

Reviewer's report

Title: Sarcoma risk and dioxin emissions from incinerators and industrial plants: a population-based case-control study (Italy).

Version: 3 **Date:** 13 June 2007

Reviewer: Jean-Francois Viel

Reviewer's report:

General

The authors have made notable improvements to the manuscript, but some changes are still needed. This article is of importance in its field, and I strongly encourage the authors to pursue their editorial work.

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

1. Page 5, control sampling and residential history, and page 8, 5 first paragraphs: This major issue must be addressed. In fact the residential history is split in three parts: Residents since 1960 or from birth (if born before 1986); residents who moved before 1970; residents who moved after 1969.

Cases 168 4 9

Controls 384 21 25

This part of the manuscript is very confused. This results from several facts:

- cases who moved in the Province after 1969 were secondarily excluded (since no information was available on incinerators from other provinces), as well as childhood cases born after 1986 (when exposure had ceased),
- the residential history was reconstructed before controls who had a malignant tumour were excluded,
- some numbers don't match. For example: "42 controls, matched with the 14 cases [were] excluded from the study" (page 8, line 10). These 14 cases encompass 9 cases who have moved in the Province after 1969. The latter corresponded to 25 [and not 27] matched cases (see responses to the reviewers). The total therefore cannot amount to 42 controls (but 40 at most).

These crucial section of the manuscript must be totally rephrased, clearly stating the inclusion criteria for controls. It could be "people without malignant disease, residing in the Province since 1960 (beginning of the exposure), or having moved there before 1970 (since exposure is unlikely in other Provinces before 1970), or born in the Province before 1986 (end of the exposure period) were included as controls". Clarity and simplicity must absolutely be achieved (see also comment n° 18).

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

2. Page 2, line 7: Numbers at the beginning of a sentence are not allowed. "205" must therefore be replaced by "two hundred five".

3. Page 2, lines 22 to 25: This sentence, crucial since it appears in the abstract, must be clarified.

The risk of developing a sarcoma is 3.3 times higher (95% CI: 1.24 – 8.76) among women with the highest exposure level.

A significant excess is not observed "in the International Classification of Diseases, 9th Revision (ICD-IX) 171 site, both sexes (OR = 3.27, 95% CI: 1.35 – 7.93)" but for cancers of the connective and other soft tissue (ICD-9 171) (both sexes, OR = 3.27, 95% CI: 1.35 – 7.93).

4. Page 4, line 2: Data by themselves are not controversial, although they may be heterogeneous. However, their interpretation may be controversial.

5. Page 5, line 1: What is "NOS" ?

6. Page 5, lines 15-16: Which information has been "back-dated"? Do the authors mean they have used file: from censuses carried out before 1990? This point must be clarified.

7. Page 5, last sentence: What is a "protected location"?

8. Page 5: What about any ethical clearance or data confidentiality committee approval?

9. Page 7, third paragraph: The ISCLT3 acronym must be detailed (Industrial Source Complex Model in Long Term mode, version 3), as in the abstract.

10. Page 7: An atmospheric dispersion layer on Figure 1 map would be welcome.
11. Page 8, lines 20-22: A figure reporting the Odds Ratio curve as a function of the level of exposure used as continuous variable, would help. The 4 and 6 fg/m³ cut-off points are probably rounded. This should be clarified.
12. Page 8, last sentence. This analysis seems to be purely spatial. The maximum cluster size must be specified, as well as the number of Monte Carlo replications.
13. Page 9, line 8: The median value of length of exposure (32.74 years) is striking since it is closed to the maximum value (36 years, see Figure 2, while assuming that only adults present in 1961 are included in the study). A figure reporting the exposure distribution would help.
14. Page 9, lines 9-13 and Table 1: For the sake of clarity and homogeneity with Tables 2 and 3, average exposure classes should be presented in columns (rather than in rows).
15. Page 9, lines 14-15: Do the authors mean “In both sexes, risk increases in relation to the level of exposure, but reach statistical significance only for women, with an OR of 2.41 among the most exposed (95% CI: 1.04 – 5.85) and a significant test for trend ($p < 0.04$)”?
16. Page 9, 16-19, and Table 3: ICD-IX codes are used, while International Classification of Diseases for Oncology, second edition (ICDO-II) morphology codes are described in the “selection of cases and controls” section. Why two classifications? If ICD-IX is to be retained, the code descriptions must be given in the “selection of cases and controls” section.
17. Page 9, lines 21-23: These sentences are confusing. The spatial cluster is not “formed by 19 cases and 9 controls” since the controls are only used to reflect the background population. I would rephrase as following: “The spatial cluster consisted of 19 cases whereas 8.35 were expected, corresponding to a significant Relative Risk (RR) of 2.43 ($p=0.048$)”. By the way, 19 divided by 8.35 does not give 2.43 but 2.28... Pinpointing this cluster on a map would be nice.
18. Page 10, first paragraph: These are results of a sensitivity analysis that have nothing to do in a discussion section. They reflect the authors’ quandary regarding the control sampling and residential history (see major compulsory revision). However, this sensitivity analysis is a nice idea, but it must be announced in the “Material and methods” section and its results moved to the “Results” section. Moreover, the opportunity to extend this sensitivity analysis by including resident who moved after 1969 (and therefore address the major issue raised by the other reviewer) should be carefully envisaged by the authors.
19. Page 11, third paragraph: Are any Kappa values or agreement measures available to support these statements?
20. Page 12, last line and page 13, first line: It is strange to discover this statement in the second-half of a discussion, whereas it should appear in the early beginning of the introduction.
21. Page 16, references 6 and 7: Accession dates to these Web sites must be specified.
22. Page 22, Table 2 : Two decimal places are enough for p-values.
23. Page 23, Table 3: Two decimal places are enough for p-values. ICD-IX codes must be explained in foot-notes.

Discretionary Revisions (which the author can choose to ignore)

What next?: 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: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.