

## Reviewer's report

**Title:** Cluster detection methods applied to the Upper Cape Cod cancer data

**Version:** 1 **Date:** 12 March 2005

**Reviewer:** Dan Wartenberg

### Reviewer's report:

#### General

The goal of this paper, to compare three methods for detecting disease clusters, is an admirable but complicated one. This is not a new question, but one approached from a data-based rather than simulation-based strategy. The authors suggest that using real data is preferable to simulated data, as this approach can address issues related features of real data often not captured by the simulations. The downside of this approach, however, is that one can develop after the fact interpretations and explanations for virtually any patterns found in any data set (or in any single simulation run). It can be risky to generalize from the analysis of one or a few data sets.

The authors compare methods with three different strategies: global clustering, local clustering, and a combined index. They apply the methods to one data set of breast cancer incidence on Cape Cod and apply them separately for three latency periods (actually, residence at the same location): 0 years, 15 years and 20 years. They do not adjust for potentially confounding variables.

The authors report that the three methods are discordant for each test. That is, at the  $p < 0.05$  level, the scan statistic does not find any clustering, the M-statistic finds only the 15 and 20 year latency data sets statistically significant, and the GAM approach finds all data sets statistically significant.

The discussion section is only partially tied to the results and talks more generally about the formulation, strengths and limitations of each of the methods. While useful, the connection between the data used and the interpretations made is so loose I do not appreciate the need for the data. I think the paper would be improved by either focusing more on the theoretical aspects of the methods and conducting simulations tailored to demonstrate the strengths and limitations, or by using additional data sets with different characteristics as a means for trying to understand the variation in methods and results. Additionally, the discussion of adjustments for covariates is not relevant to the performance of the methods as it was not used for this study. It is, of course, relevant to the choice of methods for real-world applications but is perhaps not related to the goals of this manuscript.

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

While this study is interesting, and the writing is clear, the bulk of the paper describes the methods and then discusses these methods somewhat independently of the data. It has an interesting and useful goal but, as noted earlier, I believe the paper needs either more data with associated interpretations, or no data with further discussion of the formulation of the methods and how that could guide their usage. In addition, I believe the paper would be improved by addressing some of the other issues mentioned in my comments.

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

-----  
Discretionary Revisions (which the author can choose to ignore)

There are four main issues that this paper raises but does not resolve. I suggest the authors address some or all of these. First, is the goal of this type of analysis assessment of statistical significance or evaluation of the risk difference? Many epidemiologists believe that measures of effect are more relevant than statistical significance values for substantive interpretation; others disagree. The methods used in this study differ with respect to how their results are presented and interpreted, and this paper should address the issue and make specific recommendations. Second, the authors should specify more clearly what type of pattern or patterns they are looking for, and use the appropriate methods to describe or identify it. Since these methods differ in their approaches, it was not clear to me why all three methods were used instead of one or two focused to the specific goals of the study. It is well known that different statistical methods have different sensitivities to different types of patterns. Third, it would be helpful if the authors provided some evaluation of false positives and false negatives, either their own or from the literature. For this study, one cannot determine (other than heuristically) whether the difference results reported are due mainly to differences in sensitivity, such as false positive and false negatives, or due to more substantive differences in the formulation of the methods. That is, can one method perform better than the others at detecting a specific pattern, say a single circular cluster, but worse at other patterns, or is one always more sensitive? Fourth, the authors should state whether the goal of the analysis is mainly descriptive, for example to identify subregions for further analysis, or is it hypothesis testing, for example to evaluate an independently developed observation.

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**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 to the best of my knowledge I have no competing interests.