Reviewer’s report

Title: Space-time clusters of breast cancer using residential histories: A Danish case-control study

Version: 1 Date: 30 December 2013

Reviewer: Kate Hoffman

Reviewer’s report:

The authors present a spatial analysis of breast cancer in Denmark using case-control data for approximately 9,000 women. Among the studies strengths are the use of multiple control populations, statistical methods allowing for the control of potential spatial confounding, comparison of results using multiple statistical methods, and assessment residential history data which allowed the authors to consider latency and susceptible periods of exposure. Results of this work suggest clustering of breast cancer in Copenhagen. The authors hypothesize that spatial patterns are driven by differences in HRT, alcohol use, or socioeconomic factors in the cluster area.

Major revisions:

1. One important piece of information that is missing in these analyses is the magnitude of the effect of living in a cluster. Are women living in northern Copenhagen 4 times as likely to develop breast cancer or 1.05 times as likely? It is difficult to interpret the findings without this information. If not available with the statistical package, perhaps an inside the cluster vs outside the cluster odds ratio could be calculated and would provide some additional insights. Additional concerns and points of clarification are listed below:

2. Abstract: Given that the paper investigates time and space clustering, the abstract should include some information about the timing of clustering. In the abstract it is not clear whether clustering exists at a specific time point (calendar or age).

Minor Essential Revisions:

3. Line 64: A word appears to be missing. Increased risk?

4. Line 66-70: This sentence is difficult to follow. Breaking into two sentences may improve clarity.

5. Line 86: Whereby? Thereby?

6. Line 90: “previous spatial analyses have not lead to compelling new hypothesizes about environmental risk factors...” This seems somewhat contradictory to the previous sentences referencing the work of Vieira et al. and work presented in the discussion section.

7. Line 125: Why were residential history data limited to 1971? Previous research indicates that the residential location at the time of birth and menarche are more predicative of breast cancer risk than later addresses (as in Han et al. 2004);
however, these residential locations would be missing for a large percentage of the study population (based on the 1971 criteria). The authors should justify the cut-off, which may well be based on data availability, and discuss the implications in the discussion. While age seems less important than calendar year in these analyses, it may be because very few study participants have address information at the relevant age.

8. Line 134: “Seven percent had a less precise geocoding, because they matched at the municipality level, which means that the coordinates of the centroid of the municipality was assigned to these addresses.” Were the results sensitive to the inclusion of these individuals? While it seems unlikely with only 7% of study participant geocoded top the city center, this type of geocoding could mask or create clustering.

9. Residential histories: A map of residential locations or the density of residential locations of cases and controls would be helpful as it would provide information about population density which may be helpful in interpreting results.

10. Line 223: How did the authors arrive at these covariates? This should be described in the methods section.

11. Line 248: Is there a standard method for selecting k? A model fit statistic? More detail is needed to understand how the authors arrived at their k selection (and results). How might the choice of k impact areas of low and high population density differently?

12. Results: Some of the 3183 were not geocodeable and were not included in analyses. Include the number of cases and control in each figure (or in the text). Including the number of cases in each cluster would also be helpful.

13. Line 310: “When age was used as the underlying time scale, application of each of the control groups identified clusters in the area north of Copenhagen at several levels of k, also when the control groups were combined. The cluster areas existed when participants were in their 40’s to 60’s (results not shown).” It would be interesting to see these maps as a comparison.

14. Results: I commend the authors on the number of sensitivity analyses conducted, but, for readers less familiar with the region, it becomes quite a bit to keep track of in going through all the analyses. Perhaps a table with columns for the clusters and rows for analyses (e.g. control population 1, combined controls, adjusted, etc.) indicating which analyses showed clustering would be helpful.

15. Line 344: “The combined control group continued to identify two significant cases in Odense after the adjustment, but the second control group did not.” Could these cases be driven by the first control group? Consider adding this detail.

16. Discussion: Are there differences in breast cancer screening across Denmark? Differences in screening may results in higher incidence in communities with better access to care despite their having a similar incidence of the disease to other communities.

17. Table 1: Please include the percentages for the child birth variable.
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests