Welcome to the Webinar!

INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Special Literature Review: 2014 Awards for Outstanding Research Articles in Biosurveillance Finalists

Presenters:

Sylvia Halasz, PhD, MSc, YP.com
Judy Akkina, MPH, PhD, USDA, APHIS
Howard Burkom, PhD, JHU/APL
Amy Ising, MS, University of North Carolina at Chapel Hill
Mike Conway, PhD, University of California, San Diego
Céline Dupuy, Anses, Laboratoire-Lyon
Katie Suday, PharmD, MS, Department of Veterans Affairs
Al Ozonoff, PhD, Boston Children's Hospital

Please use #ISDSlitreview in all webinar- and Conference-related Tweets

Webinar Logistics

- All participants are in listen-only mode
- Webinar is being recorded – Slides and recording will be available shortly following the webinar
- Questions using GoToWebinar Control Panel:
  - Type your questions into the ‘Question’ portion
  - Electronically ‘raise your hand’
- May also ask questions via Twitter using #ISDSlitreview

Award Committee

- Lead: Al Ozonoff, PhD, Boston Children's Hospital
- Alicia Livinski, MA, MPH, National Institutes of Health
- Amy Ising, MS, University of North Carolina at Chapel Hill
- Bill Storm, MPH, Ohio Department of Health
- Céline Dupuy, Anses, Laboratoire-Lyon
- Howard Burkom, PhD, Johns Hopkins University Applied Physics Laboratory
- Judy Akkina, MPH, PhD, USDA, APHIS
- Katie Suda, PharmD, MS, Dept. of Veterans Affairs
- Mike Conway, PhD, University of California, San Diego
- Sylvia Halasz, PhD, MSc, YP.com

Brief Overview of Process

- Articles Screened: ~700 articles
- Initial Pool: 84
- Nominated: 69
- Finalists: 10

Judging Criteria

INTERNATIONAL SOCIETY
FOR DISEASE SURVEILLANCE

- Impact on field of biosurveillance
  - Potential to change the way that surveillance is practiced
  - Potential to attract scientists from other disciplines
  - Educational value to a broader audience
- Scientific achievement
  - Innovation of methodology or application
  - Rigor of methods
  - Quality of scientific writing

Comparing control strategies against foot-and-mouth disease: Will vaccination be cost-effective in Denmark?

A Boklund, T Halasa, LE Christiansen, C. Enøe

Presenter: Sylvia Halasz, PhD, MSc, YP.com

Twitter: #ISDSlitreview

Brief Background/Purpose

Objective of the study:

1. Simulate outbreaks of foot-and-mouth disease (FMD) in Denmark using two simulation tools.
2. Do sensitivity analysis within the basic intervention scenarios:
    - Depopulation of detected herds;
    - Depopulation in ring zones around the detected herds;
    - Depopulation + suppressive vaccination (VacToKill);
    - Depopulation + protective vaccination (VacToLive).
3. Cost-benefit analysis and strategy recommendations.

Scientific achievement: Innovation of methodology or application

The study uses simulation results as a basis for cost-benefit analysis of possible intervention scenarios.

There are very few studies combining simulation and cost-benefit analysis.

Scientific achievement: Rigor of methods

1. Real data and realistic assumptions were used in the cost-benefit analysis.
2. The sensitivity analysis is statistically sound.
3. Variables that could not be modeled in the tools are pointed out.
4. The conclusions go into detail not only about the recommended strategies, but also about the importance of different variables (some of which are modeled, others are not).

Scientific achievement: Quality of scientific writing

The presentation is clear;

the different intervention strategies meticulously defined;

the assumptions about disease spread, lost income, etc. are clearly stated.

Scientific Achievement Category

INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation. J R Soc Interface. 2013;10(83).

FC Dórea, BJ McEwen, WB McNab, CW Revie, J Sanchez

Presenter: Judy Akkina, MPH, PhD, U.S. Department of Agriculture, APHIS

Twitter: #ISDSlitreview

Scientific achievement: Innovation of methodology or application

- Innovative in applying the methods to veterinary diagnostic laboratory data and for the purpose of developing syndromic surveillance systems to animal health

Scientific achievement: Rigor of methods

INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

- Used both simulated and real data and multiple syndromes to evaluate performance of the algorithms
- Evaluated 2 pre-processing methods
- Evaluated 4 algorithms on 5 different outbreak signal shapes
- In addition to standard sensitivity and specificity measures used ROC curves to determine sensitivity
- Complex study with many phases

Scientific achievement: Quality of scientific writing

- Extremely well written and understandable, even when describing complicated statistical methods and other aspects of the complex study
- Excellent figures illustrating what is described in the text
- Very educational related to using algorithms, different algorithm methods, and how to evaluate them

One-Health Simulation Modelling: A Case Study of Influenza Spread between Human and Swine Populations using NAADSM

S Dorjee, CW Revie, Z Poljak, WB McNab, J Sanchez

Presenter: Sylvia Halasz, PhD, MSc, YP.com

Twitter: #ISDSlitreview

Scientific achievement: Innovation of methodology or application

Objectives of the study:

1. Investigate the feasibility of using the North American Animal Disease Spread Model (NAADSM) as a tool to model the spread and control of zoonotic diseases;
2. Study the transmission dynamics of influenza at the swine-human interface using pH1N1 as an example;
3. Investigate the utility of applying targeted vaccination against influenza at the animal-human interface.

While NAADSM was developed to model disease spread between different livestock units, this study also includes households of people as a type of "herd".

Scientific achievement: Rigor of methods

Table 1. Description of study populations and probability density functions of the size of units used for the simulation of influenza spread between swine and human populations in a county of Ontario, Canada

Population units	**Total no. of units**	**Distribution of size of units**	**Total no. of individuals**
Swine herds (SH)	488	Uniform (min = 500; max = 2500)	733 107
Swine-worker households (SWH)	733	Truncated Poisson (mean = 3, min = 2; max = 7)	2325
Rural non-swine worker-Households (RH)	7879	Truncated Poisson (mean = 3, min = 1; max = 6)	25 521
Urban households(UH)	21 095	Truncated Poisson (mean = 3, min = 1; max = 6)	54 038
Total	30 195		814 991

Scientific achievement: Quality of scientific writing

- The presentation is clear;
- The assumptions are rigorously specified;
- The results are presented first in a statistical frame of reference;
- Then the conclusions are summarized from a public health point of view.

Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data

T Jombart, A Cori, X Didelot, S Cauchemez, C Fraser, N Ferguson

Presenter: Howard Burkom, PhD, Johns Hopkins University Applied Physics Laboratory

Twitter: #ISDSlitreview

Title, Author(s) of article: Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data.
Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N.
PLoS Comput Biol. 2014;10(1):e1003457. doi:10.1371/journal.pcbi.1003457.
http://dx.doi.org/10.1371/journal.pcbi.1003457

WIKI Research Area: 2) Statistics & Analysis

WIKI subcategory: Disease Modeling and Applications

Brief description: The authors use pathogen genome sequences and collection dates for specimens from a given outbreak to find best estimates of key descriptors of disease transmission, including effective reproduction number, newly introduced cases, and ranges of infectivity including the identification of superspreaders. Key assumptions are 1) that the infected populations are densely sampled (at least 25% of infected), 2) distribution of the generation time is known, 3) distribution of the time from infection to sample collection is known, 4) each host is limited to a single pathogen genome. The reconstruction may be improved with specific information such as known introduced cases. Software for the reconstruction method and for data simulation is available through the R package *outbreaker*.

Methods: Reconstruction of likely transmission trees is achieved with a stochastic Bayesian model implemented with MCMC sampling.

Datasets: Effectiveness of the reconstruction is tested on simulated sets of genome sequences. The method is also tested on genome data collected in Singapore from the 2003 SARS epidemic.

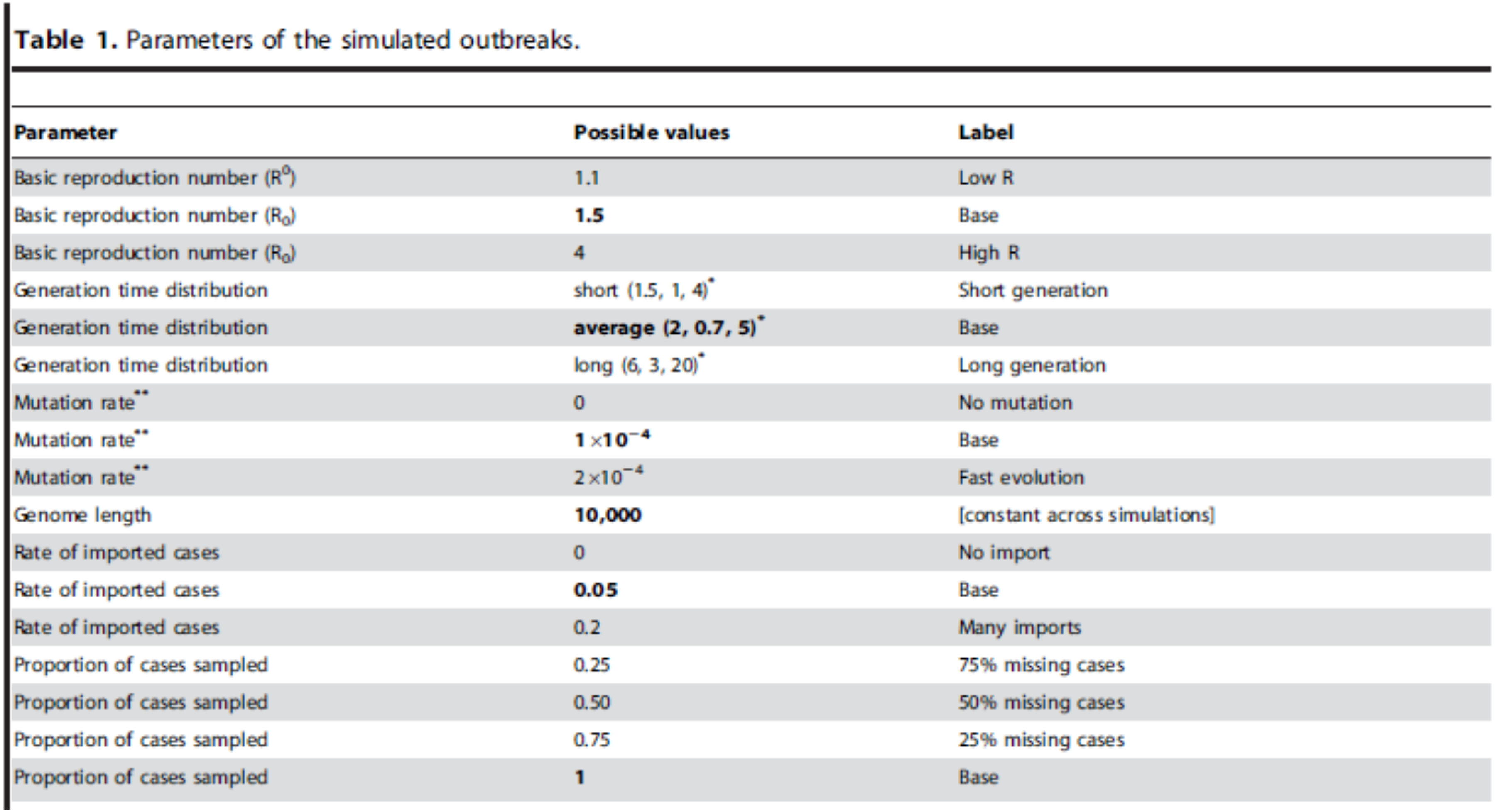


Table 1. Parameters of the simulated outbreaks.

Parameter	**Possible values**	**Label**
Basic reproduction number (R^0)	1.1	Low R
Basic reproduction number (R_0)	**1.5**	Base
Basic reproduction number (R_0)	4	High R
Generation time distribution	short (1.5, 1, 4)*	Short generation
Generation time distribution	**average (2, 0.7, 5)***	Base
Generation time distribution	long (6, 3, 20)*	Long generation
Mutation rate**	0	No mutation
Mutation rate**	**1×10^{-4}**	Base
Mutation rate**	2×10^{-4}	Fast evolution
Genome length	**10,000**	[constant across simulations]
Rate of imported cases	0	No import
Rate of imported cases	**0.05**	Base
Rate of imported cases	0.2	Many imports
Proportion of cases sampled	0.25	75% missing cases
Proportion of cases sampled	0.50	50% missing cases
Proportion of cases sampled	0.75	25% missing cases
Proportion of cases sampled	**1**	Base

Posterior distribution and full likelihood. Our model is embedded within a Bayesian framework. We denote Υ the observed data, A the augmented data, and θ the model parameters. The joint posterior distribution of parameters and augmented data is defined as:

$$P(A,\theta|D) = \frac{P(D,A|\theta)P(\theta)}{P(D)} \tag{1}$$

which is proportional to:

$$\mathrm{p}(\{s_i,t_i,\alpha_i,\kappa_i,T_i^{\mathrm{inf}}\}_{(i=1,\ldots,N)}|\mu,\pi)\times \mathrm{p}(\mu,\pi) \tag{2}$$

where the first term is the likelihood of the data and augmented data, and the second, the joint prior distribution. Likelihood computations are described below. Priors and estimation procedures are described in Supporting Methods.

Simulation of disease outbreaks

Model. Outbreaks were simulated using the function *simOutbreak* in the package *outbreaker*. Each simulation starts with a single infection in a population of n susceptible hosts. For simplicity, the same function was used for w and f. R_0 is the fixed basic reproduction number, and S_t the number of susceptible hosts at time t. The probability for a susceptible individual to become infected on day t is:

$$p_t^{\mathrm{inf}} = 1 - e^{-\sum_i R_0 w(t - t_i)/n} \tag{9}$$

At each time step, the number of new cases is drawn from a binomial distribution with S_t draws and a probability p_t^{inf}. Infectors of a case infected at time t_i are sampled from a multinomial distribution with probabilities:

$$\frac{w(t - t_i)}{\sum_i w(t - t_i)} \tag{10}$$

In addition to endogenous cases, external cases are imported at a constant rate.

Bayesian Reconstruction of Disease Outbreaks

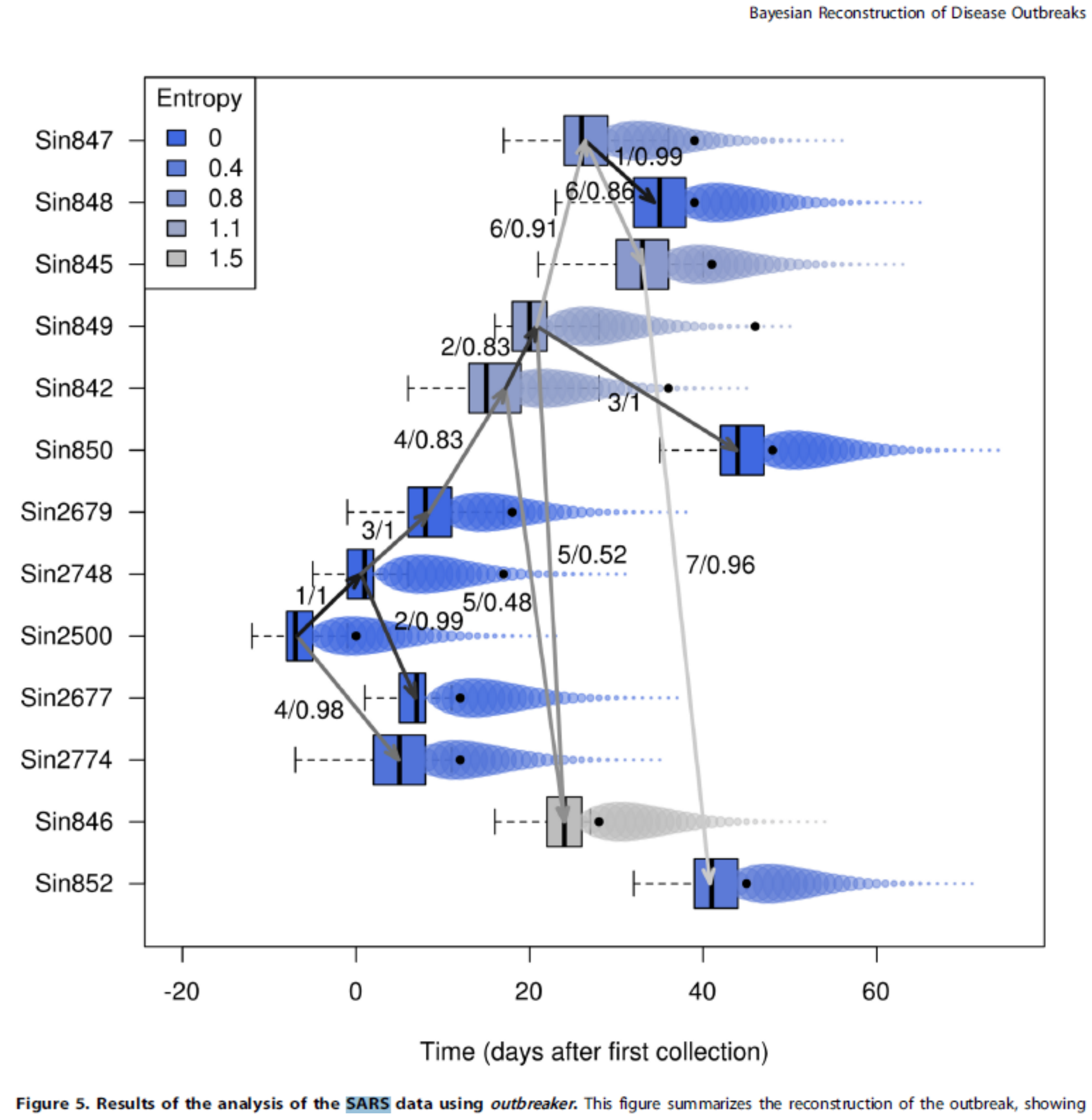
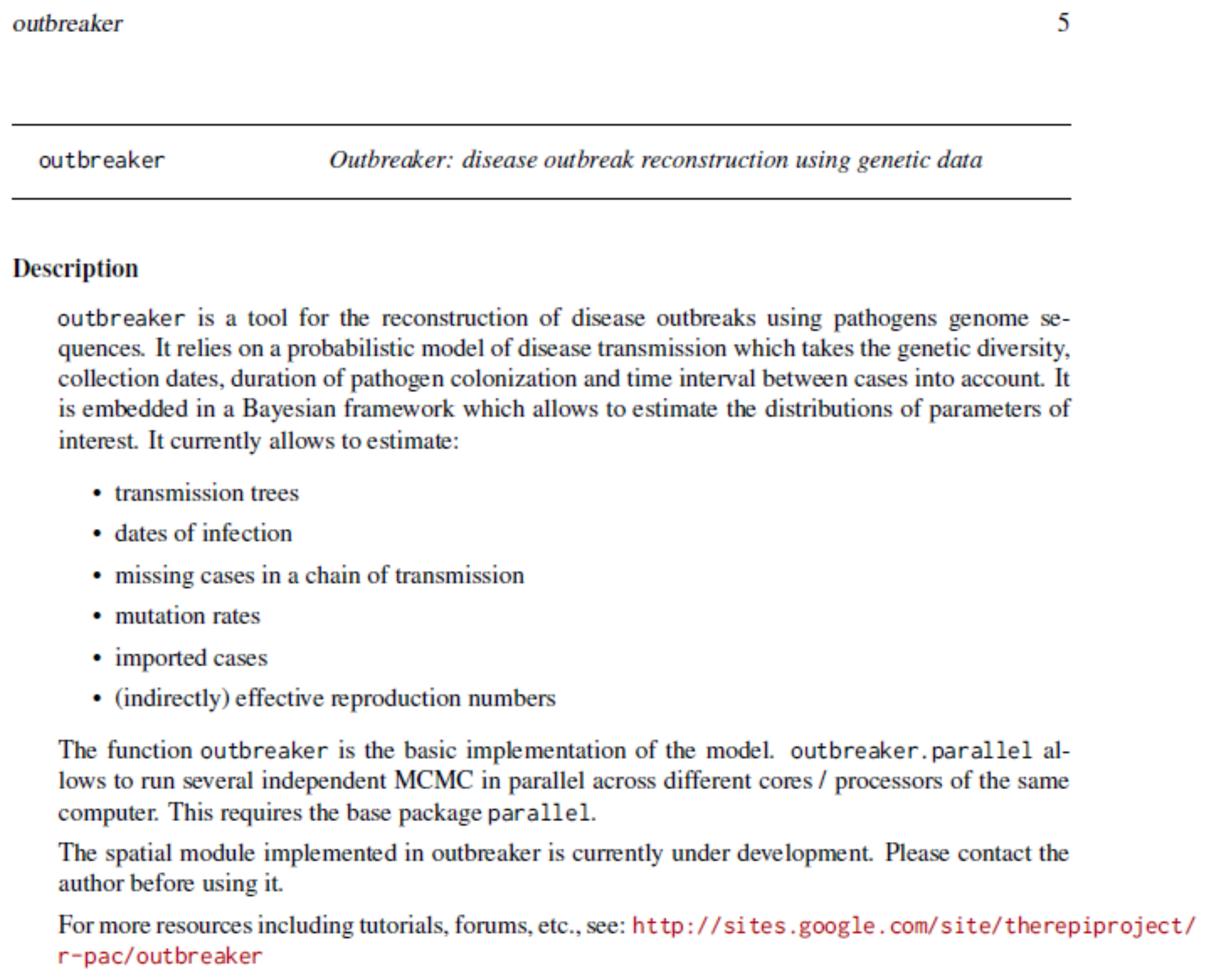


Figure 5. Results of the analysis of the SARS data using *outbreaker*. This figure summarizes the reconstruction of the outbreak, showing



`outbreaker` *Outbreaker: disease outbreak reconstruction using genetic data*

Description

`outbreaker` is a tool for the reconstruction of disease outbreaks using pathogens genome sequences. It relies on a probabilistic model of disease transmission which takes the genetic diversity, collection dates, duration of pathogen colonization and time interval between cases into account. It is embedded in a Bayesian framework which allows to estimate the distributions of parameters of interest. It currently allows to estimate:

- transmission trees
- dates of infection
- missing cases in a chain of transmission
- mutation rates
- imported cases
- (indirectly) effective reproduction numbers

The function `outbreaker` is the basic implementation of the model. `outbreaker.parallel` allows to run several independent MCMC in parallel across different cores / processors of the same computer. This requires the base package `parallel`.

The spatial module implemented in outbreaker is currently under development. Please contact the author before using it.

For more resources including tutorials, forums, etc., see: `http://sites.google.com/site/therepiproject/r-pac/outbreaker`

Scientific Achievement Category

INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products

M Myslin, S-H Zhu, W Chapman, M Conway

Presenter: Amy Ising, MS, University of North Carolina at Chapel Hill

Twitter: #ISDSlitreview

Brief Background/Purpose

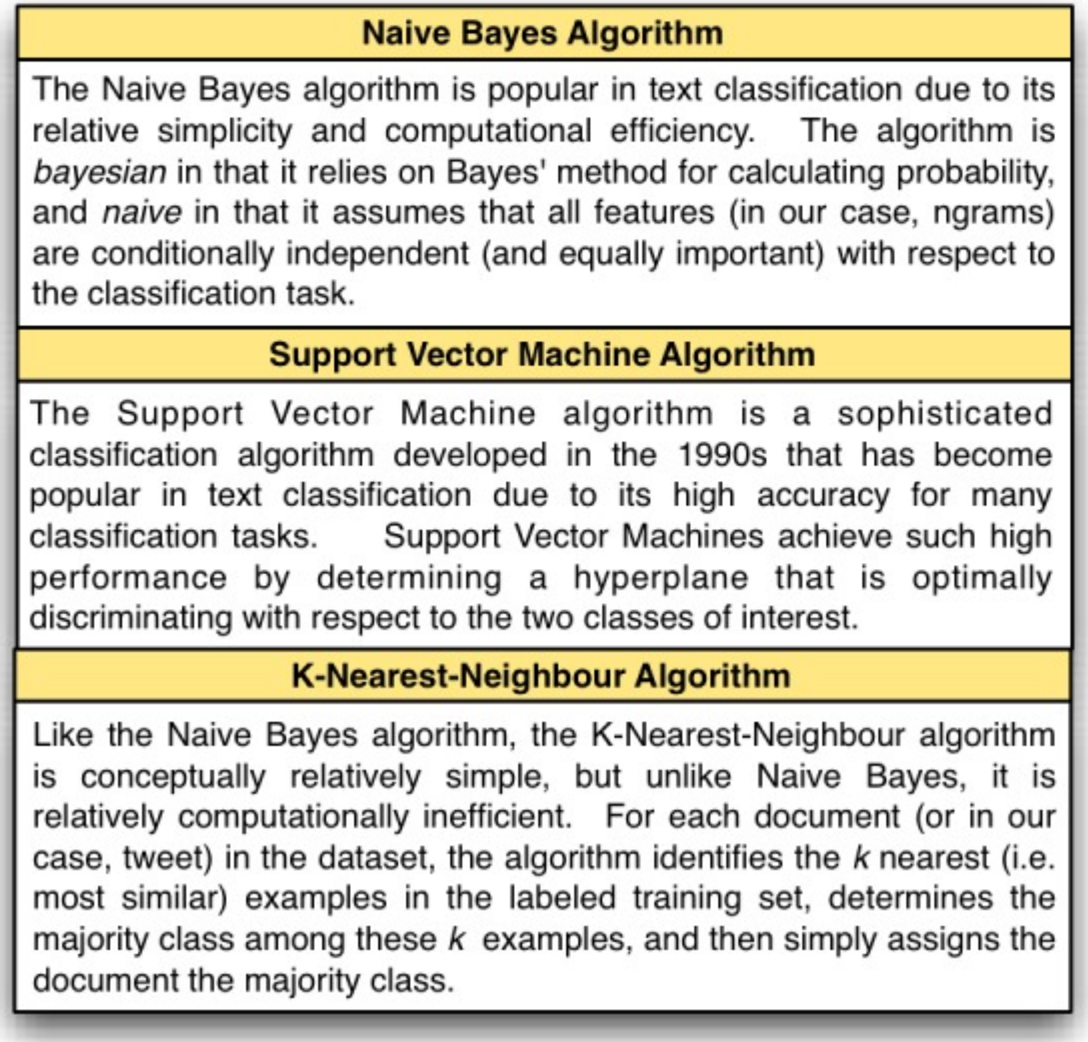
- Using machine learning techniques to analyze tweets to classify posts automatically according to their relevance to tobacco-related content categories.

Scientific achievement: Innovation of methodology or application

- improve the signal-to-noise ratio in Twitter data by automatically filtering out irrelevant content / comparison of 3 machine learning algorithms
- Focus on tobacco (vs. other topics typically analyzed using Twitter data)

Scientific achievement: Rigor of methods

- Comparison of three different machine-learning algorithms



Naive Bayes Algorithm

The Naive Bayes algorithm is popular in text classification due to its relative simplicity and computational efficiency. The algorithm is *bayesian* in that it relies on Bayes' method for calculating probability, and *naive* in that it assumes that all features (in our case, ngrams) are conditionally independent (and equally important) with respect to the classification task.

Support Vector Machine Algorithm

The Support Vector Machine algorithm is a sophisticated classification algorithm developed in the 1990s that has become popular in text classification due to its high accuracy for many classification tasks. Support Vector Machines achieve such high performance by determining a hyperplane that is optimally discriminating with respect to the two classes of interest.

K-Nearest-Neighbour Algorithm

Like the Naive Bayes algorithm, the K-Nearest-Neighbour algorithm is conceptually relatively simple, but unlike Naive Bayes, it is relatively computationally inefficient. For each document (or in our case, tweet) in the dataset, the algorithm identifies the *k* nearest (i.e. most similar) examples in the labeled training set, determines the majority class among these *k* examples, and then simply assigns the document the majority class.

Scientific achievement: Quality of scientific writing

- Well-written, accessible

Impact on Field of Biosurveillance Category

Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales

T Andersson, P Bjelkmar, A Hulth, J Lindh, S Stenmark, M Widerstrom

Presenter: Mike Conway, PhD, University of California, San Diego

Twitter: #ISDSlitreview

Brief Background/Purpose

- Comparing data sources for syndromic surveillance
- Three approaches:
  - Telephone triage service
  - OTC pharmacy sales
  - Web queries
- Compared against 9 biggest GI-illness outbreaks in Sweden between 2007 and 2011
- Summary of findings:
  - Telephone triage service identified 4 outbreaks
  - OTC pharmacy sales identified 2 outbreask
  - Web query data identified 0 outbreaks

Impact on field: Potential to change the way that surveillance is practiced

- Clear advantage of telephone triage data compared to over data sources.
- Exemplar for how surveillance can be practiced in smaller developed countries with homogeneous National Health-type systems

Impact on field: Educational value to a broader audience

- Clearly outlines 3 practical surveillance methods

Impact on Field of Biosurveillance Category

Bayesian outbreak detection algorithm for monitoring reported cases of campylobacteriosis in Germany

J Manitz, M Höhle

Presenter: Céline Dupuy, Anses, Laboratoire-Lyon

Twitter: #ISDSlitreview

Brief Background/Purpose

- Methods for outbreaks detection are mainly of frequentist nature
- Difficulties to take into account uncertainty in prediction and estimation
- Investigation of a Bayesian approach: comparison with existing methods through simulations based on the example of notified campylobacteriosis cases in Germany (2002-2011)
- Reasonable computation time

Impact on field: Potential to change the way that surveillance is practiced

- A new methodology that could be investigated on other diseases: Bayesian outbreak detection algorithm (BODA)
- Bayesian hierarchical model algorithm
- Take into account uncertainty in prediction and estimation
- Possible implementation using R software

Impact on field: Potential to attract scientists from other disciplines

- Could be applied on syndromic surveillance

Impact on field: Educational value to a broader audience

- New approach for time series analysis

Impact on Field of Biosurveillance Category

Reassessing Google flu trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales.

DR Olson, et al.

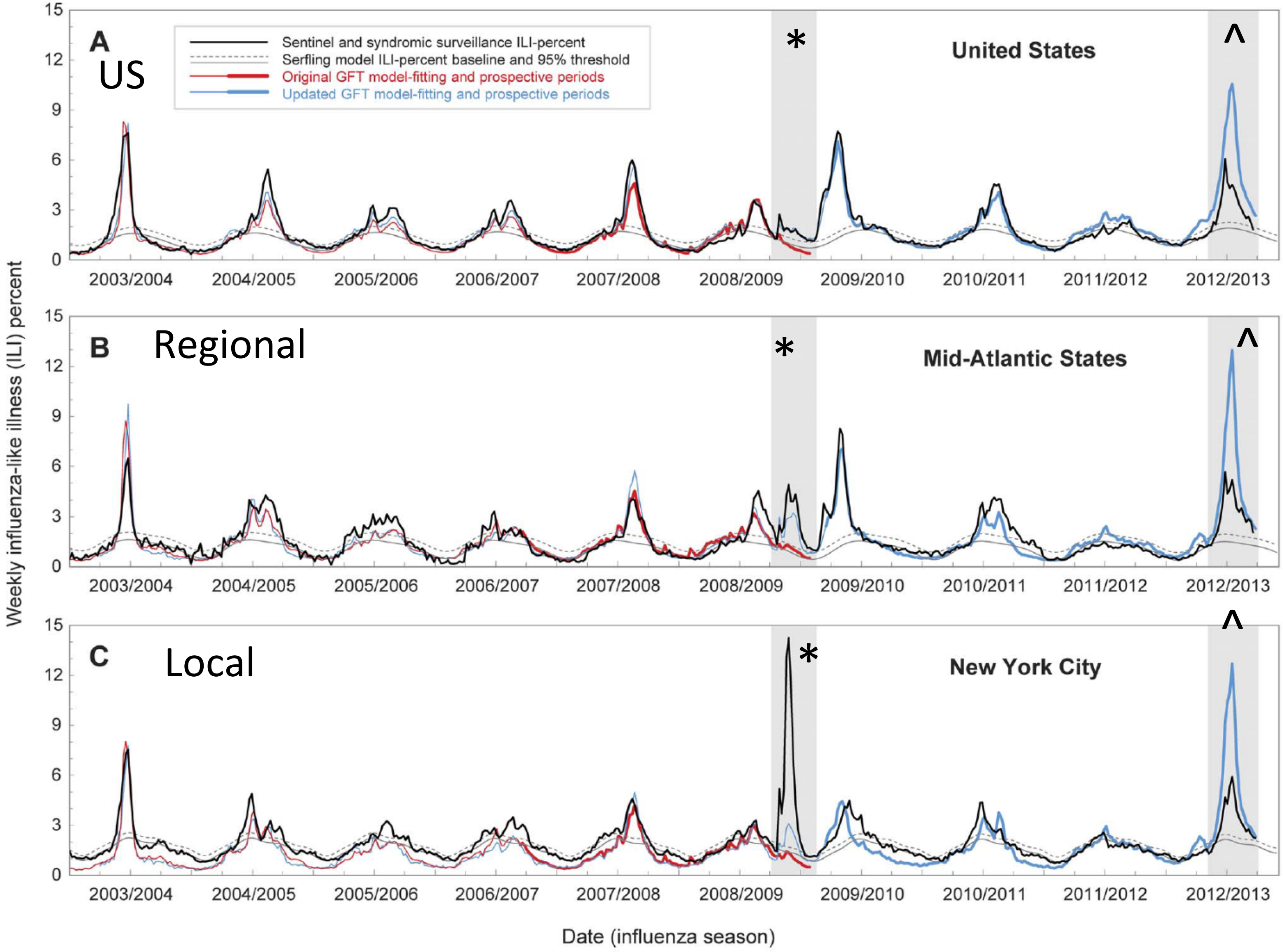
Presenter: Katie J. Suda, PharmD, M.S., Department of Veterans Affairs

Twitter: #ISDSlitreview

Brief Background/Purpose

- Google flu trends (GFT) model was revised in the Fall 2009
    - 2009 H1N1 pandemic was not detect by original GFT system
    - Model was updated with less restrictive criteria and additional search terms
- Purpose: Compare traditional ILI public health surveillance and GFT estimates for 10 influenza seasons (6/1/2003-3/30/2013) at three geographic levels (US, mid-atlantic region, NYC)

Results



* Original GFT underestimated H1N1

^ Updated GFT exaggerated

Impact on field: Potential to change the way that surveillance is practiced

- Public health professionals use Google flu trends
    - May reflect national estimates better than local estimates

Impact on field: Potential to attract scientists from other disciplines

- Additional work needs to be done with internet data to predict public health events
    - Local level
    - Atypical times
    - Measure intensity
    - Development of networks
    - Improve timeliness
    - Developing countries
- Additional support needed for computerized local disease surveillance systems

Impact on field: Educational value to a broader audience

- Other surveillance methods may be better for measuring intensity
- As compared to GFT, local public health surveillance systems are more accurate and as timely

Impact on Field of Biosurveillance Category

INTERNATIONAL SOCIETY
FOR DISEASE SURVEILLANCE

Using quantitative disease dynamics as a tool for guiding response to avian influenza in poultry in the United States of America

KM Pepin, E Spackman, JD Brown, et al.

Presenter: Howard Burkom, PhD, JHU/APL

Twitter: #ISDSlitreview

Brief Background/Purpose

- This is a review article on the use of quantitative studies, including disease dynamic models, to inform policymaking and response in the USA for avian influenza in poultry.
- The authors looked at current control policies, quantitative population data available, surveillance and experimental data, available models, and identified quantitative research priority areas for spillover to poultry from wildlife populations, transmission within poultry flocks, and and between poultry flocks.
- Through their review they identified priorities for future quanitative research which would utilize currently available tools and data to help address current knowledge gaps.

Impact on field: Potential to change the way that surveillance is practiced

- A key areas of improvement is the need for better data collection and type of data collected during an outbreak to better inform response and preparedness in the future. This may be accomplished through data standardization for data collection during outbreaks.

Impact on field: Potential to attract scientists from other disciplines

- Policy
- Veterinary surveillance
- Disease modeling

Impact on field: Educational value to a broader audience

- If new to the field of avian influenza or one needs a refresher, this is a very well written and researched overview of current data, policies, gaps and recommend priorities to inform decision making and planning.

Impact on Field of Biosurveillance Category

INTERNATIONAL SOCIETY
FOR DISEASE SURVEILLANCE

A Decision Support Tool to Compare Waterborne and Foodborne Infection and/or Illness Risks Associated with Climate Change. Using quantitative disease dynamics as a tool for guiding response to avian influenza in poultry in the United States of America

Schijven J, Bouwknegt M, de Roda Husman AM, et al.

Presenter: Al Ozonoff, PhD, Boston Children's Hospital

Twitter: #ISDSlitreview

Brief Background/Purpose

- Climate change will change risk of infection from waterborne or foodborne pathogens
- Policy makers and planners should have best available information regarding future risk of disease
- Known relationships between temperature, precipitation, and disease risk are encoded into a decision support tool to inform planners

Impact on field: Potential to change the way that surveillance is practiced

- Continued focus on effects of climate change on public health
- Decision support tool for policy makers and planners which may influence surveillance practices

Impact on field: Potential to attract scientists from other disciplines

- Climate change is a timely, active, and interdisciplinary topic
- Risk management, land use, public health policy, watershed maintenance, and animal health are potentially related topics

Impact on field: Educational value to a broader audience

- Nice summary and review of key relationships between climate and disease

- Evaluation and thoughtful sensitivity analysis

Upcoming ISDS Events

- ISDS Members: Deadline to vote for the 2014 Awards for Outstanding Research Articles
  - – August 14th (11:59pm EDT)
- 2014 ISDS Conference: Abstract Submission and Award for Outstanding Student Abstract Application Deadlines
  - – September 3rd (11:59pm EDT)
- Rick Heffernan Award for Public Health Practice: Nomination Deadline
  - – September 5th (11:59pm EDT)
- Meaningful Use Community Call
  - – September 5th (1:00 – 2:00pm EDT)
  - – Contact syndromic@syndromic.org for more information.