Reviewer's report

Title: Detecting and diagnosing outbreaks for the enhanced management of emergency departments in Queensland, Australia

Version: 1 Date: 4 November 2012

Reviewer: Howard S Burkom

Reviewer's report:

Thank you for the opportunity to review this manuscript. Using a novel method applied by the second author to traffic crashes, the authors create a multivariate method for hospital-based disease surveillance with a surveillance tree approach that combines several methods: a regression model applied to total visits, expected proportions of individual cells applied to the modeled counts, and EWMA smoothing. From my reading, their approach is novel and potentially fruitful for public health surveillance. The paper also has a number of shortcomings that need to be addressed:

Major Compulsory Revisions:

1. The principal issue is that the explanations are too casual and require related technical experience from the reader. I have worked on similar ideas recently but needed to read the methods sections several times and eventually had to consult Reference 1 for basic details that should be clear from the text. In particular, examples from the hospital data using several different variables are needed to clarify the formation and growing of the surveillance tree.

2. The text oversells with a number of unsupported claims and implications. The claims should be either deleted or justified with evidence.

p. 2, “However these tools can never capture the full complexity of disease instances...”. Do the authors claim that surveillance trees do capture the full complexity?

p. 3, How do surveillance trees combine “aspects from all the aforementioned methods”? What aspects are meant?

p. 7, line 4: Please delete or tone down the statement that the variables “are exploited to their full potential”.

p. 7, line 17: Please explain the “inclusion of domain knowledge at the two different scales”. For prediction of total visits, is the selection of independent variables for regression considered “domain knowledge”? To me, domain knowledge here implies medical expert advice, not inferences from historical data. What domain knowledge is involved in the cell estimates? These estimates seem to be driven only by empirical proportions. The only evident domain knowledge was in the selection of code groups.

3. Please address a few specific technical issues.
a. The Factors code group described on p. 4 looks interesting, and I have not seen it in other literature. However, the description as “generalist codes that appeared to be particularly prevalent in the winter crisis period” is not sufficient. The authors should consider providing codes underlying all of the groups in Appendix tables, but examples of the Factors group are definitely needed with explanation. This implementation of domain expertise should be shared.

b. Please clarify whether the “disease groups” noted in subsequent discussion refer to the same 3 code groups on pages 4-5. If so, the language should be changed; these groups are not all disease groups. If not, what is meant by “disease groups”?

c. Please explicitly describe what is meant by the phrases “training phase”, the “training period”, and the “training data”. The section “Training the Model” should give an orderly explanation with careful definitions.

d. The independent variables of the total visits regression model should be specified, not just described as variable types as on page 6.

e. The notation Yt is used to denote both total counts (p. 7) and cell counts (p. 8). Please make clear how the expected value and variance of total count predictions are calculated. The variance computation should account for the regression model. How is the variance of a cell prediction calculated? The variance calculation does not seem to account for variability in the cell proportion--please clarify.

f. I could not understand the results section on Diagnostic Ability. Do the groups g1 and g2 in that section represent the 3 code groups described earlier? Figures 7 and 8 and their captions did not explain. An example is needed.

4. An important language issue is that statistically significant findings of the surveillance tree method should not be considered “outbreaks”. There are numerous reasons why certain tree nodes that are not outbreak-related can appear significant, some given by the limitations discussed by the authors. A number of epidemiologists have objected to this usage in my own work and in that of others. Please make it clear that the statistical alerts are valuable for directing the investigations of infection control staff and other epidemiologists, but they will often not indicate what the health monitors consider to be outbreaks.

5. The idea of Figure 10 is the type of example that the paper needs, but more illustration is needed. Please give details of a couple of the nodes whose results are represented in the Figure.

Minor Essential Revisions

8. Typographical errors and careless use of language are excessive and distracting. The manuscript needs a copy-edit, especially for punctuation, but also for occasional misspellings, and noun/verb agreement (for example, the word “data” should be treated as plural). In addition, demonstrative pronouns without clear antecedents put an unnecessary burden on the reader. A reader whose native language is not English would have trouble following the text as written.
Minor Essential Revisions

p. 1, line 10 of abstract: From the results on pp. 12-14, concerned with the number of simulated events detected, the nondescript term “effectiveness” should be replace by “sensitivity”

p. 2, Conclusions: The authors’ idea is not just “multivariate surveillance”. Many authors use this term to refer to surveillance with multiple data types or code groups. A phrase like “partitioned surveillance” might distinguish this approach.

p. 2, line 14 of Background: Please change “an underlying change in the numbers…” to “a change in the process underlying the number of patient presentations”. Isn’t this the idea?

p. 2, line 18 of Background: Please reorder to say “a new strain of flu might...affect a particular age group, say pre-schoolers, more dramatically”.

p. 3, line 5: Insert the missing word in “…behavior of subgroups is likely to [be] correlated….”. I won’t list all of these, but these are why a copy-edit is needed.

p. 3, line 16: The statement: “However, multivariate control chart methods …do not directly incorporate the means to break down a signal to determine which sub-groups contributed to a detected increase” will not explain the problem to anyone who doesn’t already understand it. Make the point that these methods find change points in sets of time series but do not identify the responsible component series.

p. 4, line 9: The surveillance variables correspond to the rows of the table, not to its dimensions.

p. 6, line 1: In “total counts to be modelled over time”, do the authors mean total visit counts? Please clarify.

p. 7, line 1: The authors have just given a limitation that should just be stated as a limitation. Please delete the statement “This is not really an issue because if the `true' process is not understood in the domain area then delivering outliers relative to this model will help to understand these interactions.” The outliers do not necessarily explain because they may arise for many reasons.

p. 9, line 5: The phrase “we calculate the value of the test statistic for all sub-regions that can be generated by taking binary partitions along any surveillance variable” requires examples, as in the first reference.

p. 10, line 13: Please add the missing words in “Once [the model] is trained, then giving an incoming stream…”, or else the statement does not make sense.

p. 14, bottom: What is meant by “the true outbreak type”? This section needs clarification and rewording.

Discretionary Revisions

No authentic outbreak is mentioned in the Illustrative Examples, so I don’t see how the authors can refer to the “natural clustering of disease in subgroups” in
the Conclusions. Only simulated data effects of outbreaks are described. One underlying principle that could be expressed either in the Discussion or Conclusions is that the simple EWMA approach will always have the advantage seen in the first comparison if one knows exactly which time series to test, but testing every possible time series is not a practical choice, as described on page 3.

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

**Quality of written English:** Not suitable for publication unless extensively edited

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

**Declaration of competing interests:**

I declare that I have no competing interests,