Reviewer's report

Title: The relation between smokeless tobacco and cancer in Northern Europe and North America. A commentary on differences between the conclusions reached by two recent reviews.

Version: 2 Date: 27 May 2009

Reviewer: Carl Phillips

Reviewer's report:

This is an excellent paper in every way. It is well written, the analysis appears flawless (modulo a few comments I offer below), and makes a critical contribution: Instead of engaging in the inefficient and unscientific process that usually dominates this field, dueling monologues that make no attempt to explain contradictions, this paper provides real scientific argumentation. I commend the authors for doing this and the journal in advance for publishing it (I hope). It is too bad that this comparison is in a separate paper from the original analysis, since it means that some readers will not discover it, but I suppose that was necessary to keep things more organized.

On first reading this paper, my biggest criticism is that it was far too conciliatory. There is an unfortunate tendency to in epidemiology to say "my analysis is ok; your analysis is ok", and actually pointing out errors and problems is considered "unscientific" (never mind that aggressively pointing out errors is one of the necessary conditions for something to be a real science). I think some of this tendency is apparent here, and offer a few comments on this below. But after letting a few weeks pass and my re-readings becoming more careful, I have decided that the criticisms are mostly at the appropriate level and the anti-scientific conciliatory voice was not as bad as I thought. (My apologies to the authors for delaying their publication by being late with my review – I review bad papers in one day, but I always want to wait on good papers to thinking hard and offer every suggestion I can.)

As a note to the editor (the authors undoubtedly already agree): The Boffetta paper that the authors are analyzing is terribly flawed. Boffetta et al. tack an extremely misleading synthetic meta-analysis onto a simplistic review paper. They never explain their methods, let alone attempt to conform to accepted standards. They make capricious choices about what numbers to include and exclude, and make several glaring errors. I have pointed out elsewhere (see below) that the Boffetta paper, which does not make any real contributions to the science, appeared intended to be a go-to resource for litigation against smokeless tobacco manufacturers. With these observations in mind, the Lee and Hamling analysis should not try to offer a "balanced" assessment, as if two groups of scientists made legitimate choices but they think theirs are a bit better; Boffetta et al. made some terrible and obviously biased choices, and real science calls for pointing these out.
(Disclosures: I wrote a list of the flaws in the Boffetta paper last year, and planned to publish something based on it, but never got around to it. Thus, I am a bit jealous. However, since L&H is far better with its context of their other work than what I might have written, and certainly makes mine redundant, I freely offer, below, the few observations from my list that they did not mention. I encourage them to add those points to what they write if they agree they are correct.

I disclose that I am an expert witness for the defense in a trial that the Boffetta et al. paper seems designed to influence (one of the authors, Hecht, is a witness for the plaintiff, though he did not disclose that).

Finally, my professional opinion is that synthetic meta-analysis is generally inappropriate for heterogeneous observational studies, and has negative epistemic value when added to a systematic review. Thus, I will never be a big fan of part of the original Lee and Hamling paper (which I was not a reviewer for – for obvious reasons I suppose), though their systematic review alone is extremely valuable. That said, I note that the original L&H meta-analysis appears to be as good as those get, and is a tour de force, far better than most Cochrane papers or IARC monographs. For purposes of this review, I accept it for what it is in this review except for the first three comments below. Also, since Dr. Lee already knows I am reviewing his paper, I will eliminate game-playing by pointing out that I am Carl V. Phillips – I am not sure whether this journal sends out the reviews with the reviewer's name attached.)

I know this is a reference to the other paper, but it really makes me cringe when the authors suggest that the smoker/non-smoker combined estimate is necessarily better than non-smokers only because it "provided greater power". It might indeed be better, but implying that greater power alone is necessarily better (combining more data that is less compatible) reflects one of my serious concerns with synthetic meta-analysis. Why is it necessary to imply that in this paper?

I do not propose to offer an opinion about the right way to do a synthetic meta-analysis, but I will note that one of the L&H vs. Boffetta differences appears rather arbitrary. Boffetta et al. use only published quantitative estimates, while L&H derive additional estimates if they can from the published record, even if the original authors did not report them. But why stop there? There are other papers that look at cancers and multiple exposures that report statements like "no significant association was found with smokeless tobacco". And there are other studies still that looked at the association but did not mention their results. Leaving those out may be a practical necessity, but it is still epistemically arbitrary. Since almost all such papers found a low (null or protective) exposure – high associations get reported – this means that the L&H estimates are still biased upward.

The practical implications of this are (a) I think L&H should spend half a sentence in the discussion mentioning that to the extent that protective and non-significant results were not quantified at all in some studies, even their results are a bit high
and (b) there should be some acknowledgment that the derived-estimate approach includes more and does seem better than published ORs only, but is still drawing an arbitrary line.

I consider this the most significant methodologic point: The authors emphasize in various places, including the abstract, that they attempted to avoid using data subsets, which (as they do not mention) is one of the most common tools used by those who would exaggerate the risks studied here, as I have pointed out in several analyses. However, the authors fail at this in one critical place: As I have pointed out (e.g., http://www.epiphi.com/papers/phillips_pbis_poster.pdf – I can put a copy of this in my permanent working paper archives if the authors are persuaded that it should be cited – which, btw, I am inclined to make a pitch for: Since I have been writing about what is wrong with the usual interpretations of this literature for years, it seems only fair that I be referenced somewhere), the oft-cited result from Blot et al. is an analysis of the small subgroup, nonsmoking women snuff users.

I know of no way to get the result for the larger number of nonsmoking exposed men using public data, but it can be surmised that it is much lower than the OR of 6 for women, which is one of the largest ORs reported in the entire literature. The result for the population as a whole is 1.0, and the result for all men is protective. I realize that there is little chance the inclusion of the biased subsample number from Blot will be changed in the original Lee and Hamling paper, which I assume is finalized and which no one asked me to review. But since the present paper is an analysis of choices, and it repeatedly emphasized the "no subgroups" rule/goal, I believe that in this paper the authors should explicitly acknowledge that the number they used is an apparently upwardly biased subgroup number that they had to use because the other subgroup (men) was suppressed from the reported results.

Similarly, I have pointed out the biased interpretations of the pancreatic cancer results. There is something very strange about the Alguacil result (even with the correction L&H offer) given that controlling for some variables that should have lowered the estimated OR dramatically increased it from a protective association to a positive association. One of the problems with any of these analyses is that such problems are generally ignored when the numbers are just cooked together as if they are all correct but for random sampling error.

Several times, the authors note that Boffetta et al. fail to define particular methods (e.g., "properly adjusted" was also undefined), but never mention the elephant in the room: Boffetta et al. carry out an analysis that usually demand meticulous descriptions of methods and choices, without even sketching their methods. This should be mentioned within the first page or two of the paper to provide critical background for the reader. Not only is this a legitimate criticism of Boffetta that should not be suppressed, but it is critical for the reader trying to understand why the present authors use speculative language in many places.

The language about the patently arbitrary and inconsistent choices Boffetta et al. made is subdued in a few places to the point that it is wrong. It is well known that sometimes "being polite" is a way of saying "avoiding unpleasant truths" or even
"lying"; science should not avoid unpleasant truths. Some of this might be the understated British style, and I picked up on that after a few readings. But some places a change is called for. On p. 6 when Boffetta et al. are said to have "overlooked" some studies, this is polite to the point of lying; there is no doubt that they were aware of at least the Rosenqvist study, as is it considered among the more useful studies on the topic and they have cited it elsewhere. Stronger language (at least "ignored") is in order.

When the authors describe the Stockwell study (ref 18) that Boffetta despite the fact that smoking was not controlled for, they might want to note the following: Because of the nature of that data (recording only one tobacco exposure per person, using a selection method for which usage is "primary" that is not clearly defined), it is not only that it is impossible to control for smoking. The data in that dataset is such that analysis that compares ST users to non-users of tobacco will have no smokers in the unexposed group, but may (presumably does) have smokers in the ST-exposed group. Thus the problem is worse than simple confounding (which is only a problem if the two exposures are correlated): this method guarantees that there are more smokers in the exposed group.

The authors point out the anomaly of Boffetta et al. singling out the tongue cancer result from the Stockwell study. They might be interested in my assessment of why this was done: As I have pointed out note (see http://www.tobaccoharmreduction.org/wpapers/009.htm – look at Sup 2a and scroll down to the original letter to Lancet Oncology’s editor) this anomalous reporting seems easiest to explain in the context of the aforementioned lawsuit, where trying to create epidemiologic evidence that ST causes tongue cancer, rather than cancer of other oral sites, benefits the plaintiffs.

The paragraph heading "Inclusion of a biased result" is a bit problematic, since every result is likely biased to some extent, and quite a few of the results that are included (by L&H, not just Boffetta) are rather clearly biased. Perhaps a better heading would be the more specific, "Unnecessary inclusion of a confounded result".

The observation that all of the departures from the L&H optimum by Boffetta result in bias in the same direction should be better emphasized. This is not an observation that should be allowed to slip past the reader. The authors should point out that given how many disagreements exist, the chance that this is a coincidental result of two different but neutral methodologies is extremely small. Thus the most reasonably interpretations are that (a) L&H have tried to bias their result downward or (b) Boffetta et al. tried to bias their result upward. The reason for acknowledging the first of these is to preempt the ad hominem attacks by the anti-tobacco extremist faction that will inevitably be directed at Lee and Hamling. The possibility that they intentionally biased their analysis should be explicitly acknowledged as an explanation for the difference, and then explicitly responded to. This will not stop all ad hominem attacks on their study (nothing will), but it will give everyone access to a rejoinder.

Having acknowledged that without further analysis we cannot know who is
biased, the authors should then remind us of the analysis. That requires reprising
the content of the rest of the paper, of course, but it should be reprised
specifically in this context. My suggestion would be, in two or three sentences,
reminded the readers that (a) the L&H method is systematic, well-defined, and
difficult to criticize, while Boffetta's is ad hoc and unstated, (b) it is difficult to
imagine any justification for such choices by Boffetta et al. as omitting a major
study, omitting some relevant results from studies they use, and choosing the
nonsmoker analysis sometimes and the controlled-for-smoking analysis
sometimes, always choosing the larger. Having said this explicitly, the authors
can continue with the five particular cases.

It may be that one or more of the authors or editors might feel that this goes too
far into the debate rather than pure science. But it is necessary to be realistic
about the politics. Even though the L&H analysis is obviously superior to
Boffetta's, it is the former that will be tarred as "biased", and that tarring will be
considered "science" by certain "journals". Some preemptive truth on that point in
an unbiased journal is in order.

When summarizing the results for oral/pharyngeal cancer, I would urge the
authors to add the following observations to their existing points about post-1990
studies and adjustment for alcohol. Boffetta et al. followed the odd procedure of
dividing the results between Scandinavia and the United States, rather than
choosing a more epidemiologically relevant division. (This, by the way, appears
to also be motivated by the current lawsuit or future hypothetical American
lawsuits, where the plaintiffs want to draw this artificial line.) Having made that
division, they report a high degree of heterogeneity for the American result,
making it clear that this amalgamation of studies is not appropriate. L&H should
note this. L&H propose solutions to this heterogeneity problem, but fail to
mention one other that Rodu and others have pointed out: Removing the results
that appear to be entirely exposure of Appalachian women to the local variety of
dry snuff in the early- and mid-20th century (Winn, women from Georgia in Blot)
also eliminate the heterogeneity and the elevated risk (within American studies
as well as when the remaining American studies are combined with the
Scandinavian studies).

The authors may not agree with Rodu and others that this is the important
division; if so, they should certainly make that clear. But they should
acknowledge that some experts in the field think that this explains the
heterogeneity, and that it offers a similar improvement in heterogeneity test
statistics to the choices they propose.

At the start of the Conclusions, when the authors note the many differences that
produce the different estimates. They should reiterate that making the most
obvious and indisputable corrections to Boffetta's biases result in most of the
correction, and the bits that reasonably people might disagree on (e.g., whether
to draw the arbitrary line at reported estimates or derivable-from-the-article-text
estimates) are not what matters. They point out a few of the former immediately,
but do not actually make this point. (Also, I wonder whether, structurally, those
bits belong in the previous section, alongside the five glaring errors list. I have no
strong feelings about it, but I did wonder.)

Later in this passage, L&H suggest that Boffetta chose a particular result because it was the one significant result. Is there really evidence of this, or is it just as plausible to say they chose it because it was the biggest?

In the final paragraph, the authors say (they attribute it to Boffetta, but they express agreement) that the health effects of ST "need" to be better characterized. The word "need" is a bad one, meaningless without an "in order to..." phrase. Obviously there is no such existential need. The "in order to..." is important because clearly the claim is not true for every such. For example, there is no such need in order to be able to conclude that ST is approximately two orders of magnitude less harmful than smoking. Nor is it necessary to conclude that the risk for oral cancer is small enough that it cannot be reliably measured. The authors should either amend this to saying "it would be reassuring" or "it would be helpful" to know more, or specify exactly what they have in mind that we need to do more research in order to know.

The final sentence of the conclusion should appear in the abstract. This is an important scientific conclusion, and an important part of the value of the paper.

The combined COI and Acknowledgements statements leave the most critical bit of information about the role of the funder ambiguous: Who initiated the research? If the authors initiated the research using existing funding or asking for funding to support it, then they should explicitly say that it was author initiated. This may be what is intended by "independent scientific assessment", but if that is true then it should be made clear. (Also, the "independent" and "views of the authors" bits really should go without saying, but I do not object to them being there.) However, if the funder suggested the research, that should be noted. That does not change the quality of the work, but readers who are concerned about the influence of funding should be told this fact, which is much more important than merely who wrote the check.

Further on COI, the authors should note that they have the incentive to defend their own work. Anyone should recognize that for the present paper (as opposed to the original analysis), this is a far more compelling motive than is funding or anything else. I will disclose that this is a pet issue of mine, and I have written several papers that make similar points. But I really think the thoughtless fixation on funding would be dramatically reduced if authors were required to admit to the real motives, and I urge authors like L&H who like me are the target of that thoughtless fixation to lead the way by mentioning their stronger motives *first*. 