Author's response to reviews

Title: Modeling and detection of respiratory-related outbreaks using chest radiograph data

Authors:

Peter F Craigmile (pfc@stat.ohio-state.edu)
Namhee Kim (kim.1862@osu.edu)
Soledad Fernandez (soledad.fernandez@osumc.edu)
Bema Bonsub (bonsub@columbus.rr.com)

Version: 2 Date: 2 April 2007

Author's response to reviews: see over
Response to Reviewer 1: Howard S. Burkom

Many thanks for the constructive review. We hope that our revision addresses your points.

Responses to Major Compulsory Revisions

1. In the revised manuscript, we now include the motivation for the inclusion of covariates, as per your suggestion in the Background section. In the simulations we now compare our approach to two other models to show the effect of including covariates and random seasonal terms. We rephrased the sentence about “changes in the future”.

2. Thank you for your suggestions on restructuring the Background and Discussion. We have made those changes in the revision.

3. The first day of the outbreak is now randomly picked from a uniform distribution in the period December 5th, 2003 to June 22nd, 2004. To reflect this change, we used the first ten months of data as the training set, and the last ten months are the test set. See the bottom of page 11.

4. We agree that this Figure was not very well explained. In the revised manuscript, we include more detailed discussion and a new simplified Figure (Figure 5) which we hope explains the concept of the time-varying quantiles more clearly. We include a small discussion of the two choices of threshold.

5. The low sensitivities were due to an error in the simulation of the prodromal and fulminant durations for each individual. Sorry about that. The simulations results have changed as a consequence of this and sensitivities are high.

Responses to Minor Essential Revisions

1. We amended the abstract to highlight the general applicability of time-series-based methods.

2. We are more careful with the use of the term “outbreak”, highlight the key fact that what we observe is the presence of outbreak as expressed through the extra chest radiograph counts in the ED. We hope the language is more precise now.
3. The health-utilization model has been improved in the revised manuscript, thanks to your suggestions – please see pages 7–8. Now, the probabilities of entering the ED for the prodromal and fulminant stages are not the same. In the prodromal stage: $P_d$, the weekday probability, is 0.25 while $P_w$, the weekend probability, is 0.40. Since in the fulminant stage the anthrax symptoms are those of a heart attack, the probability of entering the ED is higher and same for weekdays or weekends, 0.8. We also assumed that a small percentage of the subjects are misdiagnosed and thus need to re-enter the system. We assumed a maximum of three visits to the ED during the same attack, i.e., maximum of two misdiagnoses. We assumed that 10% are misdiagnosed once, and only 5% are misdiagnosed twice. The probabilities of going to the ED for these misdiagnosed cases are larger, 0.05 larger for the second visit and another 0.05 for the third visit. We also include a probability of drop-out in the model. We have removed the phrases regarding enclosed areas, and emphasized that indeed we use the individual-based infection model.

4. We have amended the discussion of the calculation of the specificity. As per your suggestion we now calculate the specificity using all the data, in the absence of the outbreak signature.

Responses to Discretionary Revisions

- We now calculate the (average) ROC curves. We believe this greatly improves the discussion of the detection methods.
Response to Reviewer 2: Simon Hendrik H. Heisterkamp

We thank you for reviewing the paper. Here we include the responses to your review.

Responses to Major Compulsory Revisions

1. An in-depth discussion of our choice of distribution and scale (including why we do not use an offset model is given at the top of Page 5 of the revision). While in the future we seek to further investigate Poisson-based models, our aim at the moment is to understand what is typically done in the literature (e.g., Reis et al., Ivanov et al., Burkem et al., Shmueli, and Zhang, et al. consider forms of Gaussian time series model on the original scale, and detection methods based on filtering). We added a new figure (Figure 3) which contains the time series residual plots. These plots show that, in fact the Gaussian distribution assumption is appropriate for these data. We discuss more about, $\sigma_k^2$, the innovation variance of the time series errors in the revised manuscript.

2. Upon reflection, the strongest seasonal terms are at the weekly scale. We have amended our discussion in the paper on this point. We have down-weighted our conclusions about yearly seasonal variation; instead we motivate the scientific rationale for including temperature as a covariate.

3. Our time series models are indeed intended to capture the variations in the influenza-based background signal, including other outbreaks. What we are investigating is that we can detect an anthrax outbreak above this background (e.g., an anthrax attack), that appears according to a certain stochastic distribution. We have rewritten the Background to make this clearer.

4. The health-utilization model has been improved in the revised manuscript, using ideas from the other reviewer – please see pages 7–8. Now, the probabilities of entering the ED for the prodromal and fulminant stages are not the same. In the prodromal stage: $P_d$, the weekday probability, is 0.25 while $P_w$, the weekend probability, is 0.40. Since in the fulminant stage the anthrax symptoms are those of a heart attack, the probability of entering the ED is higher and same for weekdays or weekends, 0.8. We also assumed that a small percentage of the subjects are misdiagnosed and thus need to re-enter the system. We assumed a maximum of three visits to the ED during the same attack, i.e., maximum of two misdiagnoses. We assumed that 10% are misdiagnosed once, and only 5% are misdiagnosed twice. The probabilities of
going to the ED for these misdiagnosed cases are larger, 0.05 larger for the second visit and another 0.05 for the third visit. We also include a probability of drop-out in the model.

The reviewer is correct, we didn’t use a region-based health-utilization model (like Buckeridge et al.’s) to take into account for spatial correlations. We used, instead, the Brookmeyer individual-based infection model. We included a sentence on page 6 indicating this. Temporal correlations were not considered since anthrax is not contagious, so the only mechanism to get the disease is that individuals inhale the spores.

5. A clarification. Our training model is fit only using the ten months of data. The outbreaks are added to data for the second half (the remaining ten months). Our simulations (at least partly) reflect new data coming into the system. We have amended the text to make this point clearer adding a sentence in the Methods section. Also this is explained in the Simulations section.

Responses to Minor Essential Revisions

1. See above for the response regarding seasonal terms.

2.-3. See above for the discussion on choice of distribution and scale.

4. Thank you for your comment. We included the word "infected" in the sentence, as you suggested. We think that this assumption is realistic because we are only "adding" outbreak data to the observed data. People that have other chest problems are considered in the background data.

5. The \( \{E_{k,t}\} \) process was defined in the Appendix. It is the one step prediction error process, which depends on the data, and the one-step-ahead predictions based on the time series model – it is not the expectations – sorry for the misunderstanding. Including this term did not add any interpretation to the main body of the text, and so we removed it, leaving it only in the Appendix. Instead we added more explanatory discussion of the use of filter-based outbreak detection (see Pages 8 and 9 of the revision).

6. We have included further information about diagnostic checks for the model.