Author's response to reviews

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

Authors:

Rikke Baastrup Nordsborg (baastrup@cancer.dk)
Jaymie R Meliker (jrmeliker@gmail.com)
Annette Kjær Ersbøll (ake@si-folkesundhed.dk)
Geoffrey M Jacquez (gjacquez@buffalo.edu)
Aslak H Poulsen (aslak@cancer.dk)
Ole Raaschou-Nielsen (ole@cancer.dk)

Version: 2 Date: 7 March 2014

Author's response to reviews: see over
Dear editor,

We are pleased to have the opportunity to re-submit a revised version of the manuscript entitled: **Space-time clusters of breast cancer using residential histories: A Danish case-control study.** Below are our responses to the reviewers’ thoughtful comments, which have been used to improve the manuscript.

**Reviewer # 1**

Discretionary Revisions

1. Line 64. The word “risk” is missing at the end of the sentence.
   *The word “risk” has been added (line 66).*

2. Fig. 2a. The caption should be improved to clarify the content of the insert map.
   *The figure legends of figures 2 and 3 have been improved to clarify that the inserted maps show the Odense area.*

**Reviewer # 2**

1. The paper is too long and must be reduced in the “methods” part
   *We agree and have shortened the methods and results sections of the paper.*

2. Is there any clinical particularities in the breast cancer of women leaving in the cluster with the highest incidence: younger age at diagnosis, more frequent family history of breast cancer? more aggressive course of the cancer? It is very important to have these clinical data to give details in the Discussion part. What are the arguments for a genetic cause in the presence of this cluster? Part of environmental factors?
   *This is an important suggestion. We examined age and extent of tumour at date of diagnosis separately for cases that had lived inside (for at least 5 years) and outside the northern Copenhagen cluster area. Cases inside the area were on average 7 years younger at date of diagnosis and had a lower frequency of metastases compared to cases outside the area. This is not surprising since studies of social inequality in breast cancer stage and survival have shown that affluent women are diagnosed at earlier stages and have better survival rates compared to deprived women, and the particular area indeed has a wealthy population. These details have been added to the results and discussion sections (line 300-302 and 361-396).*
Data on family history of cancer and clinical details are not available in the Danish Cancer Register, thus we could not address this question further.

Familial aggregation of breast cancer could potentially create space-time clusters, if members of the same family lived close by one another for a considerable period of time. However, only 5-10% of breast cancer is thought to result directly from inherited gene defects, and considering the size of the cluster detected in the present study it seems very unlikely that it would be due to familial aggregation.

3. The authors should separate the results into 2 parts: constant space-time cluster and non constant space-time clusters. The results will be more easily to read. We have not separated the results as suggested, because we think it is more logical to present the results according to the different methods that were applied. However; we have produced a table (table 2) that summarizes the results of the different cluster analyses, and we have rewritten the results section to make it easier to read.

4. Can the authors briefly give arguments (exhaustivity rate?) of the completeness collection of cancers in the population-based Danish Cancer Registry?

The Danish Cancer Registry is virtually complete, because 1) it is mandatory to report all new cancer cases to the register, and 2) the register is linked with the Danish National Patient Register, the Danish Pathology Register and the Danish Registry of Causes of Death, which secures that virtually all cases are recorded. The completeness is emphasized as study strength in the discussion (line 434-436) and we have added a sentence on this in the methods section (line 105-107). For further details please see reference 21 in the manuscript.

5. Regarding the Table 1, the authors must give the detailed statistical results in the comparison between cases and controls regarding the frequency of child birth, the repartition of the qualitative variable on number of children (0,1,2 and #3), the age at first child birth, the aggregated educational level in area and the aggregated average income in area. Indeed it was not evident that there is a significant difference between cases and controls regarding these data. We have tested the differences between cases and controls according to these variables and added the test p-values to table 1.
Reviewer # 3

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.

*This is a good suggestion. Cases and controls are individually matched on age and followed-up for cancer in the cancer register, therefore it is possible to calculated IRR instead of OR. We have calculated IRR (called relative risk in the manuscript) of breast cancer associated with living inside vs. outside the cluster area in northern Copenhagen and found that living inside the cluster area for at least 5 years resulted in a relative risk of 1.39 (95%-CI 1.11-1.74). This result has been added at the end of the results section (line 298-300).*

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).

*The temporal extent of the cluster has been added to the abstract (line 42).*

Minor Essential Revisions:

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

*“Risk” has been added (line 66).*

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

*The sentence has been divided in two (line 68-71).*

5. Line 86: Whereby? Thereby?

*The sentence has been reworded (line 88).*
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.
*These sentences have been modified to avoid this disagreement (line 92).*

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.
*The 1971 cut-off is indeed due to data availability, which we now explain in the methods section (line 126). We also address this limitation in the discussion section (line 453-455).*

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.
*We do not expect this type of geocoding to create any clusters, because this geocoding method did not depend on case-control status. We have repeated a number of analyses without the addresses that were geocoded to municipality center, and results were similar to those obtained with all addresses. We now mention this geocoding uncertainty as a limitation of the study in the discussion section (455-458).*

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.
In the overview map in figure 1, we have added the density of the residential addresses of the study population by municipality. The density is highest in the capital area and low in rural areas especially the western part of Denmark. Since this is a population-based sample, the pattern reflects the general population density in Denmark.

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

We searched the literature for established risk factors for breast cancer, however only information on some of the factors were available from registers and databases, which determined if we could include them in the study. We have added this information to the methods section (line 205).

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?

There is no standard method for choice of k, therefore we based our choice of k on the results from the simulation study (Sloan et al 2012). We have added this information in the methods section (line 179). The population density and choice of k will influence the size of detectable clusters. In densely populated areas a k of e.g. 20 would correspond to a small area, while in a sparsely populated area 20 nearest neighbours would live further apart, hence a detected cluster would cover a larger area.

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.

Some addresses of cases and controls were not geocodeable, but all cases and controls contributed to the study with at least one geocoded residential address. The number of cases in each of the clusters has been added to the figures 2-4.

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.

The results were very similar regarding the spatial location of clusters, and we did not include these maps because the manuscript is already relatively comprehensive. However, we now emphasize that the spatial location of the cluster based on the age time scale was similar to that detected with the calendar year time scale (line 264).

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.

We agree that the results section is extensive, and we have tried to shorten it and emphasise the central findings. We have created an additional table as suggested (table 2) in which we summarize the findings by area, method and control group, we hope this will provide some clarification of the results.

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.

This seems unlikely, as the first control group did not identify any significant cases in Odense in any of the analyses. To make this clear we have changed the text to: “The combined control group continued to identify two significant cases in Odense after the adjustment, but applying the two control groups separately did not” (line 291-293).

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

There was no organised breast cancer screening programme established in the municipalities north of Copenhagen at the time when cases of the present study were diagnosed, but we cannot exclude that affluent women are more likely to seek breast cancer screening on their own initiative than deprived women, which could have contributed to the observed cluster. We now mention this in the discussion (line 426-430).
17. Table 1: Please include the percentages for the child birth variable. 

The numbers have been added to the table.

Additional minor changes of the references and the reference list have been made.

Sincerely,

Rikke Baastrup Nordsborg

Postdoc, M.Sc., Ph.D.

Danish Cancer Society Research Center
Strandboulevarden 49
Dk-2100 Copenhagen
Phone: +45 35 25 76 86
E-mail: baastrup@cancer.dk