Reviewer’s report

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

Version: 7 Date: 10 October 2013

Reviewer: Howard S Burkom

Reviewer’s report:

I still feel that the work presented in this manuscript is valuable and worthy of publication. The most substantial criticisms have been answered. The Associate Editor may choose to accept the manuscript if the authors will supply the additional explanations requested below. If not, I would be willing to look at another draft at short notice, even via email.

Of major points 1-11 from the previous review, I wish further clarification of points 2, 3, 7, and 8.

Major Compulsory changes are requested only for point 2.

Regarding questions 2 and 3:

Q2. On p. 11, the node-specific threshold is purported to be “independent of the properties of the node”, including the mean value…Please explain.

Q3. Is the pruning as straightforward as dropping a node if its “smoothed” z-score is larger than the maximum of (threshold, parent’s smoothed z-score)? Or is there more involved?

Only question 2 was answered, but I will assume a “yes” response to question 3. The added explanation of the threshold derivation is obscure. What “threshold model’ was created for each partition? To me, details of the threshold computation are essential. I stress again that the conversational style and technical slang make the work difficult for readers whose first language is not English.

Minor Essential Revisions:

Q7. In the description of the injection procedure on p. 12, we are told that “1000 dates in 2009 were selected at random with replacement and using these as start points, in-control counts were simulated forward in time according to the models developed on the training data. To each of these data sets we added artificial hotspots.” Does this mean that 1000 copies of the time series were created and a different interval of injections added to each copy? Authors have done such injection in a variety of ways, and the methodology may influence the results. Please elaborate. Also, please add a statement about the importance of deciding which strata to spike and any procedures used.

I am confused by the statement “To each of these samples we added artificial hotspots at the beginning of the series (so the hotspot starts at a random point
during 2009) but then its existence is hidden from the methodology.” I had thought that the entire 2009 data interval was used for testing, with hotspots added at a different randomly chosen location in each of 1000 repetitions. The statement that hotspots were added at the “beginning of the series” suggests that the testing begins at the date randomly selected for spiking, not at the beginning of 2009. Is no warm-up period used for the model in the test data? I ask this question because the model involves day-ahead forecasts.

[8. Clarification is needed for the statement “The univariate control chart is trained to have approximately the same false alarm rate as the Surveillance Trees (135.32 and 134.65 respectively).” Is the idea that the resultant false alarm rate was 135.2 for the univariate method and 134.65 for the Surveillance Tree? Were these rates calculated from the full training sets, with no outlier removal? Please explain what was done.]  

The explanation was what I already knew and did not address the question. Another way of asking is: how were the charts trained? Were the reported rates the result of applying trial thresholds to the full training sets with no outlier removal, until the resultant alarm rates were nearly equal? The authors should not assume that BMC readers know what “training a chart” entails. I am not asking for a primer on control charts, but only brief explanatory statements.

Thank you also for addressing many of the minor language issues, though please correct the spelling in “ad[đ]ressing the computational challenges” on page 6 and in “there w[j]ere 7 branches left” on p. 19.

Regarding the use of demonstrative pronouns without clear antecedents, there is an apparent disagreement between me and Mr. McGregor. The BMC Associate Editor may agree with him, and I don’t wish to delay publication over this disagreement. The list below has the instances that I feel need correction, with suggested additional wording. Please consider these, and if some are not what the authors intended, the need for explicit wording will be supported.

p. 3 “This is crucial when monitoring…” => “Accounting for such underlying effects is crucial when monitoring…”

p. 5 “This holds true for the method presented in this paper.” => “These phases are applicable to the method presented in this paper.”

p. 15, “We decided to use the same in-control time-to-signal … but recognise that this offers an unfair comparison (in favour of the univariate chart). => “…but recognise that this criterion choice offers an unfair…” [and also] “This is because both methods are trying to …” => “The presumable advantage to the univariate chart is that both methods are trying to…”

p. 16 “Note that this is the situation for syndromic surveillance…” => “Note that a variety of assigned diagnosis codes across categories is commonplace in syndromic surveillance…”

“We illustrate this in Figure 5a…” => “We illustrate this weakness of univariate monitoring in Figure 5a…”
p. 17 “This is illustrated in Figure 5b and …” => “This effect on a subgroup without changing the total is illustrated in Figure 5b and …” [and also]
“In both cases this is due to…” => “In both cases this detection success is due to…”

p. 18 “This is not surprising as this is the timing of the usual seasonal increase which has higher variance” => “This variability in detection times is not surprising because of the year-to-year variance observed in onset of the seasonal increase”.

“This is illustrated in Figure 10” => “This uncertainty in capturing the truly affected population is illustrated in Figure 10”

p. 19 “when perfect matching is not achieved this is due to over or under specification …” => “imperfect matching is due to over- or under-specification …”

p. 20 “We could consider therefore that this is a different hotspot than” => “We could consider therefore that the hotspot responsible for the second week’s displayed variables is different from the hotspot causing the first week’s display, and hence the disease process may have changed.”

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

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.