upon the comment of the reviewer, we have decided to dismiss the quality score related stratification.

5) Analysis page 9. Can the authors clarify the first sentence here "Estimates for the whole population were chosen in preference to those for data subsets", as I’m not sure which data subsets (cancer type?) they are referring to.

The corresponding description in the methods section (see paragraph 2 on page 10) has been amended to make it clear that estimates for smokers and nonsmokers combined were used if available.

6) Analysis page 9. I'm not clear what the authors mean by the statement "When multiple effect estimates were reported (e.g. for different degrees of confounder adjustment), the one the original authors HAD RELIED on was chosen". How did the reviewers determine which estimate the original authors relied on?

The original description in the methods section was confusing and has been amended. Where results for different degrees of confounder adjustment were available, we in fact used the most adjusted estimates (see paragraph 3 on page 10) – not that many studies allowed much choice.

7) Page 10. Last sentence of first paragraph. Can the authors explain how they calculated effect measures adjusted for confounding, if they were not originally calculated by the study authors? How many (& which) studies does this apply to?
The corresponding description in the methods section has been amended (see top of page 11). In addition, table 3 has been enhanced and now gives the details for each estimate.

8) Page 10. Heterogeneity is clearly important, but there seem to be a lot of sub-groups here, and they won’t have much power with only 32 studies, as well as increasing the potential for spurious findings. These should therefore be rationalized, and a clear statement of how many sub-group analyses were actually performed is needed here.

Apart from an analysis by study quality (now not considered in the paper), and an analysis by whether or not the estimates had been reported by the author or were derived from information presented (mentioned in the text of the results section), all the subgroup analyses that were performed are those reported in Table 5.

The sets of factors considered for the subgroup analyses were derived a priori as being of most relevance. Although in the event some subgroups contained few studies we feel it more appropriate to present all the findings than to pick and choose. The methods section now contains a clear description of which subgroup analyses were conducted (see main paragraph on page 12).

9) Page 10. The authors have chosen to assess heterogeneity using the chi-squared test, but this is known to have low power to detect heterogeneity, particularly where there are relatively few studies. In general, a cut off of p = 0.1 is used rather than the standard 0.05 for this reason. Can the authors confirm which cut-off they used here?

While choosing a more liberal alpha level leads to more sensitivity, specificity suffers at the same time. In order to protect a multiple alpha level (either 1 or 5 percent) a test-wise alpha-level far below 1 percent would need to be adopted, rather than a more liberal one. In the context of an exploratory analysis, it seems to be far more appropriate to report exact p values and consider the corresponding tests as descriptive tests, rather than performing confirmatory tests at a specific alpha level.

Exact p-values are reported for all heterogeneity statistics in tables 4 and 5. As it happens to be the case, only one (within-group) p-value between 0.05 and 0.1 was observed for the heterogeneity tests contained in table 5 (referring to CC studies with population controls, p=0.0514), which implies that using a 5% or a more liberal alpha-level of 1% would not have led to different conclusions.

10) Given the comments above, I would recommend that the authors calculate and interpret $I^2$ statistics for heterogeneity (Higgins et al. BMJ 2003) for each meta-analysis performed, in addition to the chi-squared statistics presented.

The $I^2$ statistic is now provided in addition to the traditional heterogeneity statistics (see tables 4, 5 and 6).

11) Page 10 results. It would be helpful to list the studies excluded and reasons for exclusion in a table. It is not generally recommended to exclude studies simply because they are reported in brief (as abstracts or letters), provided they obtain sufficient quantitative information to include. Therefore, these two reports should be included if at all possible.

Initially RW identified 1313 papers by systematic searches in MEDLINE, EMBASE, CANCERLIT and TOXLINE and also by reference lists of individual papers and reviews.
Scanning titles and abstracts of these papers made it clear that all but 63 were irrelevant for various reasons, such as study in animals, clinical trial, etc. but no details were kept of the reasons for rejection. Subsequent detailed inspection of the 63 made it clear 31 were of no value, and as the text of the first paragraph of the results section (see pages 12 and 13) already succinctly lists the references and reasons for rejection for each, we see no reason for giving this information more lengthily in a table. We agree that abstracts and letters are not themselves a reason for rejection, and now make it clear that the two in question merely commented on papers rejected for other reasons.

The second to the last sentence of the first paragraph of the results section was changed as follows: “Two reports [45,46] were letters commenting on studies already rejected for other reasons and two were reviews. [47,48]”

12) Page 11 first para - seems to suggest that two studies were excluded due to “limited power”. I’m not clear was this was defined and assessed?

On page 8 of our original manuscript we pointed out with respect to exclusion criteria that studies with less than five expected exposed cases were excluded. On page 734 Accortt et al.[5] point out that “There were no oral cancer deaths among exclusive smokeless tobacco users (table 5). In a cohort of this size followed for approximately 20 years, only one death would have been expected.” In the study of Bundgaard et al.[6], three cases and one control had used chewing-tobacco.

13) Page 17 - I’m not very clear how the more detailed heterogeneity analysis was undertaken? Please briefly explain. Are these further sub-group analyses?

We are not quite clear about this comment. In comments 9 and 10 the reviewer pointed out that we should use the I² statistic for heterogeneity analysis. This statistic is based on the chi-squared statistic that we have used so seems to imply that the heterogeneity analysis was indeed transparent. In addition, the methods section references Deeks et al. [our reference 22 in the original manuscript] where the method is described in detail. In order to avoid redundancy, we have not added more details here.

14) Page 18 Effect of study region. I don’t think its reasonable to comment on the effect size for studies conducted in the UK and Brazil when there are so few of them! I suggest deleting this comment.

The method of statistical testing is designed to account for varying levels of power. It seems to be hard to justify a certain level at which this principle is put out of force. If a difference based on a small number of studies is reaching the level of significance, this implies that the effect which has been tested must have had considerable magnitude. Although the precision of estimates for “other countries” is small, the point estimates are differing from those referring to the US. Again, what would be the right point at which quantitative principles should be set aside for qualitative ones?

15) Page 18 I think the paragraph on sex specific risk estimates should be deleted. There are insufficient data in women to draw conclusions.

We agree with the reviewer that it would be highly desirable to have more data in women. However, from the present finding we cannot completely exclude the possibility
that there exists a higher risk in women as compared to men (cf. pages 21 and 27 of our current manuscript). We think that it would not be appropriate to pass over this issue.

16) Page 19. The explanation of publication bias seems unnecessarily long and detailed. I think it would be sufficient to simply point out the presence of asymmetry. It should also be made clear that there are other possible explanations for funnel plot asymmetry apart from publication bias.

This point has been made and commented on already above (comment 2). We have changed this paragraph and replaced the third and following sentences by: “The observed asymmetry of the funnel plot suggests that publication bias cannot be excluded.” [See page 21.]

17) I appreciate this is an open-access journal, but the discussion and conclusions (8 pages in total) seem very long and could be shortened dramatically without any loss of information. I've made several suggestions for shortening below. The potential for residual confounding by alcohol consumption is a valid point worth stating clearly, but far more succinctly. For example much of page 22 could be deleted (from "If one could assume...... to unhealthy habits in general") in my opinion (the authors agree that the inferences they are discussing are unlikely).

The discussion section has been considerably rewritten, with some of the original text removed. What remains seems to us worth saying.

18) Page 23. Winn's study in North Carolina is important, because there are so few studies of smokeless tobacco among women. However, I cannot think of any reason why smokeless tobacco users would be less likely to report alcohol consumption (I suspect the opposite -if a woman is prepared to report one risk behavior she is more likely to report another) so suggest deleting these comments.

The cited reference (Nilsson[7]) has a different view on this. It should be noted that the comments of Nilsson were not general but rather specific with respect to the study of Winn et al. Obviously, this view is exactly opposite to the one of the reviewer. In the last sentence we point out we do not know. [See page 25.]

19) Page 23. Its speculative to suggest that differences in alcohol consumption between the US and Sweden might account for differences in risk estimates, as this is based on an ecological comparison. I think this section should therefore be deleted -the authors agree that the question can't be resolved from the data available. It seems more parsimonious to me to suggest that any difference in risk between these countries may be due to either chance or differences in the forms of oral tobacco used or both.

What the reviewer proposes in her last sentence indeed is more parsimonious, but it is not equivalent to what we were saying. We are clearly pointing out that the issue “cannot, however, be resolved from the data available.” [See pages 25 and 26.] We think that ignoring a country-specific difference of close to 20 percent in alcohol consumption would not be appropriate. Also (although we are aware of the problems associated with ecological data), ignoring ecological evidence just for the reason that it is not individual data, is far from being the generally approach in epidemiology. In fact – remember John Snow – quite a few important epidemiological milestones are based completely on ecological data.

20) Page 25 Paragraph on female user of smokeless tobacco. Again, this seems unnecessarily long and detailed given the lack of sex-specific information. I suggest including the first two sentences only.

We have rewritten this paragraph into a shorter form. [See page 27.]
Whether or not smokeless tobacco use can play a role in helping smokers quit and reducing population risk overall has been fiercely debated in recent years. I feel that this paragraph is rather unbalanced, putting only one side of the debate. For example, there are many authors who point out that the precise risks of smokeless tobacco are still uncertain, or who feel that smokeless tobacco is unproven as a smoking cessation aid (have there been adequately powered randomised controlled trials?), and dispute the "Swedish experience"; the Swedish studies referred to here are based on ecological data. The authors should summarise the arguments for both sides of this debate in this paragraph.

We agree with the reviewer in her assessment concerning the precise risks of smokeless tobacco, especially with respect to "overall health risks" as still being uncertain. It should be noted, however, that there are conclusions and results pointing into the direction of ST of being of less overall health harm [references 78-80 in the manuscript]. The 2002 report of the Royal College of Physicians[8] estimates that “…the consumption of non-combustible tobacco is of the order of 10-1,000 times less hazardous than smoking, depending on the product” (page 5). Roth et al.[9] concluded that health risks for certain health outcomes (specifically lung cancer, oral cancer, and gastric cancer) are lower for snus than for smoking. The authors additionally point out that their review was restricted to studies where both exposures were considered. “Many important diseases known to be associated with cigarette smoking were not included (e.g. emphysema)” (page 747). Certainly, there are opposite views. Hatsukami et al.[10] state that “…ST use facilitates the use of cigarettes”, but also that “Considerably more research and product regulation is necessary prior to considering smokeless tobacco as a harm reduction strategy” (page 309). We are not convinced of their argument concerning the facilitation of smoking by use of ST, but on the other hand decided to stay away from this discussion in the current context of the present manuscript. We feel that our participation in this debate in the discussion section of the present manuscript might be perceived by some as an unbalanced view. Furthermore, and more important, we believe that the debate on overall health risks associated with smokeless tobacco is far beyond the scope of the present paper.

Specifically with respect to the role of ST in smoking cessation we conclude from the reviewer’s comment that she believes that “adequately powered” randomization trials are needed. Again, we strongly agree, and this point has been put forward by others as well (e.g. by Furberg et al.[11]). Unfortunately, we are not aware of any convincing and conclusive trials with sufficient power and high methodological quality. Tilashalski et al.[12] found that 12 of 16 subjects who had quit smoking using ST were smoke-free at seven years, compared to 4 of 6 who used means other than ST. Although the follow-up interval is relatively long and the results point towards a positive effect of ST, we are not convinced of the data, mainly due to the very limited sample size. We therefore did not mention this study in our discussion. Klesges et al.[13] found that “…smokeless tobacco users were 1.33 (95%CI = 1.08-1.63) times more likely than controls (p<.01) continuously abstinent at follow up…”, after a brief tailored smoking control intervention delivered during basic military training. Here the sample size is much larger, but the follow-up is only one year. Certainly, again, there are conflicting results, e.g. by Haddock et al.[14] In a sample of nonsmoking military recruits the authors found that both past and current users of ST were about 2.3 times more likely to have become smokers after a 1 year follow-up. The study has been criticized by Kozlowski et al.[15]
Again, the present manuscript is focused on ST and oral cancer, rather than on the role of ST in smoking cessation. We acknowledge, however, that the reviewer wishes to see that we point at the ongoing discussion. We have decided, therefore, to add the first sentence given by the reviewer on this topic to the end of the paragraph, along with references to Haddock et al.[14] and Hatsukami et al.[16] [See page 29.]

“Whether or not smokeless tobacco use can play a role in helping smokers quit and reducing population risk overall has been fiercely debated in recent years [e.g. by Haddock et al.[14] and Hatsukami et al.[16]”

Table 22) Tables and Figures. I am familiar with some of the studies included in this meta-analysis from my research, but in some cases I am not clear how the reviewers have obtained the estimates they extracted. Can the authors please clarify this in the methods and confirm that all estimates extracted include only never-smokers in the reference category? I give some examples of estimates I am unclear about below:

The methods section has been amended to give greater detail of how ORs and CIs were calculated. Specific information has also been added to table 3 at the level of individual estimates. [See footnotes d, e, f, g, i, j, n, p, q and r appearing in the last column.]

It is certainly not the case that our estimates generally include only never-smokers in the reference group. This would imply either that we restrict attention to those few studies that have reported results for never smokers, thus losing most of the literature, or that we compare subjects who smoke and use smokeless tobacco with those who do neither, with a huge potential for bias due to smoking. The selected RRs/ORs compare ST users with non ST users (not with non smokers) and, except for a few studies where only results for non smokers are available, are based on the whole population of smokers and non smokers. We do however now include a new Table (Table 6) and a brief discussion concerning the limited evidence in never smokers specifically. [See page 23.]

For Lewin 1998, the authors have extracted a value of 1.01 (95% CI 0.68 to 1.52), but I think this OR includes cigarette smokers in the referent group. The valid referent group should be never-smokers and with this group the OR was 3.3 (95% CI 0.8 to 2.0) for current snuff users in this study.

Lewin 1998[17] did indeed cite an OR of 3.3 (though the upper 95% CI extends to 12, not 2), as the reviewer states. However, this is for current users of snuff, relative to a reference category of "never tobacco-users" and is based on an analysis including not only oral sites, but also larynx and esophagus. The reviewer is implying that we should compare snuff users (many of whom smoke) with those who neither smoke nor use snuff. Such a comparison is not appropriate as it would only provide an estimate of the joint effect of snuff and smoking, not of snuff on its own. Our choice of OR for the Lewin study follows general principles underlying our choice for all the studies - compare ST use with non ST use, prefer ST ever to ST current where both are available, and restrict attention to oral cancer. Our estimate is calculated based on Table 5 of the Lewin et al. publication which reports site-specific RRs in ever users which are adjusted for age, region, smoking, and alcohol intake. For oral cavity and pharynx, the estimates reported by the authors are 1.4 (95%CI 0.8-2.4) and 0.7 (0.4-1.3), respectively. Originally, these two estimates have been aggregated by means of a fixed-effect model. Based on the
remarks of the reviewer regarding the problem of multiple use of the control group, a recalculation has been undertaken based on the EN method which avoids the problem. The estimate is 0.98 (0.63-1.50). Although the difference from the overall estimate used in the original version of our manuscript is very small, the more correct estimate is being used in the revision.

In other cases, I think that important confounders have been adjusted for, which the reviewers may have overlooked. E.g. Blot 1998 performed analyses limited to non-smoking women (controlling for smoking by restriction) with an OR of 6.2, 95% CI 1.9 to 19.8. This is different from that reported in this review (OR = 3.44, 95% CI 1.09 to 10.9). It is difficult to extract an OR for smokeless tobacco use for males in this study, as almost all male users also smoked cigarettes - I'm not sure how the authors have extracted this?

Blot et al. (1988, not 1998)[18] reported on page 3283 of their publication, that “Six % of cases and 7% of controls among males had used smokeless tobacco (primarily chewing tobacco), but nearly all were also smokers.” The only relative risk estimate that can be calculated from this is an unadjusted one of 0.85 (95%CI 0.57-1.26) for any smokeless tobacco use regardless of smoking. On the same page, Blot et al reported that in women 3% of cases and 1% of controls used smokeless tobacco (primarily snuff). This leads to the unadjusted estimate of 3.44 (95% CI 1.09-10.9) that we used. We did not use the estimate of 6.2 (95% CI 1.9-19.8) in our main analyses as this relates to nonsmokers and our general approach has been to use the complete population of cases wherever possible, as this is all that is available in the great majority of studies. However, the estimate of 6.2 does appear in the new Table 6.

The authors of this review state that Mashberg 1993 adjusted for smoking, but in my view this study did not adequately adjust for smoking (because the referent category included "minimal smokers", defined as those smoking 1-5 cigarettes per day, which may dilute estimates of risk for smokeless tobacco users).

We agree with the reviewer that using a reference category with “minimal smokers” is not ideal for this study[19] and may indeed dilute estimates of risk for smokeless tobacco use. As the authors point out on page 1370, however, “…too few lifetime nonsmokers were available to form a stable reference category”. There is not much one can do about this. Therefore, we used the OR estimates provided in the article for snuff (0.8; 95%CI 0.4-1.9) and chewing tobacco (1.0; 0.7-1.4). Note that our analyses now use an estimate for smokeless (ever) of 0.96 (0.70-1.33), adjusted for smoking, race, age and alcohol. This is calculated from the estimates for snuff and chewing tobacco using the EN method described above, when discussing the Lewin study.

The authors state that Schwartz et al controlled for smoking, and alcohol amongst other risk factors, but my understanding is that the OR reported (1.0, 95% CI 0.4 to 2.3) is crude. This article is mainly concerned with sexual history, oral sex and HPV infection, and only measures tobacco use as a potential confounder of this relationship. I think that the effect size estimates for the main risk factors (sexual behaviors) are adjusted for confounders as the authors state, but not the OR presented for oral tobacco use.

We had the same initial impression as the reviewer about this study.[20] But then, two things contradict this view: a) On page 1628, the authors state that “Unless otherwise stated, ORs presented are adjusted for age, cigarette smoking (continuous pack-years), alcohol consumption (continuous average number of alcoholic beverages consumed/week during lifetime), and sex (in analyses combining males and females.” This does not leave much room for interpretation. b) Applying the rates of prior
smokeless tobacco use of 6.7% and 5.6% among men (as reported on page 1629 of the article) on the 165 male cases and 302 male controls, a crude OR of 1.197 (95%CI 0.547-2.621) can be calculated. This is different from the one reported by the authors, which further supports the notion that this estimate is indeed adjusted.

Stockwell et al. 1996 - I assume that the authors have combined the estimates for different oral cancer sites, reported in this paper. The method used to combine estimates should be clearly stated in the methods section (I see some details of method used to combine different levels of exposure, but not estimates from different sites).

A combination of the estimates provided by Stockwell et al. (1986, not 1996)[21] in their table 7 is not adequate, as they are restricted to those who were not primary smokers; it is important to note that the exposure to smokeless tobacco in users of multiple tobacco products only indicates that this type of tobacco was primarily used (“only the primary product was recorded”, p. 105). (As table 2 of the publication indicates, the proportion of smokers was 2.8 times as high in cases as compared to controls.) An odds ratio of 2.02 (95%CI 1.01-4.02) can be calculated based on the data contained in tables 1 and 2 of the publication, according to which 11 and 1451 cases of oral cancer were primarily exposed vs. not exposed to smokeless tobacco, respectively. In contrast, 31 of 8285 controls were primary users of smokeless tobacco.

The methods section has been amended to give greater detail of how ORs and CIs were calculated.

Williams and Horm (1997) The authors say no adjustments for confounding were made, but there are tables providing the relative odds for males and females after adjusting for smoking, age and race in the paper. Again, I’m not clear how estimates from separate sites were combined for this study.

This publication (actually 1977 not 1997)[22] is difficult to read and the presentation of the results is unusual. Indeed, relative odds point estimates are given for different levels of smokeless tobacco use. There are some problems with these estimates, however. The exposure levels are not stated explicitly, but rather were chosen to divide the distribution of exposures about equally. Instead of giving confidence intervals, levels of significance are provided (but no precise p-values). The effect estimates are given for specific cancer sites and are based on extremely small numbers of exposed cases. It is difficult to see how reliable estimates for a relative risk adjusted for three factors can be obtained when the number of exposed cases never exceeds eight in any analysis. Finally, the reported estimates do not seem to be consistent. For example, the authors claim the estimate of 3.88 for gum-mouth cancer in male low-level users of smokeless tobacco is significant at p<0.01 (table 8A of the publication). In contrast, the estimate for male high-level users (6.65) and for low-level female users (4.92) are claimed to be not significant, even at p<0.05 (tables 8A and 9A, respectively).

It is possible (although intricate) to extract from the tables in the publication the data required to calculate unadjusted overall effect estimates for males and females. The results are as follows:

<table>
<thead>
<tr>
<th>Smokeless tobacco</th>
<th>Male cases</th>
<th>Male controls</th>
<th>Female cases</th>
<th>Female controls</th>
</tr>
</thead>
</table>


Unfortunately, the data can be easier explained than extracted. Table 8B contains the numbers of exposed male cases in the first four rows (lip-tongue, salivary, gum-mouth, pharynx) of the two columns corresponding to “chewing or snuff tobacco” levels (superscript a in the above table). The numbers of exposed male controls are given in row 10 (superscript b in the above table). The numbers of exposed female cases as well as of exposed female controls can be obtained in the same way from the corresponding values in table 9B (superscripts e and f in the above table).

The number of unexposed male controls (d) can be calculated by subtracting the number of male exposed controls (b) from the total number of controls (which are defined in the study as patients with tobacco-unrelated cancer; cf. footnote on table 4B), which is 1788 according to table 8B. In the same way, the number of unexposed female controls (h) can be calculated from the data contained in table 9B.

The number of unexposed male cases (c) can be determined by subtracting (a) from the total number of cases with oral cancer (which is 190, according to table 8B). By the same token, the number of unexposed female cases (g) can be determined from the data contained in table 9B.

According to these data, the odds ratios for males and females are 0.91 (95%CI 0.53-1.56) and 1.54 (0.37-6.42), respectively.

Figure 1 - The labelling could be more informative i.e. any two identical labels such as 'Henley 2005' should be altered to explain the difference between the two estimates.

This has been changed.

Minor Essential Revisions

1) Usually fixed effect meta-analysis is referred to in the singular (estimating a single effect) and random effects as a plural.

In the “fixed effect” model it is assumed that the study-specific risk estimates share a (one!) common mean effect. In contrast, in the “random effects” model it is assumed that the individual risk estimate reflects a true study-specific exposure effect, which in turn reflects the distribution of exposure effects around an underlying true hyper-effect. Therefore, the use of “fixed effect” and “random effects” is correct (cf. for example Cochrane Collaboration’s open learning material (http://www.cochrane-net.org/openlearning/HTML/mod13-4.htm).

2) Is it possible to reduce the column width in table 2, so it can be presented on one page?
Table 2 is now on one page.

References


