Reviewer's report

Title: Gender differentials in the evolution of cigarette smoking habits in a general European adult population from 1993-2003

Version: 2 Date: 5 December 2005

Reviewer: Paul Hewson

Reviewer's report:

General

The analysis is based on an impressive and thoroughly collected dataset.

-----------------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Given your title ("gender differences ") why have you analysed data for each gender separately (your first statement in Statistical Analysis section). Why haven't you included gender as a term (with interactions where necessary) in your model so that you can make a formal assessment of the role of gender?

When considering age groups (35-39) at 11 different sampling points, what allowance have you made for the different grouping structure. For example, if something dramatic happened in 1940 that had an effect on age of onset of smokers who were 60 in 2000, but that factor would be absent in smokers who started in 1930 and were 60 when surveyed in 1990. Equally, this factor would be present in smokers age 50 when surveyed in 1990. I would therefore like to see much more details of the models used, and an explanation of how these possible confounding factors have been avoided. I am nervous of graphs (figure 1 and figure 2) of crude data which have not been produced in a way that is sensitive to these possibilities. In particular, I am confused by the statement "young men started to smoke earlier than older men", as the younger men become older men (agreed, only by 11 years in this study, but unless you block out the year of the survey in some way you will tend to mask this effect). I would prefer that statement to be derived from a model parameter estimated from the data.

Do you really think that a "trend analysis", consisting of a line fitted to 11 years data is a reasonable way of modelling time series data? Whilst there may be insufficient data for formal time series methods I did wonder if a simple linear trend was just too simple for 11 years data. I also felt that these linear fits had been mis-applied in a few places. (It's not by biggest worry here, but are you running the risk of multiple comparison problems?)

How reliable is your age of "starting smoking" data. How much might differences in this figure be due to recall effects- for example if you asked 20 year olds when they started might you get a different figure from 60 year olds not because they figures are different but because memories are different?
It is not clear to me what analyses have been done; the graphical presentation of results is poor. It is becoming established practice in many medical journals to quote confidence intervals. On the lower panel of figure 2 it would have been most useful to have confidence intervals plotted. I was curious as to what you make of the 60-64 smoking initiation age?

The upper panel of Figure 2 should be some kind of density plot and I fail to see the point of the linear model fits. I can't read the key (legend) properly, but the group represented by grey bars (Former smoking?????) look like they have a peak at 50-54 (a bit like a very spread out and slightly assymetric bell curve). So why is it sensible to fit a linear to these data. Given the title of the paper (Gender differentials), where can I find a model that provides a formal compairon between the two groups?

The details given on your analysis of Cigarettes Smoked per day are inadequate. What models have you fitted, what are the parameter estimates, do you have any graphical or tabular summaries of the results.

What is the point of the lung cancer incidence data???????? What's the point of your comments on socio-economic status - might that be another confounding factor in this study????

I wondered whether (however formally) you should carry out some age-adjustment. In particular, might you have lost more heavy smokers in your older age groups because of smoking related illness which would bias the results slightly (I appreciate you might not be able to correct this formally, but you could acknowledge the potential problems).

Miscellaneous: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests