Reviewer’s report

Title: Spatial analysis of bladder, kidney, and pancreatic cancer on upper Cape Cod: An application of generalized additive models to case-control data

Version: 1 Date: 16 December 2008

Reviewer: Jaymie Meliker

Reviewer’s report:

This manuscript follows up on previous applications by this research group on the use of GAMs for spatial analysis of hot-spots and cool-spots in case-control data. This is another solid contribution of the approach using data from population-based case-control studies, this time bladder, kidney, and pancreatic cancer. Overall, this is a very well-written, well-conducted, interesting analysis.

Minor essential revisions:

(1) Case-control recruitment: When the study population is first introduced, on page 4, I would like more details about how the controls compare with the cases in terms of spatial recruitment. Am I to understand that each case (e.g., bladder cancer cases) has ~14 controls living in the same town around time of diagnosis? How does the original matching on town influence the spatial distribution of the case-control populations in this study?

I worry that if the town-based matching was well executed, then these data may be over-matched on spatial factors, making it difficult to detect any underlying spatial factors that may be present. The latency analysis is one way to get around this problem, and may help explain why clustering was detected using that approach. I don’t think this recruitment strategy would be a large problem using historical residences, but by including residences at time of recruitment in the analysis, this recruitment strategy might be biasing the results.

The authors should consider and discuss how recruitment might influence their results.

(2) I wish there were greater numbers of cases available to enable something more than spatial-only analyses of temporally aggregated data. One of the great strengths of the GAM approach is the ability to identify spatio-temporal clustering using time-resolved mobility histories—it is too bad this dataset did not permit this type of temporally detailed analysis.

Understandably, the authors chose not to consider the full range of temporal variability in their data because of small sample size. However, in the discussion, I would like to see mention that residential histories in case-control studies provide a valuable resource for generating hypotheses about location and timing of exposure. More temporally resolved analyses might provide greater insights into importing time periods of susceptibility. Similarly, mapping where people live
at different ages, also might allow for observing different patterns that could lead to new hypotheses.

(3) At the end of the first paragraph on page 12, the authors state that there was a lack of clustering in the kidney and pancreatic cancer analyses. But I understood the results to indicate that clustering was present in these analyses. A hot-spot along the southern shore for kidney cancer, and a couple reddish areas for pancreatic cancer that were significant after accounting for alcohol-related behavior. Am I misinterpreting the results? This requires clarification.

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

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests