Author's response to reviews

Title: in silico Surveillance: Evaluating Outbreak Detection with Simulation Models

Authors:

Bryan L Lewis (blewis@vbi.vt.edu)
Stephen Eubank (seubank@vbi.vt.edu)
Allyson M Abrams (allyson.abrams@gmail.com)
Ken Kleinman (ken.kleinman@hms.harvard.edu)

Version: 2 Date: 5 October 2012

Author's response to reviews: see over
Oct 5, 2011

We thank the reviews provided and have made several changes to the manuscript to reflect the the concerns voiced. We recognize that this is a complicated experiment and have made some choices to keep the manuscript focused on the methodology used to harness the simulation for the purposes of evaluating surveillance systems. We feel that discussing the simulation itself and or the detection algorithm in more detail is more appropriate for other previously published works, and that detail in this manuscript would further complicate an already complicated narrative.

Thank you again for your time reviewing this manuscript,
Bryan Lewis, Stephen Eubank, Allyson Abrams, and Ken Kleinman

Responses to individual comments:

Reviewer 1: Poh-Chin Lai
Reviewer 2: Colin Robertson

**Major / Compulsory Revisions:**

**Reviewer 1:**

1. Background and general concept as illustrated in Figure 1 seemed logical. However, discussions thereafter (especially the sections on care seeking modeling, surveillance system, and artificial outbreaks) became confusing. How were these related to the processes described in Figure 1?

The disease model illustrated in Figure 1 was used to simulate the spread of the background ILI disease in the population. Care-seeking model was triggered when an infected individual experienced symptoms. If the individual sought care and happens to be a member of a healthcare system providing data to the surveillance system, their report of illness was included in the overall surveillance system. Artificial outbreaks were used to provide a standardized way to test the sensitivity and specificity of the surveillance system’s ability to detect outbreaks. This is explained in a straight forward manner in the text.

2. While it is important to provide different options for surveillance, it is also important for the authors to discuss clearly their differences in principle. It would certainly benefit the readers if more detailed descriptions on one of the six surveillance designs were provided.
These differences were described in the “Surveillance System” section. The “Natural 6” system is an operational surveillance system that is actively used by the Harvard Pilgrim Health Care (HPHC) system. As described it represents 246K members spread across 755 zip codes in the Boston Metro area. Natural 12, 18 represent this system with expanded coverage (ie 2x and 3x more members but distributed in the same geographic proportions). The Uniform systems represent an artificial surveillance system based on uniform sampling of the population by zipcode. More detail than what is included is unwarranted since their inclusion is for illustrative purposes only.

3. Why use “artificial outbreaks” to evaluate outbreak detection ability of the system when the authors seemingly have access to historical data (pages 6, 9, 10)?

Historical data is insufficient since one never truly knows how many outbreaks have been missed by the surveillance system. The purpose of this methodology is to provide a way to recreate historical like data (in large enough volumes to evaluate performance with statistical significance). The choice of, artificial outbreaks were used to provide a consistent standardized “signal” to test the performance of the different surveillance systems. This allows one to thoroughly assess the ability of the surveillance system to detect an outbreak of this size (rather than historically accurate but less well defined outbreaks). This approach is commonly used in the outbreak detection community.

4. What is the relevance of Figure 3 and how can such a map aid in intervention or control of disease outbreaks?

Figure 3, provides an information dense overview of the entire process described in this paper and was included to provide a vehicle to discuss detection of the outbreaks and the process by which they were classified. This methodology is intended to help evaluate the performance of existing surveillance systems and provide guidance on their improvement.

5. The last Figure was not labeled and not discussed in the text. However, it does appear to show interesting results comparing performances between real world data against simulated results.

This figure was uploaded in error and was not meant to be included

6. This reviewer feels that the authors tried to cover too much with insufficient detail and flawed methodology.

Indeed this is a rather complicated experiment presented in a condensed form. Initial efforts to publish this manuscript as a purely methodological piece met resistance for the lack of a concrete example. We feel that while complicated, there is sufficient information provided for a reader to understand the concept of the process we undertook, and to impart the utility of highly-detailed simulations for addressing problems that traditional epidemiology can't address. This
Reviewer 2:
1. While references are provided to previous papers describing agent-based population data simulation, some description of the approach is warranted here. As it reads now it is impossible for a reader to have a clear understanding of the types of decisions involved in parameterizing such a model from this paper alone. The language describing the population simulation is very general, and more detail is definitely needed here.

The details of the agent-based simulation are addressed in some of the supplementary materials (website). Indeed parameterizing the agent based model required many decisions, the major and more general decisions are discussed in this manuscript. We feel that including more detail in this paper (rather than refering to previously published work) would distract from the main thrust of this manuscript, explaining how such a system was employed to evaluate surveillance systems.

2. The paper strings together a series of models into a surveillance system evaluation scheme, however very little is reported about the fit of these models. This was most glaring for the baseline disease case data allocated to the simulated population. Figure 2 shows the epidemic curves look similar, but how similar were they? While stochasticity is inherent in the system, confidence bands generated from a number of simulations would provide a clearer depiction of the simulated data.

There is not sufficient historical data to do a proper analysis of how similar it is to the stochastic simulations. Additionally, the point was not to replicate the exact same counts but to provide a platform that provided realistic counts of influenza like illness. Providing an analysis of the variance of the simulations in this manuscript was not done, because there is extreme variability and it was thought that this might distract from the main point that of 100 stochastic realizations there are several that are similar to the historical record and the rest are “structurally” similar, yet absolutely different in terms of timing and exact counts.

3. I found the terms ‘natural’ and ‘uniform’ confusing. These might better be described as ‘random’ and ‘proportional’ or ‘stratified’ sampling designs, else have the terms used described in greater clarity.

These terms were chosen to convey that the “Natural” surveillance system were based on a real-world distribution of surveillance system membership, as in “naturally occuring.” We’ve addressed this concern by providing more description in the “Surveillance System” section of the Methods.
4. The spatial configuration of the outbreak injection was also confusing to me. The simple procedure of assigning cases ‘near by’ to randomly selected seeds seems at odds with the realistic simulation approach taken with the baseline data. Why not simulate realistic interactions based on demographics and activity profiles to allocate cases during outbreaks as well? Some discussion of the limitations imposed by the method employed is warranted.

Certainly the approach of simulating “true” outbreaks using similar techniques could be used, however, we wanted to limit the variances in the comparison metrics. A “fixed” artificial outbreak allows the surveillance system being evaluated to be judged against a consistent standard for the size and “spread” of an outbreak that it could be expected to detect. Extending this method a variety of standardized outbreaks could be evaluated allowing one to find the “kinds” of outbreaks the detection method and surveillance system are able to detect and which kinds elude detection.

5. The expected false positive rates need to be described in greater detail. How were these expected #s arrived at, and why were the true FP rates so high? The conclusion (paragraph 2) cites this as a validation of the agent-based model. I do not understand this reasoning. This would seem to me to be a poorly modeled baseline. Why wouldn’t the risk ratios in the s-t scan method be adjusted to represent seasonal variability? Was this done? More details needed on the parameterization of the scanning method.

We believe this is directed towards the content of the 4th paragraph of conclusions (2nd paragraph on page 13). We’ve added a quick example of the calculation that can help explain why the nominal false-positive rate would be 36.5. At random one would expect a false positive once every 1000 days when using the p=0.001 threshold, and thus for 365*100 days, we’d expect 36.5 false-positive signals. This supports that the simulation is creating spatio-temporal clusters at a higher rate than purely random distribution could. Creating patterns like this while maintaining the structural validity of the system is the strength of this technique.

The SaTScan method uses a 90 day history to generate the distribution of “normal’ patterns, and as a result is able to incorporate seasonal variability. The scan method is described in a condensed form in the Outbreak Detection section and details about the particular parameterization is discussed in the cited works (ref 20,21)

Reviewer 1

Minor Essential Revisions

1. Figure 2 is not properly labeled. Which of red or black dots represent real versus simulated data?
It was a deliberate choice to leave this out of the legend, since assessing the similarity between two pairs of data should not be biased by which is “actual” and “test”, however the caption provides this information.

2. Keys for Figures 4 and 5 were not consistent. The natural and uniform designs should have different shades with internal classes shown in different line patterns.

Indeed the figure uploaded was incorrect. More consistent one is included with this revision.

**Discretionary Revisions**

1. Perhaps more discussions on how the animated movies (page 11) can be used to inform intervention and control measures.

The movies are not intended to inform intervention and control measures, they are merely illustrative of the complicated evaluation process described in the manuscript.

**Reviewer 2**

**Minor Essential**

1. Why increase susceptibility to induce seasonality? (as opposed to increasing contact rates (e.g., school attendance) or overall viral load). Justification needed. How does this assumption impact your analysis?

Seasonality of ILI has many factors, increasing susceptibility (akin to increased infectiousness of the pathogen) is supported as one of the main drivers (see works of Shaman et al).

2. Figure 5 – the lines are different colour than they are on Figure 4 which makes interpreting difficult, revise to line symbology used in Figure 4.

Indeed the figure uploaded was incorrect. More consistent one is included with this revision.
3. Some discussion of the practical significance of timeliness of outbreak detection for influenza (what does it mean to detect an outbreak in 5 days vs 8.5 days)?

We include this analysis as an example of the many metrics that one could use to evaluate the performance of their surveillance system and use to optimize its performance. Generally the sooner one can detect the better, however, the impacts of a particular delay is specific to the surveillance system and health system using it.

Discretionary

Figure 3 – the Voronoi tessellation adds confusion rather than clarity and the histograms are not legible, even when printed full size on 8.5x11 paper. The zip boundaries (included and excluded) and cluster locations would be a more effective representation in map form.

We understand where this comment comes from, the decision to visualize this system was very carefully considered. Voronoi tessellations were chosen to enhance clarity, sorry if this seems to be counter productive, but zip-code boundaries are very irregularly shaped thus adding more confusing angles and also leads to boundary-histogram overlaps. More importantly perhaps, the SaTScan outbreak detection process relies on the centroid of the zip-code area, making a Voronoi tesselation a more natural representation of how the process actually works. Indeed the image size restrictions placed by the journal renders these visualizations difficult to see and interpret. The Supplemental materials available on the external website linked in the Results section (page 11) provides access to a much higher resolution images.