Estimation And Validation Of An Outbreak Simulator

Min Zhang, Ph.D., Xiaohui Kong, Garrick L. Wallstrom, Ph.D.

RODS Laboratory, Department of Biomedical Informatics, University of Pittsburgh

OBJECTIVE

In previous work [1], we developed a Template-Driven Simulator, which is a non-disease specific outbreak simulator that uses templates to describe the temporal or spatial-temporal pattern of an outbreak. Here we address the problem of estimating the template from outbreak data. We then conduct a limited validation of the outbreak simulation model by estimating the template using outbreak data generated from BARD [2], a sophisticated state-of-the-art anthrax outbreak simulator and detector. This limited validation confirms that the outbreak simulator is capable of generating complicated disease outbreak patterns for evaluating outbreak detection algorithms.

BACKGROUND

In [1], we described a non-disease-specific outbreak simulator for the evaluation of outbreak detection algorithms. This Template-Driven Simulator generates disease patterns using user-defined template functions. Estimation of a template function from real outbreak data would enable researchers to repetitively simulate outbreaks that resemble a single real outbreak. These simulated outbreaks can then be used to evaluate outbreak detection algorithms.

To demonstrate template estimation, we employ BARD [2], a disease-specific outbreak model for outdoor aerosol release of *B. anthracis*. It uses epidemiological and atmospheric dispersion models in conjunction with geographical and meteorological data to generate anthrax cases. The home census block group and time of visit to an emergency department are available for each simulated case.

METHODS

The spatial-temporal template is a bounded function *f* of space *s* and time *t* that describes the spatial and temporal distribution of cases. We write the function as a product of a conditional density function of time given the spatial location and a marginal probability mass function over the spatial location:

$$f_{S,T}(s,t) = f_{T|S}(t \mid s)\, f_S(s) \quad (1)$$

For each case, BARD simulates emergency department visit times using a convolution of two lognormal distributions. The lognormal parameters depend on the home location of the case to account for the dose received in that block group. Here we model the visit times in each block group by a single lognormal distribution with location-dependent parameters. That is, $T \mid S = s \sim \text{lognormal}(\mu(s), \sigma(s))$. We assume that μ and σ are smooth functions of space, and f_S is an arbitrary probability mass function.

We estimate $f_S(s)$ by the proportion of all cases that reside in block group *s*. We estimate the smooth functions μ and σ by first computing the maximum likelihood estimates of μ and σ for each block group, and then computing a spatially-weighted average of the maximum likelihood estimates.

We use data from ten BARD simulations of an aerosol release of 0.1kg of anthrax spores in the Pittsburgh region. For each set of data, we compute a Pearson χ^2 goodness-of-fit test statistic using block groups and days for bins. Due to some low bin counts, we use Monte Carlo simulation rather than a χ^2 approximation to compute p-values.

RESULTS

The p-values for the ten datasets are listed in Table 1.

Table 1. P-values for the ten datasets

P-value (dataset 1-5)	.053	.007	.299	.390	.070
P-value (dataset 6-10)	.109	.812	.997	.407	.858

CONCLUSIONS

The goodness-of-fit tests performed on ten sets of data indicate mixed results. While many of the p-values are large indicating good model fit, there are several p-values that are low. This result suggests that goodness-of-fit may depend on characteristics of the release such as the weather conditions at the time of the release. A larger simulation study and further analysis would be required to fully understand these mixed results.

The majority of the simulated data sets do result in good model fit. These results suggest that the simulator is sufficiently flexible to describe most (but not all) simulated releases from BARD. This result provides a limited validation of the simulation model. Further model validation should include estimation from real outbreak data.

ACKNOWLEDGMENTS

This research was supported by a grant from the Centers for Disease Control and Prevention (R01PH000025). This work is solely the responsibility of its authors and do not necessarily represent the views of the CDC. We thank Dr. William Hogan for providing the BARD simulation data.

REFERENCES

[1] Zhang M, Wallstrom GL, Template-Driven spatial-temporal outbreak simulation for outbreak detection evaluation, AMIA 2008 annual symposium, Washington D.C, Nov. 2008. In press.

[2] Hogan WR, Cooper GC, Wallstrom GL, Wagner MM, Depinay J-M. The Bayesian aerosol release detector. Stat Med 2007. 26(29): 5225-52.