in silico Surveillance: Informing Surveillance with Simulation

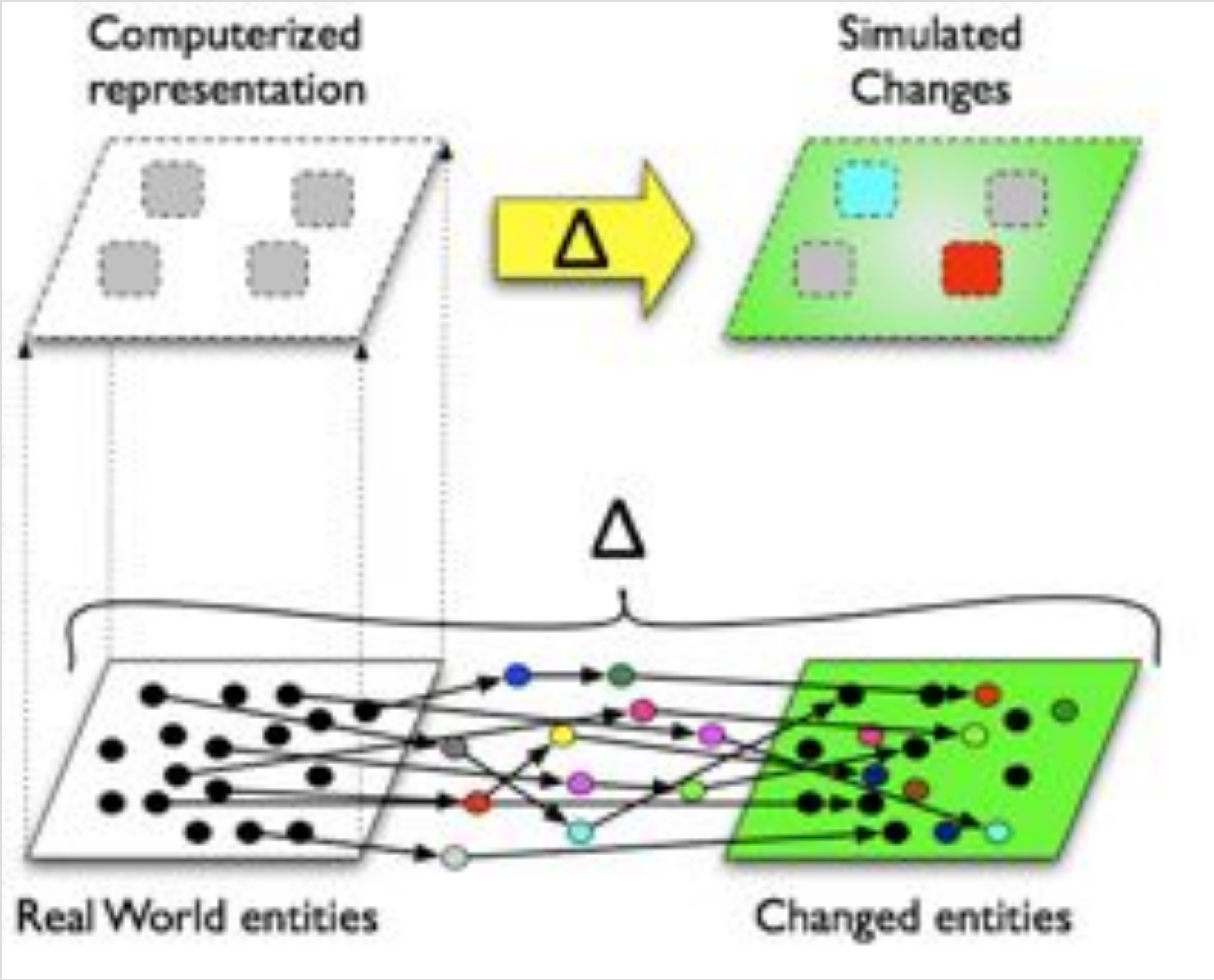
Bryan Lewis, MPH PhD
blewis@vbi.vt.edu
Network Dynamics and Simulation Science Laboratory
Virginia Bioinformatics Institute
Virginia Tech

VirginiaTech
Virginia Bioinformatics Institute
Network Dynamics & Simulation Science Laboratory

Intro to Modeling and Simulation

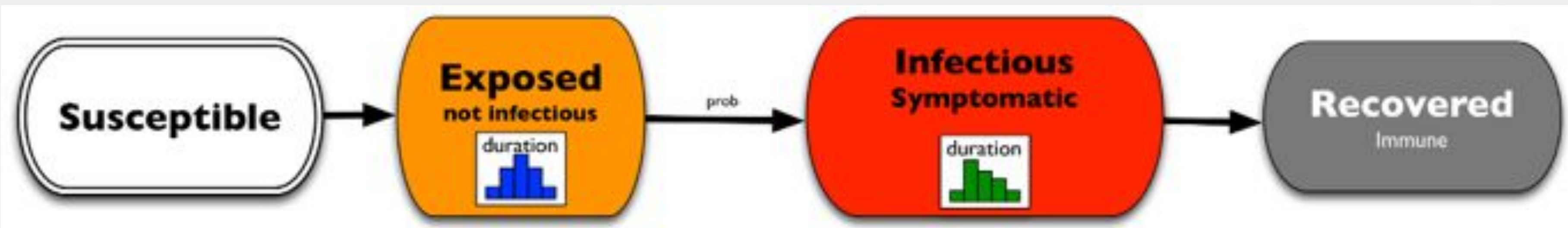
- **Goal:** Make a simplified representation of a system so one can explore its dynamics
- **Challenges:**
  - The world is rarely simple
  - Well understood systems frequently don't need an *in silico* representation to understand its dynamics
- **Motivation:**
  - Sometimes pure analysis isn't revealing enough
  - Model building provides a structure for laying out all assumptions and known information
  - Even if not 100% "right" knowing the general direction of a trend line given certain conditions is valuable

Compartmental Representation



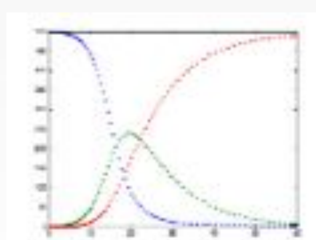
Compartmental Modeling

- For most infectious diseases: SEIR is the basic starting point (SIS, SIR, SEIRS, and other elaborations)

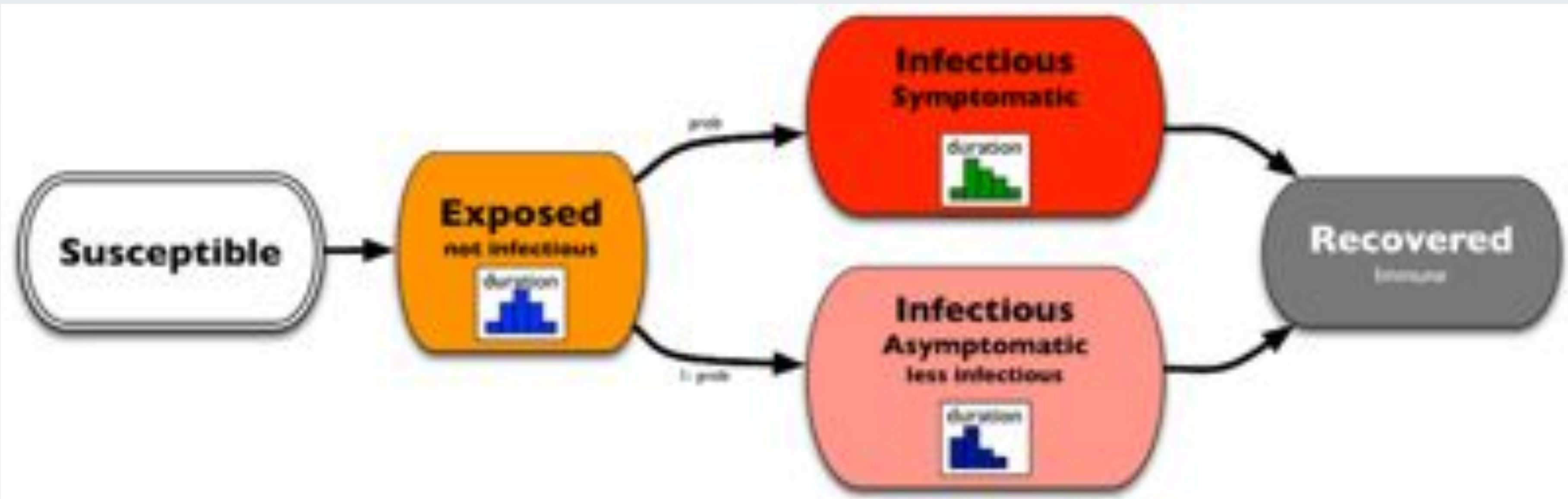


- Compartments for each "important" stage of disease
- Mathematical representation for how the population flows through these compartments

$$\frac{dS}{dt} = \mu N - \mu S - \beta \frac{I}{N} S$$
$$\frac{dE}{dt} = \beta \frac{I}{N} S - (\mu + a) E$$
$$\frac{dI}{dt} = aE - (\nu + \mu) I$$
$$\frac{dR}{dt} = \nu I - \mu R.$$

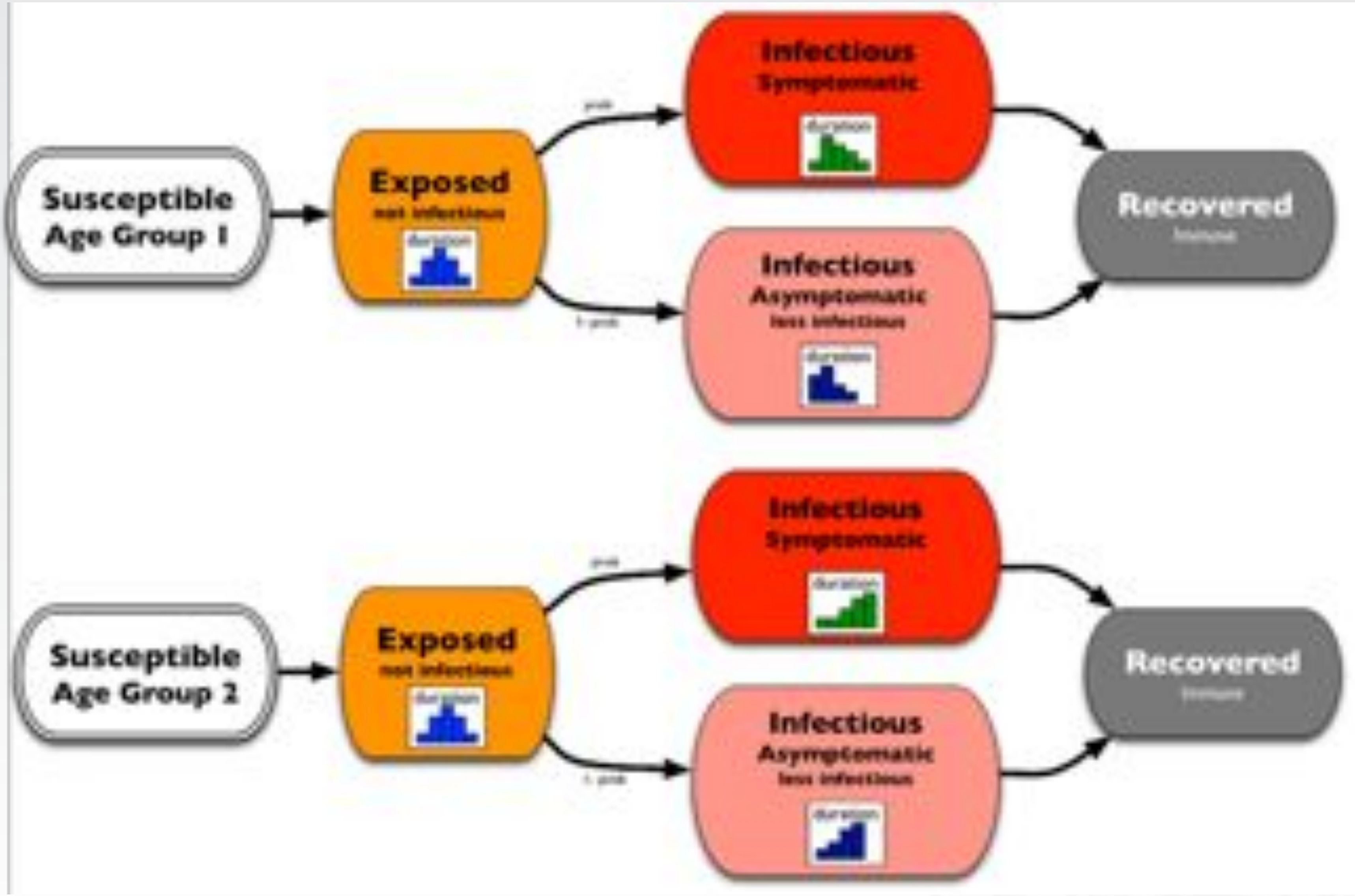


Compartmental Modeling



- However, some diseases need further elaboration

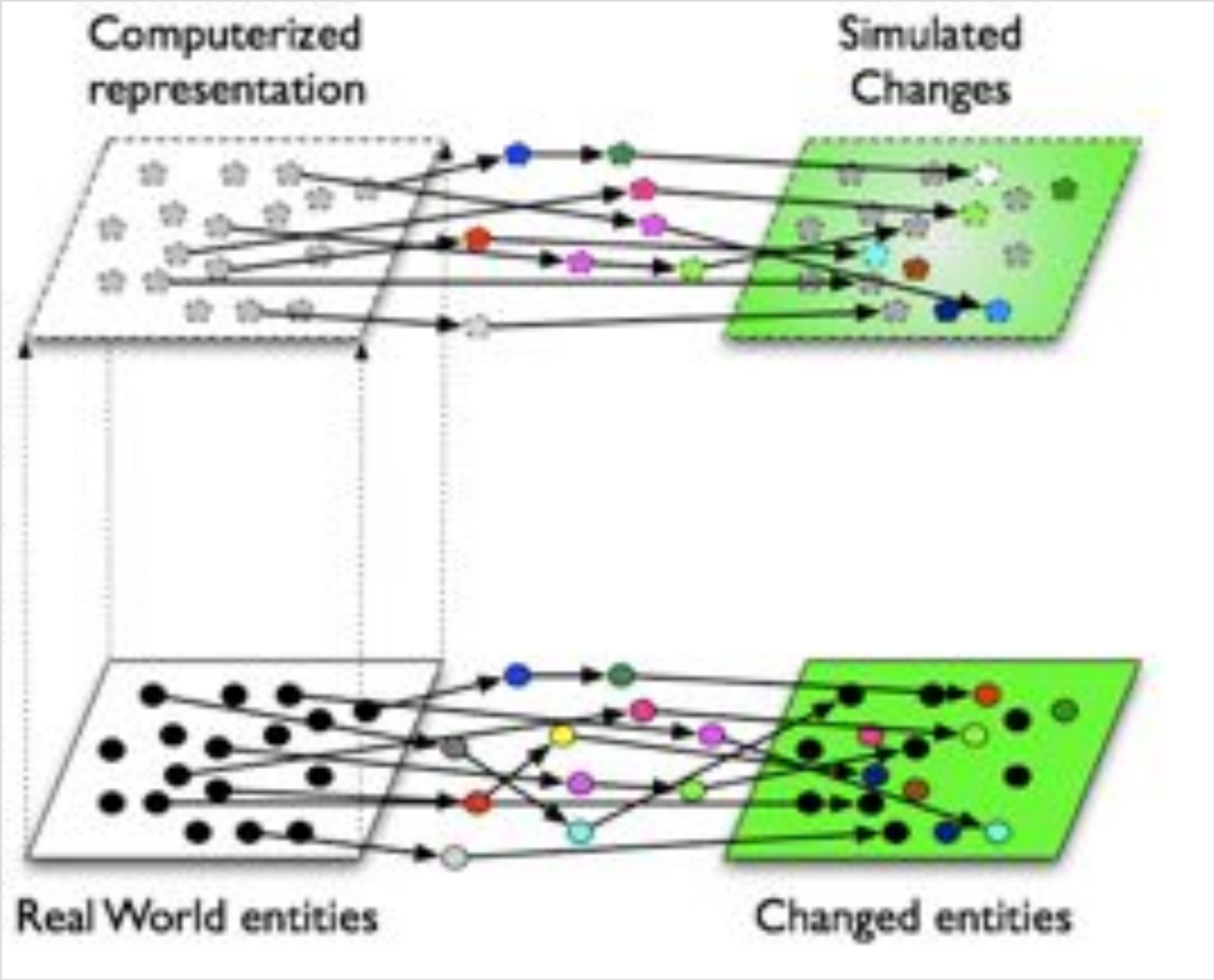
Compartmental Modeling



Compartmental Modeling

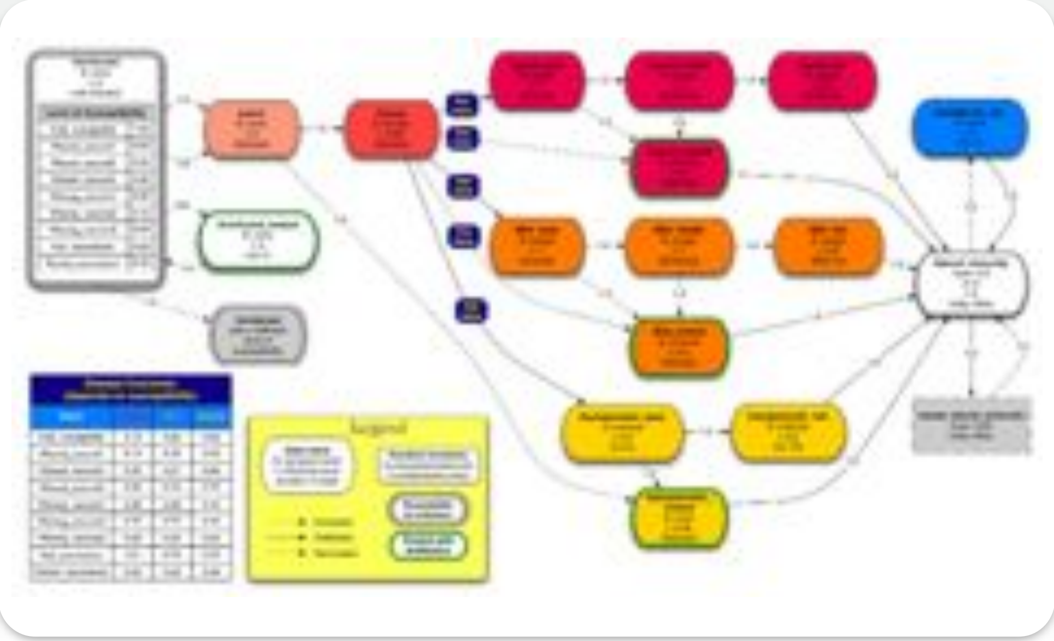
- **Advantages:**
  - While the number of compartments remain reasonable nice analytic methods exist for evaluating the system's behavior
  - Computational requirements are generally reasonable (run on a laptop in less than a day, often minutes/seconds)
  - Fewer parameters require less data and calibration
- **Limitations:**
  - Some questions / scenarios require elaboration beyond "reasonable" levels of compartments
  - Level of abstraction of the represented system limits the ability to generalize to specific cases
  - Data available don't often match up with the parameters

Agent-Based Representation



Agent-Based Modeling

- **Population:** Represented at individual level (can include fomites, vectors, environment)
- **Dynamics:** Most often are stochastic simulations with discrete time, can only implemented as software
- **Disease:** Compartments limited only by data and creativity

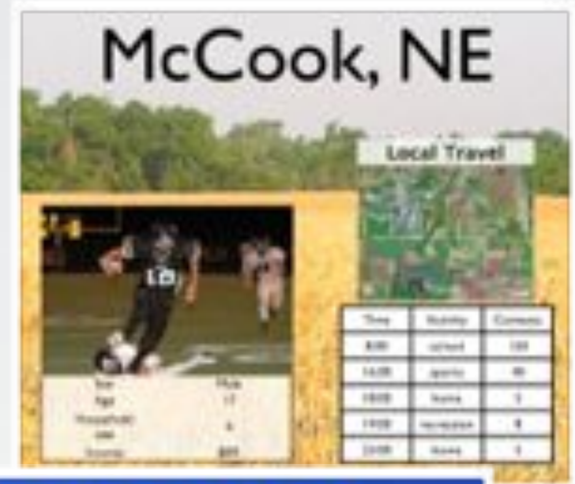


Agent-Based Modeling

- **Advantages:**
  - More structure allows the modeled system to behave more similarly to the represented system
  - Allows a more "natural" representation of the system facilitating elaboration as data often align with parameters
  - Natural representation enables analysis using tools used for analyzing real systems which also facilitates interpretation
- **Limitations:**
  - Computational requirements often require use of high-performance computing resources which aren't always available (though are becoming cheaper and easier to access)
  - More detail requires more data and more calibration
  - Complexity of system can make analysis and interpretation challenging

Building a Synthetic Population: People

- Census data to the census-block level
- Demographics:
  - Gender, Age, Household structure



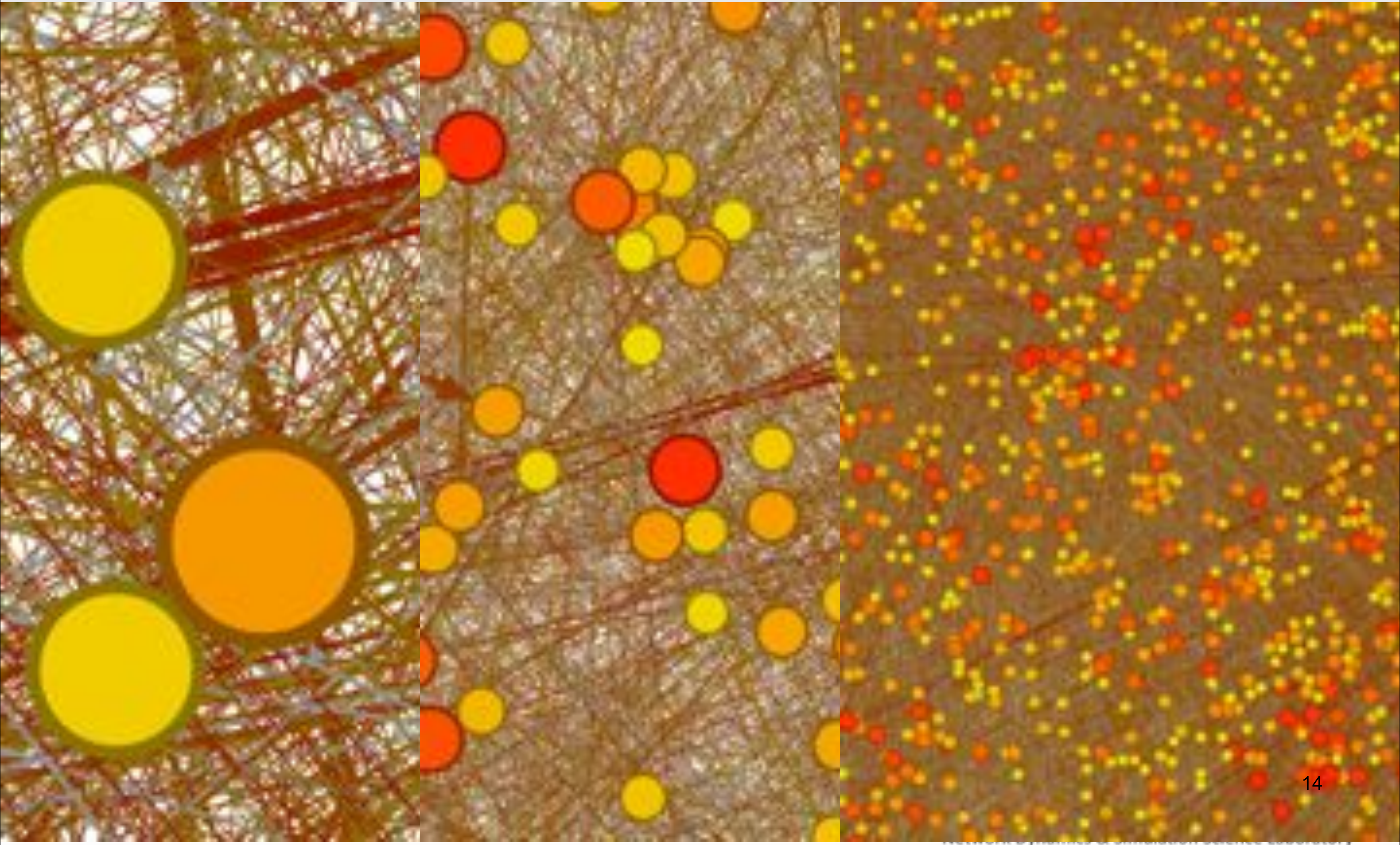
Building a Synthetic Population: Locations

- Dun & Bradstreet and NavTeq
  - Includes size of building for businesses
- Street address and precise lat-long

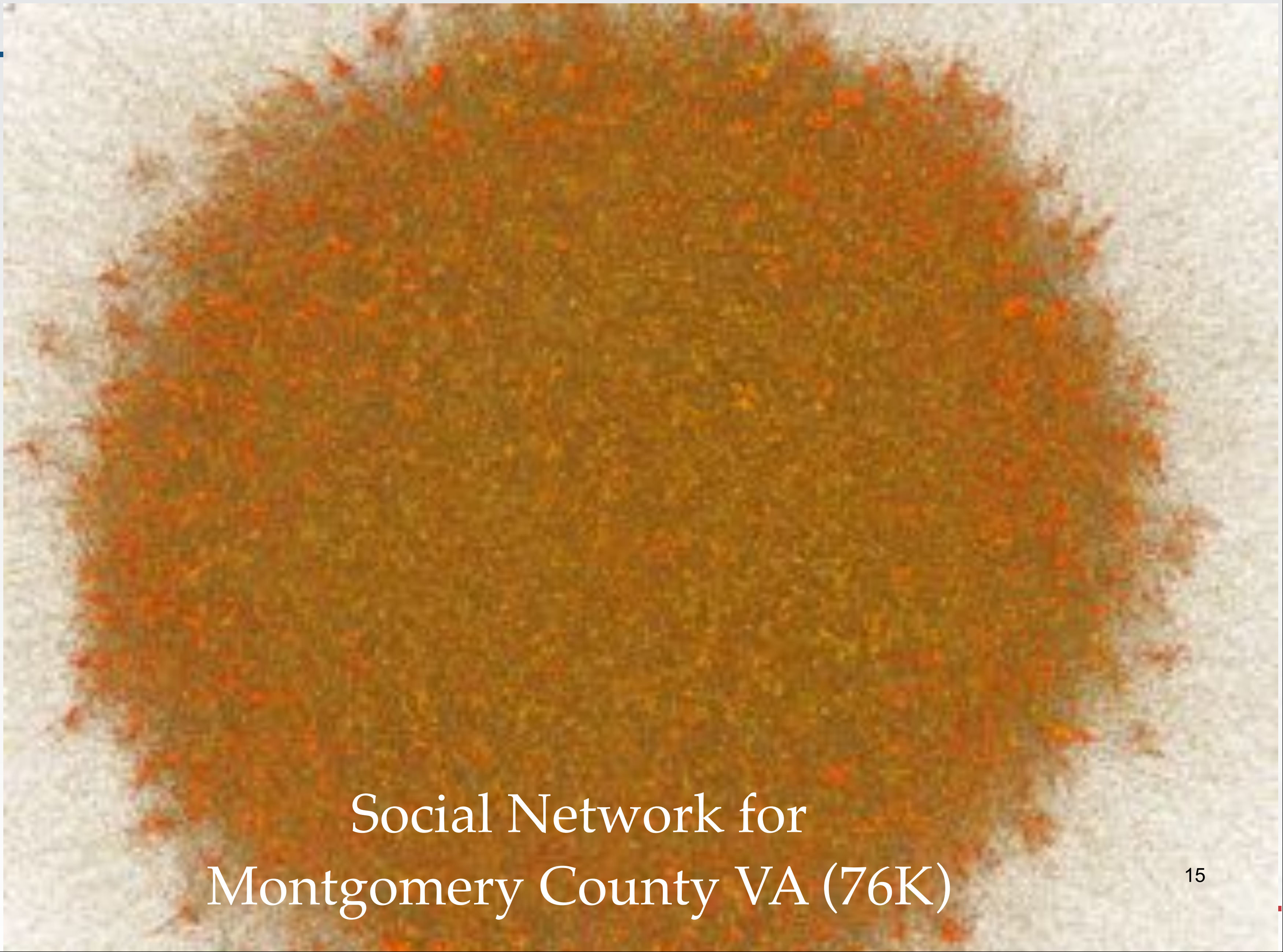
Building a Synthetic Population: Activities

- Activity Surveys
  - Daily activities type and time
  - Demographically matched to households

Synthetic Population Generates Social Networks

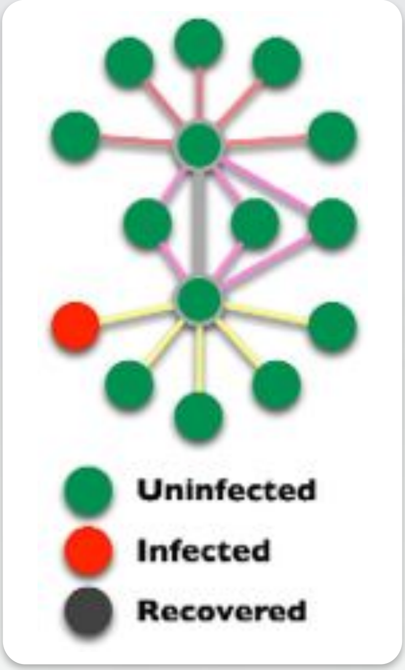


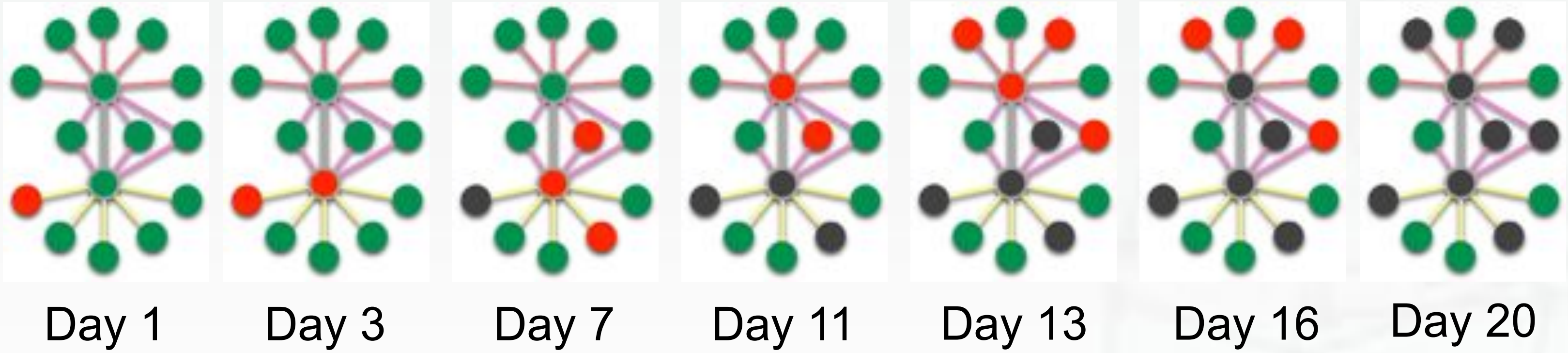
Social Networks get Very Complex



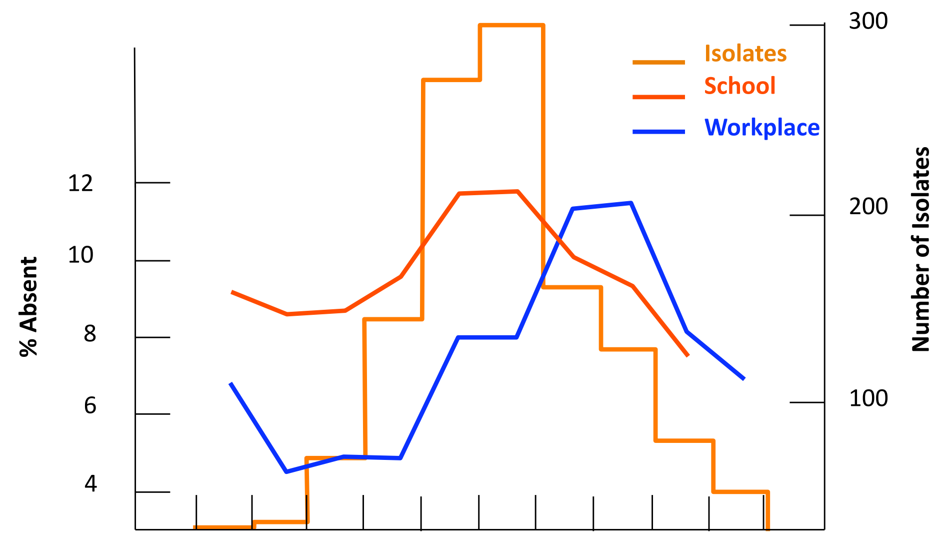
Network Based Simulation

- Starting with the social network, disease can spread through the network
  - Efficient methods exist for this approximation
- Can also map the kind of "interaction" thus alter base on interventions that affect these interactions (e.g. school closure)

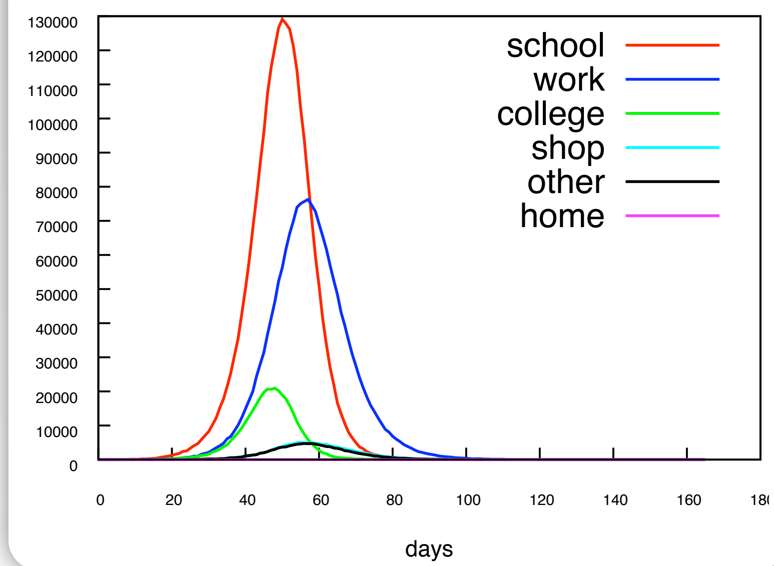




Real World Consistency



Glezen WP, Couch RB. Interpandemic influenza in the Houston area, 1974-76. N Engl J Med 1978;298:587.

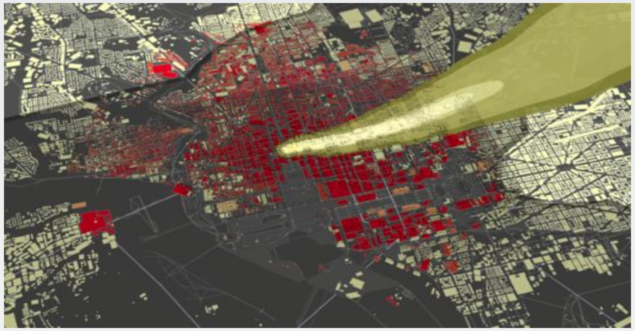


VBI Modeling Result from PNAS study

Halloran ME, Ferguson NM, Eubank S, Longini IM Jr, Cummings DA, Lewis B, Xu S, Fraser C, Vullikanti A, Germann TC, Wagener D, Beckman R, Kadau K, Barrett C, Macken CA, Burke DS, Cooley P. Modeling targeted layered containment of an influenza pandemic in the United States. Proc Natl Acad Sci U S A. 2008 Mar 25;105(12):4639-44

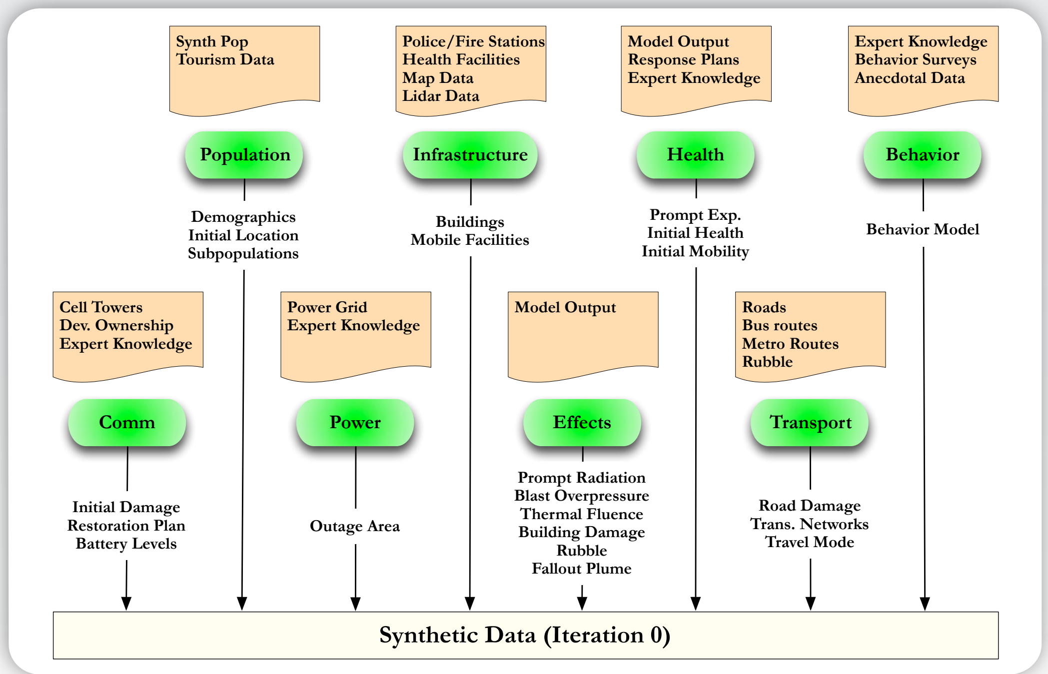
Synthetic Information Approach for Public Health

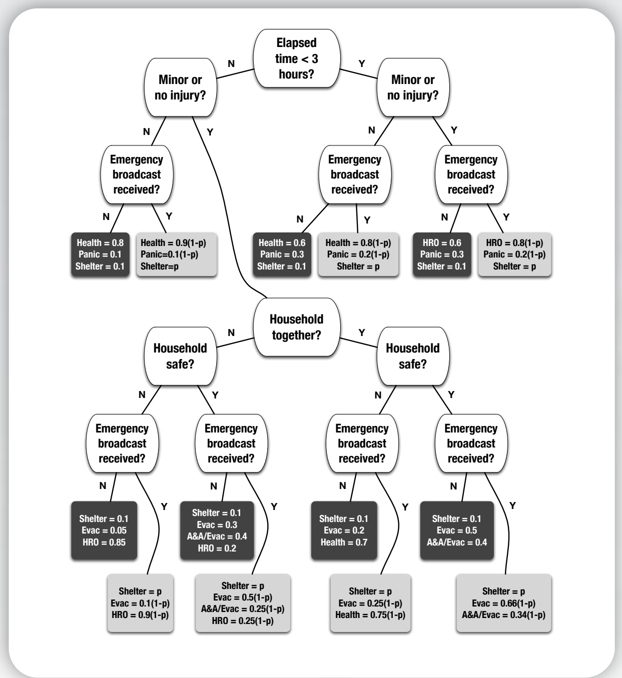
- **Demonstration Study:** Disaster Preparedness and Emergency Response
  - Large-scale destruction of urban area
  - Multiple interdependent systems fail
  - Failures cascade and cause other failures
  - Population in flux, recovering from event



- **Question:** Can the population be made more resilient by enhancing its ability to communicate?
  - Must simulate the dynamics in the short-term following the event, including behaviors, travel, health issues, etc.
  - Minor communication restoration following event requires simulation of all telecommunications in the region

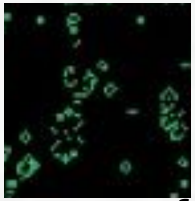
SYNTHETIC INFORMATION APPROACH





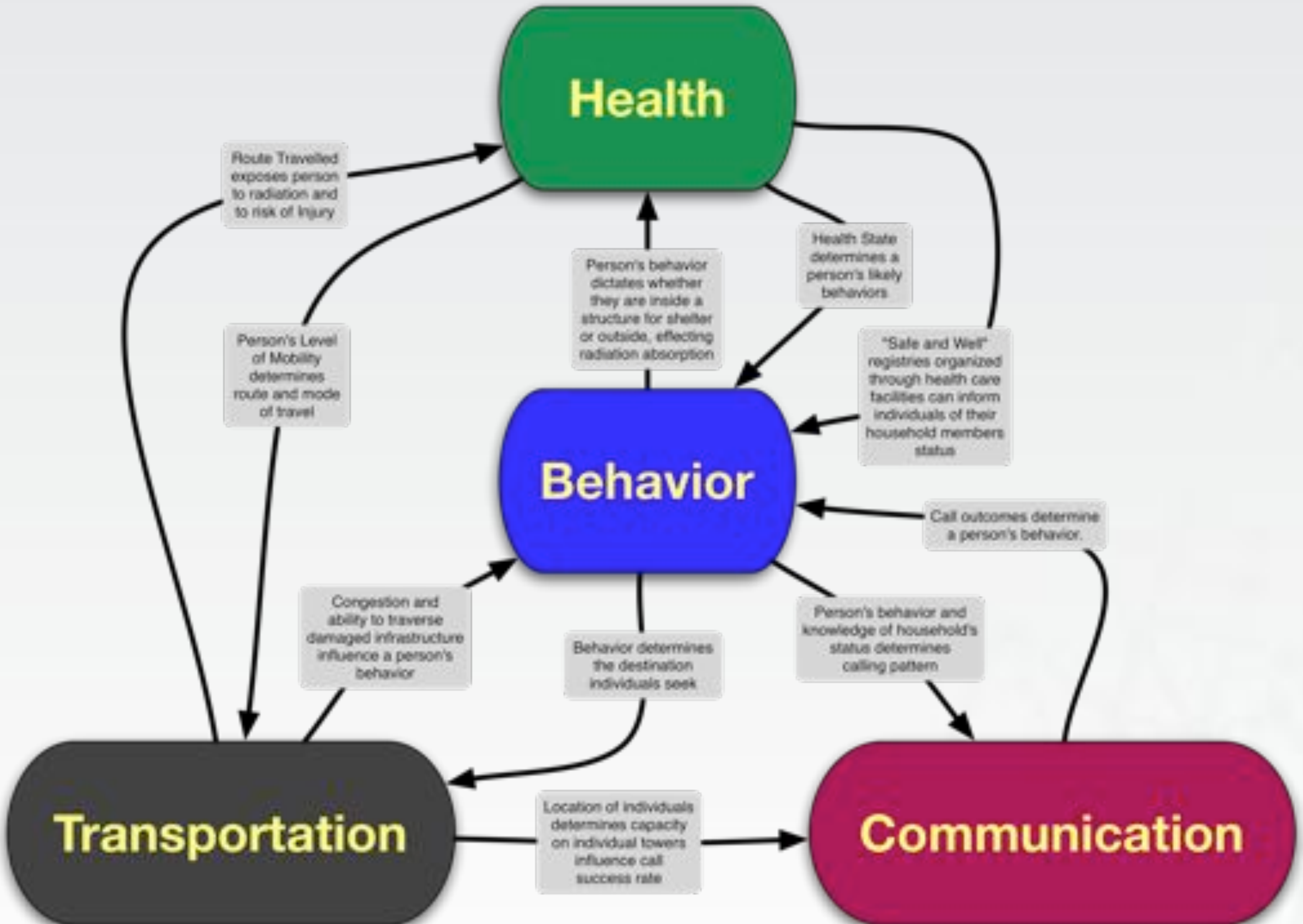
Lewis B, Swarup S, Bisset K, Eubank S, Marathe M, Barrett C. A simulation environment for the dynamic evaluation of disaster preparedness policies and interventions. J Public Health Manag Pract. 2013;19 Suppl 5:S42–8. doi:10.1097/PHH.0b013e31829398eb.

- Fusion of many types of data
- Rules at individual element level
- Novel behaviors emerge
  - Similar to a multi-dimensional and more complex “game of life”



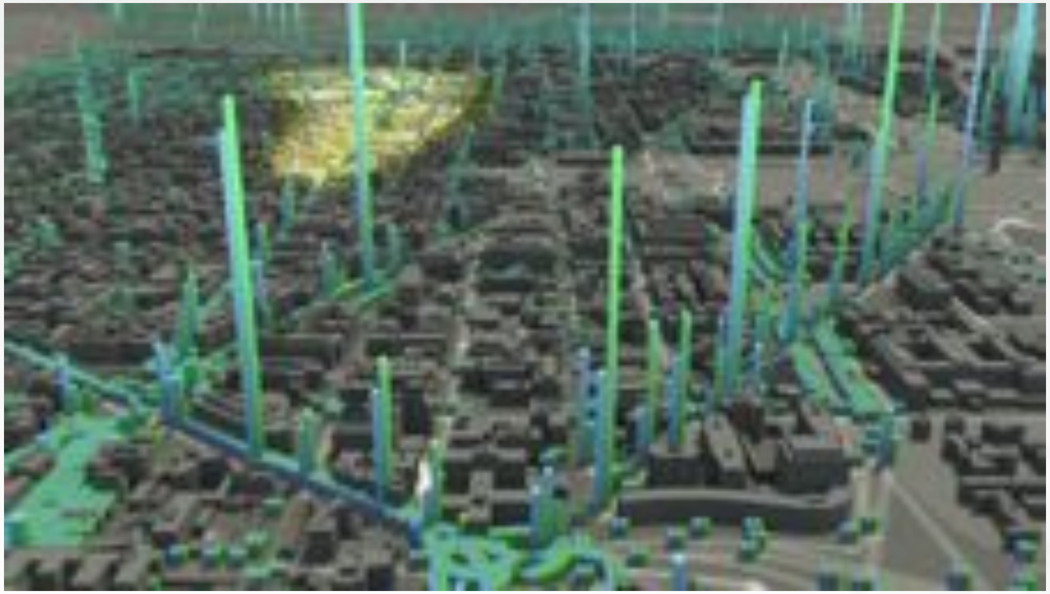
http://www.pygame.org/project-Cellular+Automata-662-.html

Synthetic Information Capturing Interdependencies



Synthetic Information Demonstration

- **Study Outcomes:** A seemingly small number of increased phone calls has benefits
  - EBRs increase sheltering in place
  - More information about family members decreases panic and coordinates evacuations
  - Health status of populations is generally better



Synthetic Information Systems

- **Advantages:**
  - Supports a variety of highly-detailed simulation techniques
  - Fusion steps allow a data structure that can provide insights through analysis alone without any dynamics
  - More structure allows the modeled system to behave more similarly to the represented system
- **Limitations:**
  - Computationally very taxing, requires use of high-performance computing resources and intensive database operations
  - Complexity of system can make designing experiments, analyzing, and interpreting them challenging

DESIGN AND EVALUATE SURVEILLANCE SYSTEMS

Case Study

EVALUATION OF HARVARD PILGRIM HEALTH SURVEILLANCE FOR DETECTING INFLUENZA-LIKE ILLNESS OUTBREAKS

Lewis B, Eubank S, Abrams AM, Kleinman K. In silico surveillance: evaluating outbreak detection with simulation models. BMC Med Inform Decis Mak. 2013;13(1):12. doi:10.1186/1472-6947-13-12.

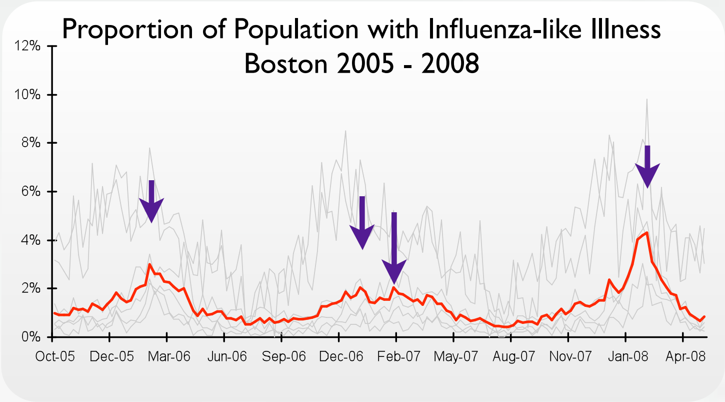
http://ndssl.vbi.vt.edu/insilicoSurveillance/

Problem: Find Real Outbreak in ILI Noise

- Many methods have been developed
  - Clustering: temporal, spatial-temporal, combined with scan statistics, etc. (SaTScan a popular free toolkit)

Nice review: Buckeridge DL. Outbreak detection through automated surveillance: a review of the determinants of detection. *J Biomed Inform*. 2007;40(4):370–379. doi:10.1016/j.jbi.2006.09.003.

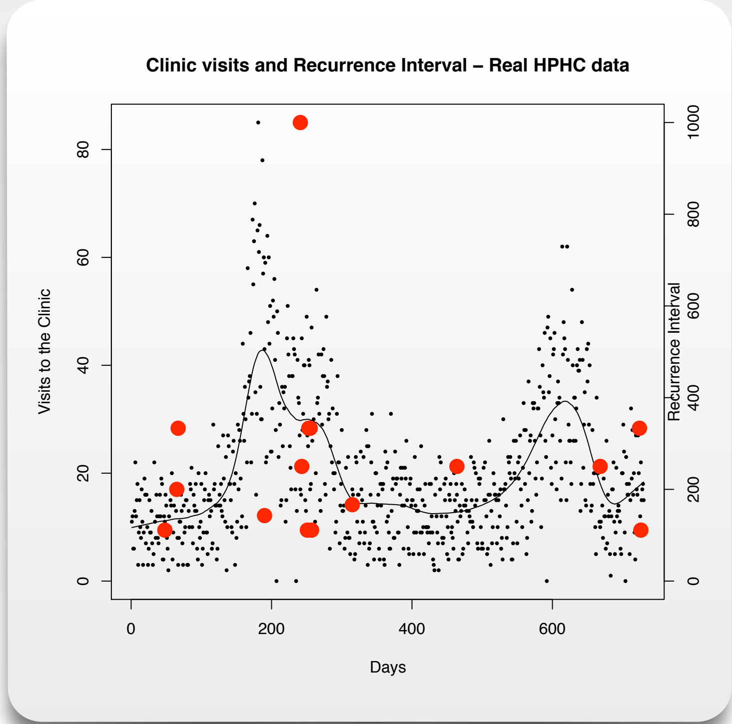
- **Challenge:** How to evaluate the performance of methods against each other, different "tunings" of the same tool, and for different disease "types"

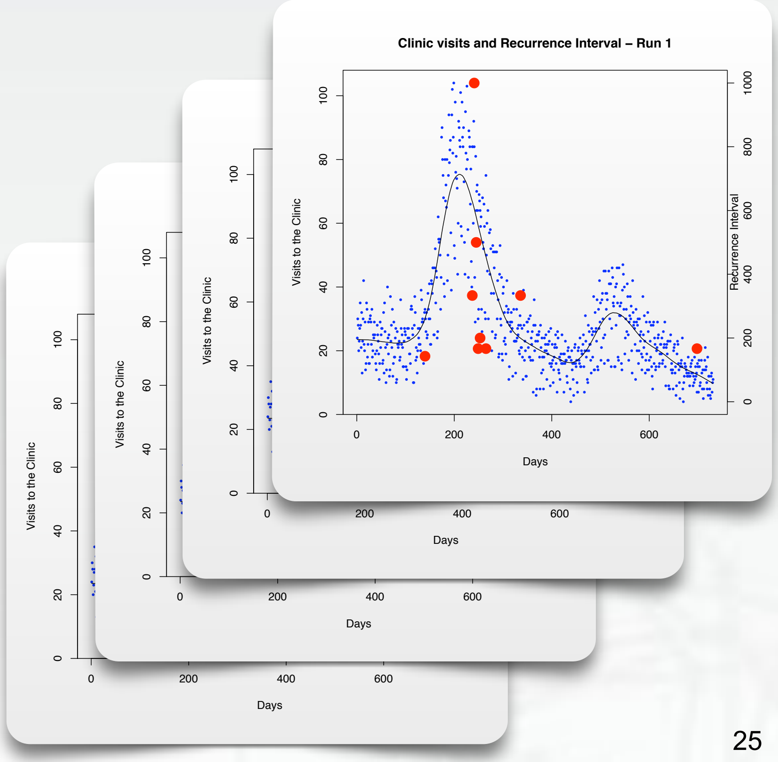


Justin Pendarvisa, Erin L. Murrayb, Marc Paladinib, Julia Gunna, Donald R. Olson. Age Specific Correlations between Influenza Laboratory Data and Influenza-like Syndrome Definitions in Boston and New York City. Presentation, 2008

Role of Simulation

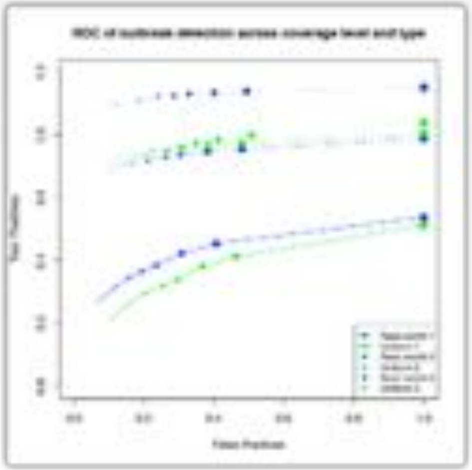
- Produce large realistic data sets for thorough evaluation of outbreak detection algorithms
  - Provides larger sample size than historical data sets

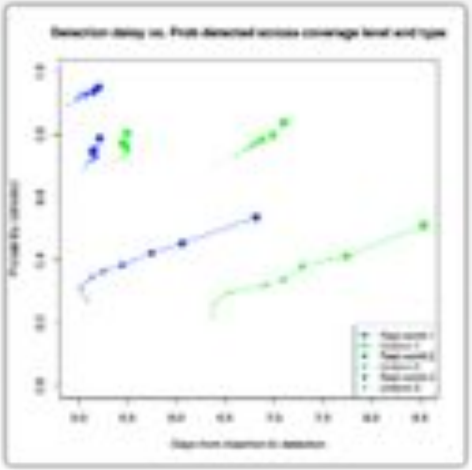




Role of Simulation

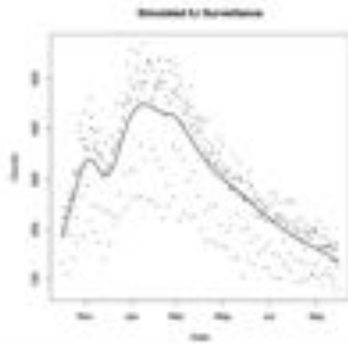
- Provides a configurable framework for studying surveillance system designs
  - Catchment population
    - Locations of clinics or surveilled populations
    - More low-quality vs. targeted high-quality
  - Performance of diagnostic tests employed
  - Optimizing for sensitivity or specificity or “type” of outbreak





1. Simulate Surveillance Data

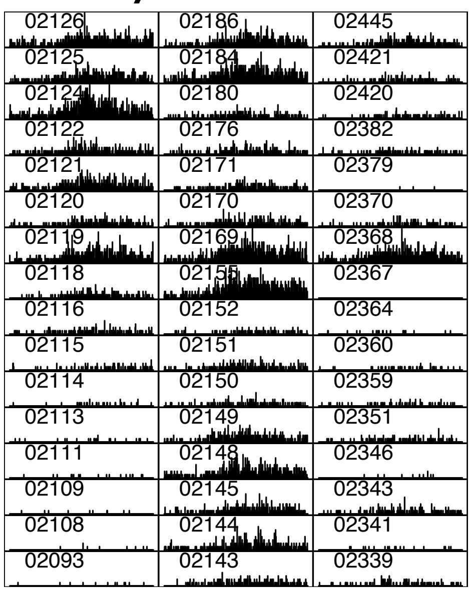
Disease model parameterized for disease of interest. Calibrate to approximate global disease levels, add other influences like seasonality and disease interventions. Health care seeking and other behavioral modeling fine tune the simulation results to the desired surveillance system

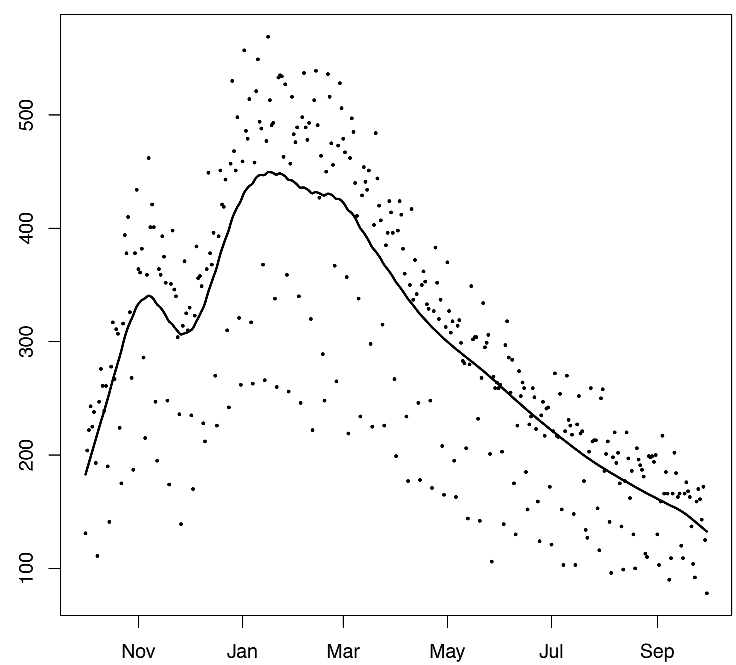


Demonstration Study Specifics:

A. Calibrate ILI disease for endemicity
B. Make global adjustments to transmissibility to create seasonal peaks
C. Determine if and when a case will seek health care using delay to care and day of week bias
D. Determine if this person is a member of the surveillance system
E. Sum surveilled cases by zip code for each day

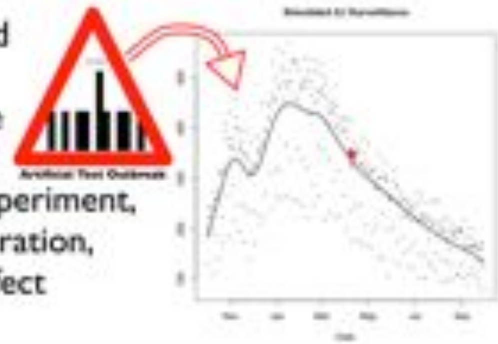
A haystack of data





2. Insert Standardized Outbreaks

Artificial outbreaks are inserted into the synthetic disease surveillance data stream. These test outbreaks factor into the specific design of the *in silico* experiment, as true-positives. The shape, duration, method of insertion, etc. can effect detection.



Demonstration Study Specifics:

A. Select a random day for artificial outbreak insertion

B. Scale outbreak case numbers to reflect coverage level of surveillance system

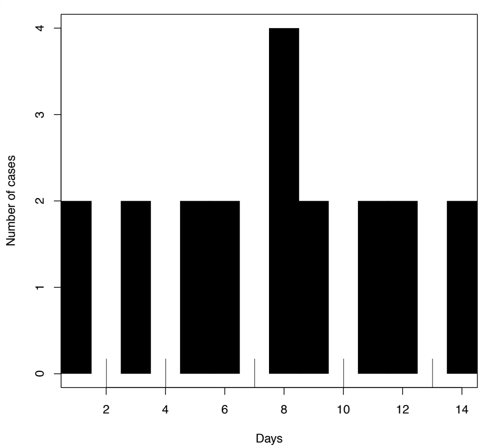
C. Select random location and 2 neighboring locations for insertion

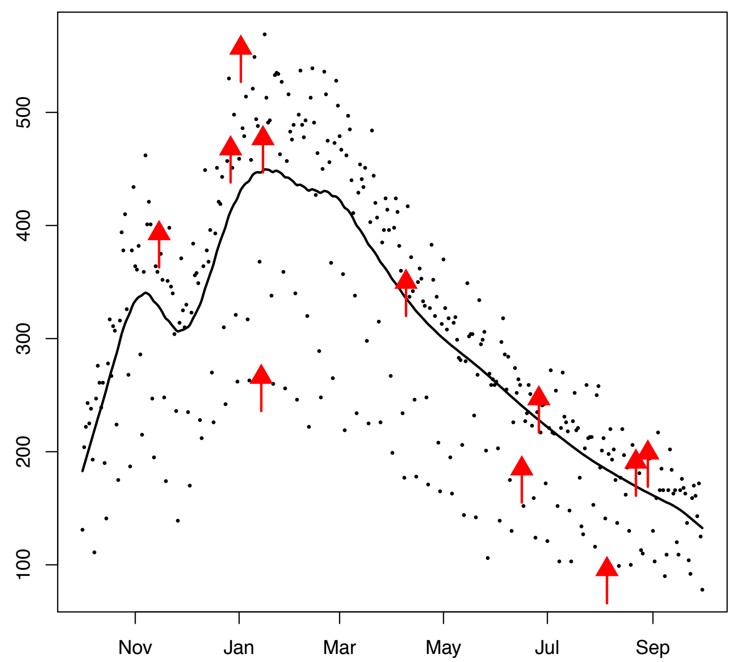
D. Add outbreak cases to surveillance data

E. Remove inserted cases, and repeat 11 times.

F. Retain all 12 data sets with inserted cases and original simulated surveillance data for analysis.

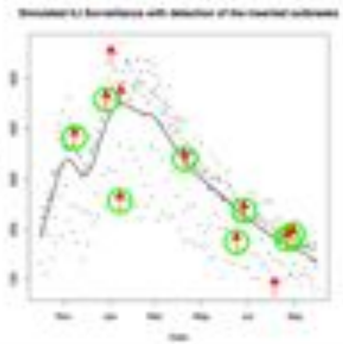
The needle





3. Detect Outbreaks in Synthetic Surveillance Data

Surveillance signals with the inserted test outbreaks are evaluated with outbreak detection algorithms. The synthetic surveillance signal without any inserted outbreaks is also evaluated, these outbreak events can be considered false-positives.



Demonstration Study Specifics:

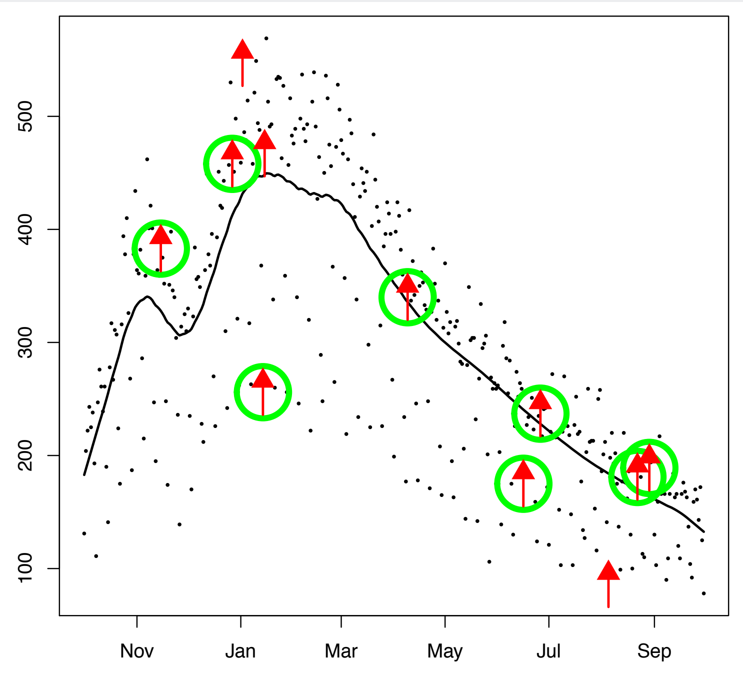
A. Perform SaTScan analysis for every day of an entire simulated ILI season as well as all independent inserted outbreaks

B. Merge SaTScan identified clusters that overlap in time and space under appropriate detection thresholds

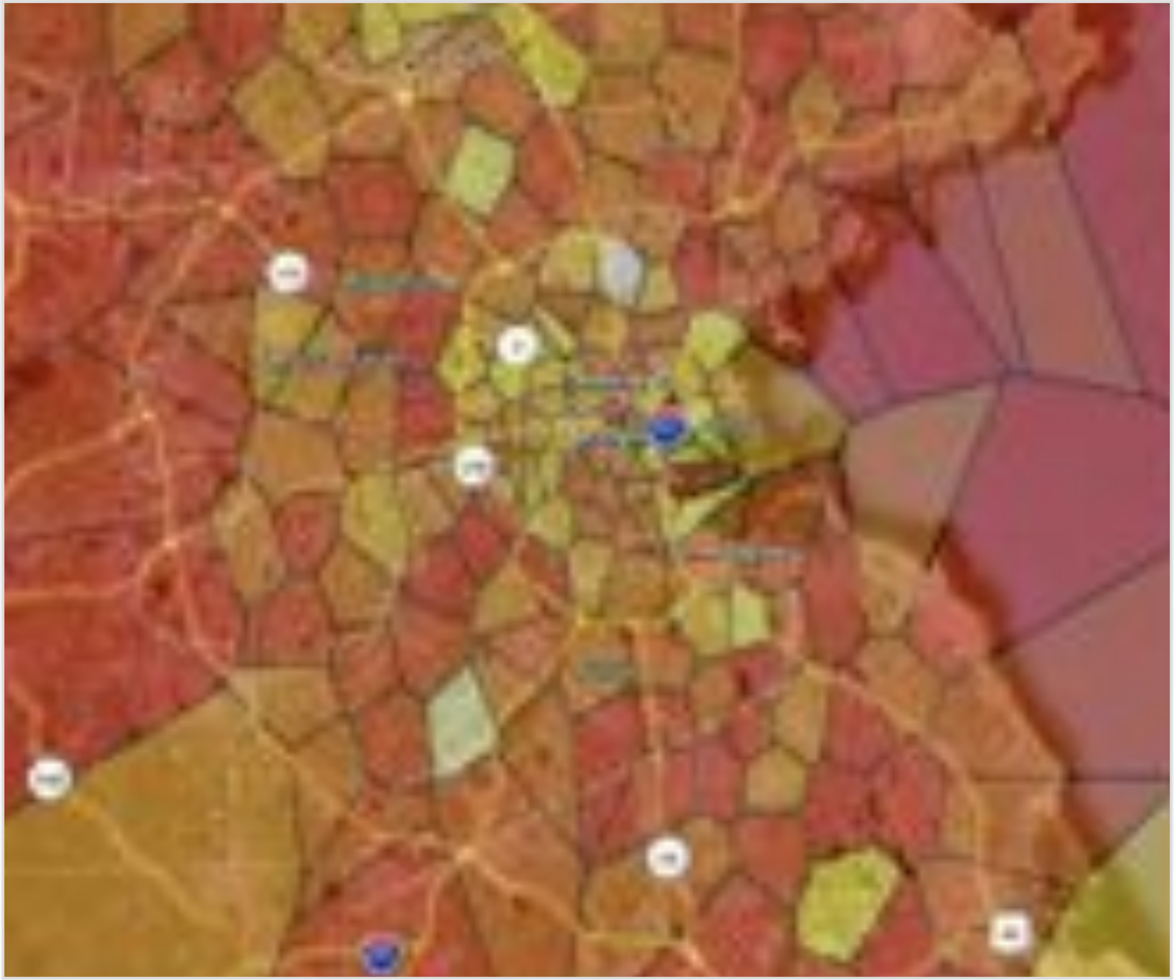
C. Identify which SaTScan identified clusters correspond to inserted outbreaks

D. Evaluate surveillance system performance

SaTScan

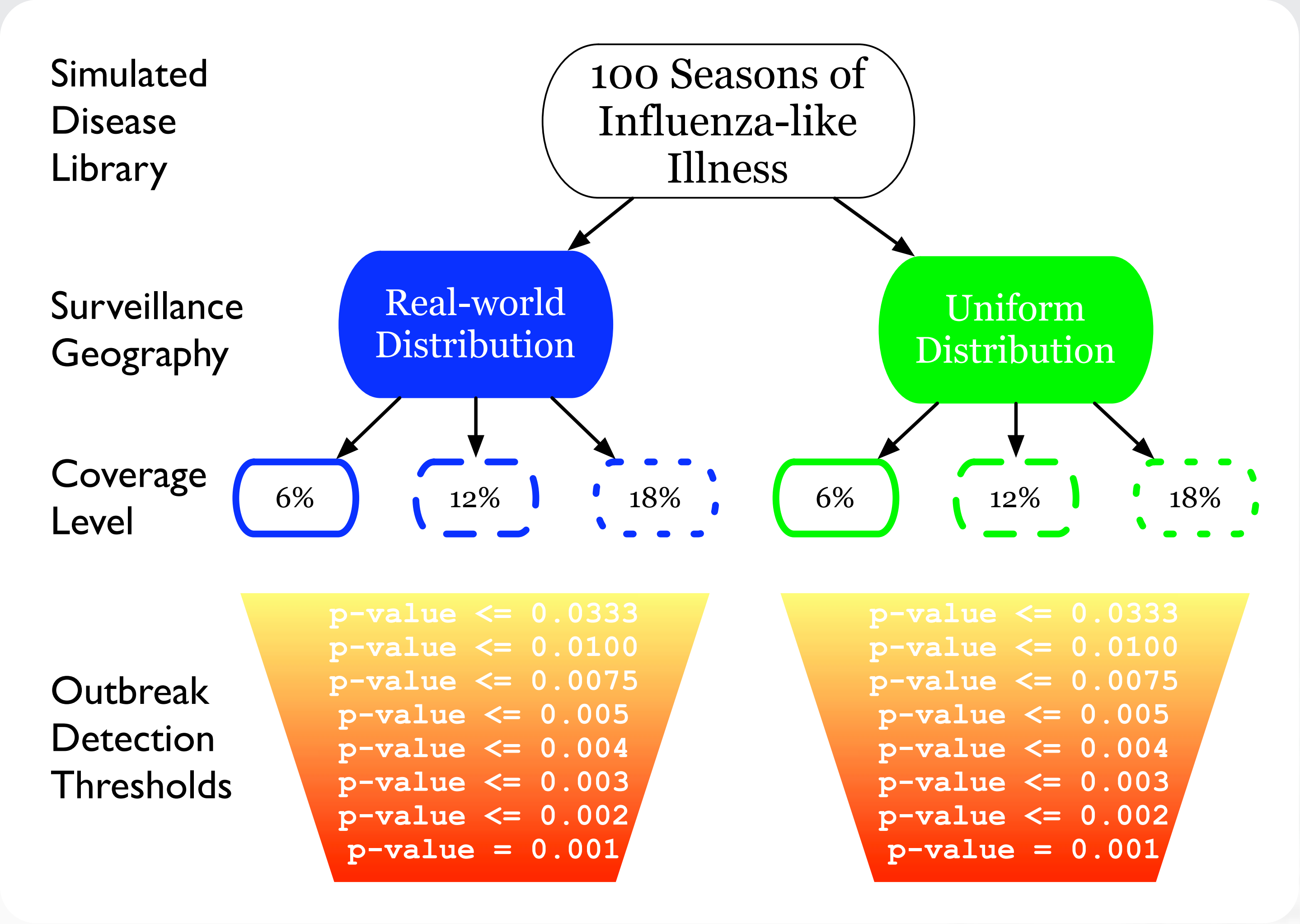


Synthetic Surveillance Catchment



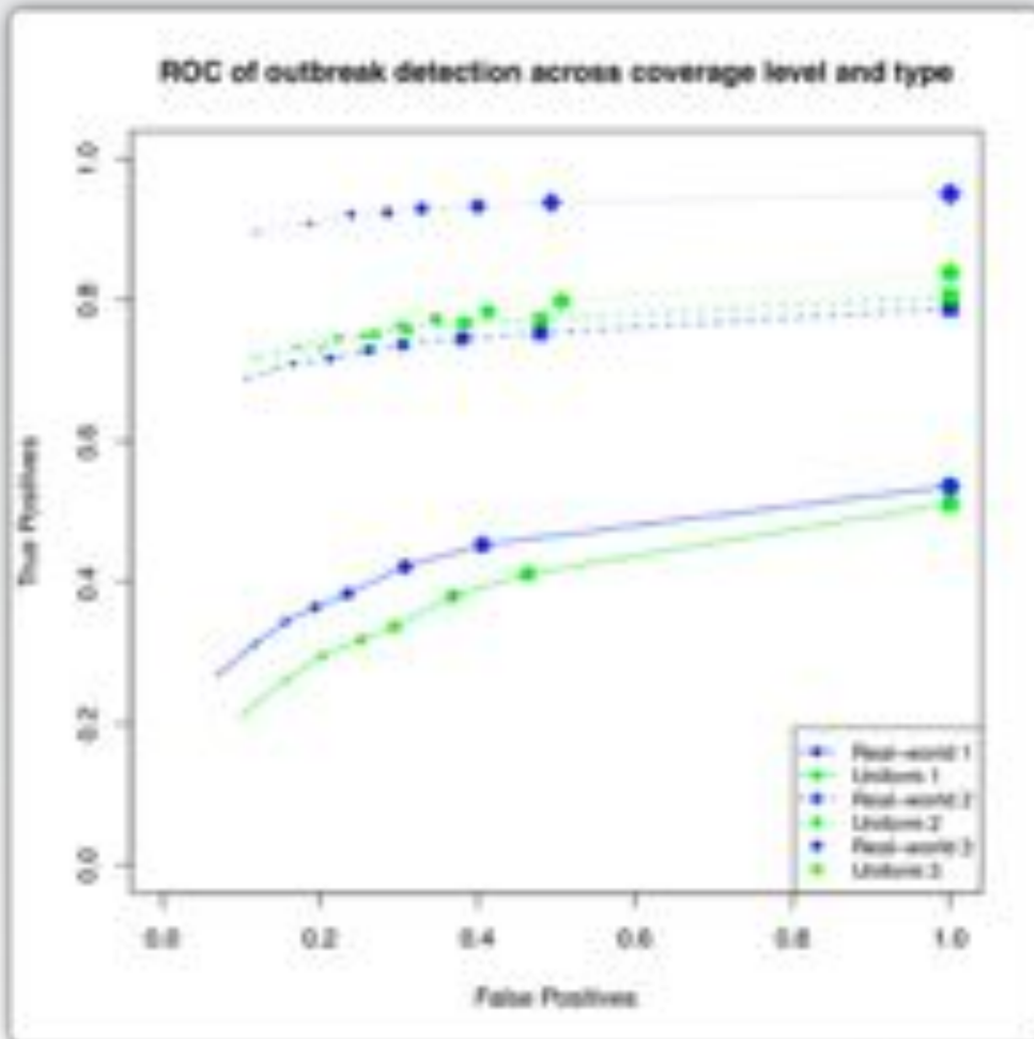
Proportion of each zip code that belongs to the surveillance system

Study Design

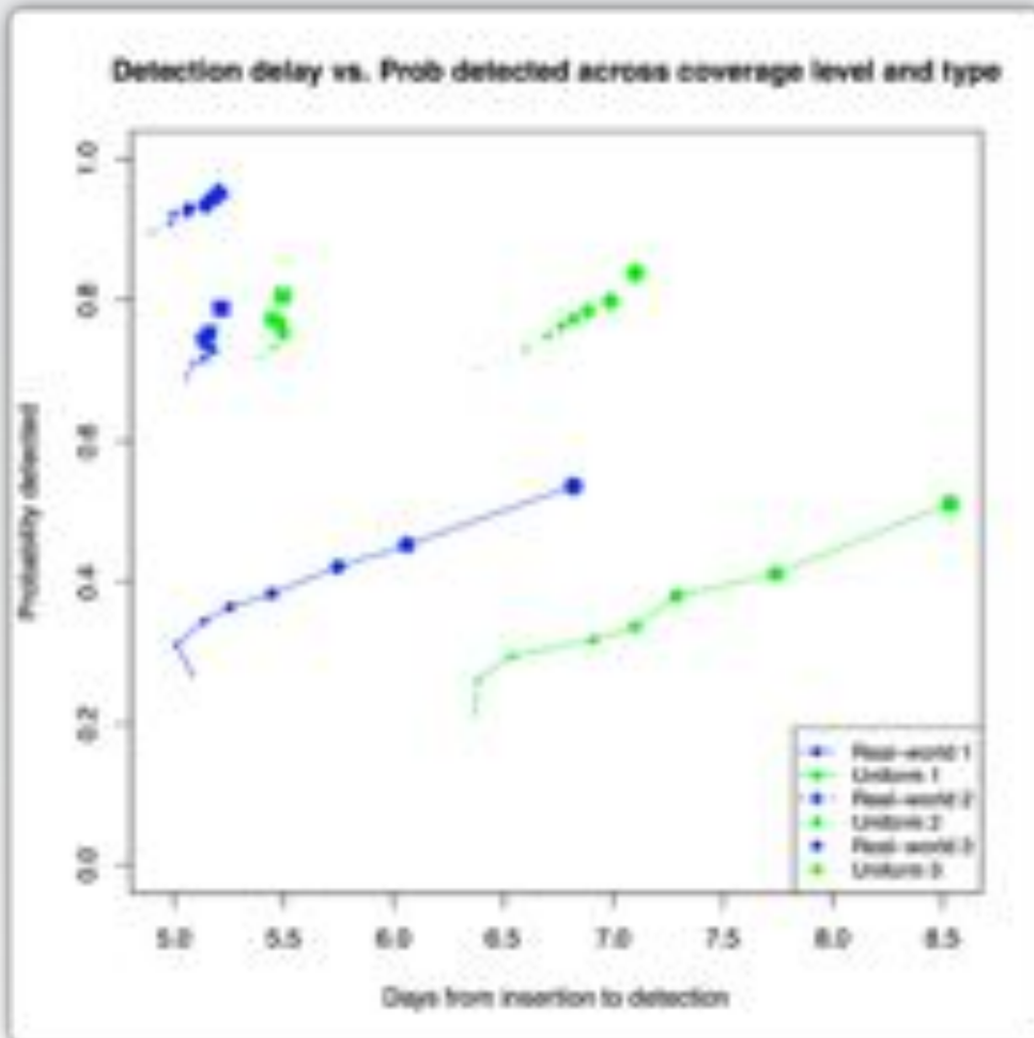


Results: Movie of Detection

Results: Detection Performance



Results: Time to Detection Performance



Impact for Infectious Disease Surveillance

- Framework for evaluating outbreak detection
  - Offer a way to improve existing methods or develop new ones
  - Explore cost-benefit of surveillance system modifications

- Provide guidelines for public health practice
  - For a given surveillance system, estimate appropriate thresholds for what qualifies as an outbreak

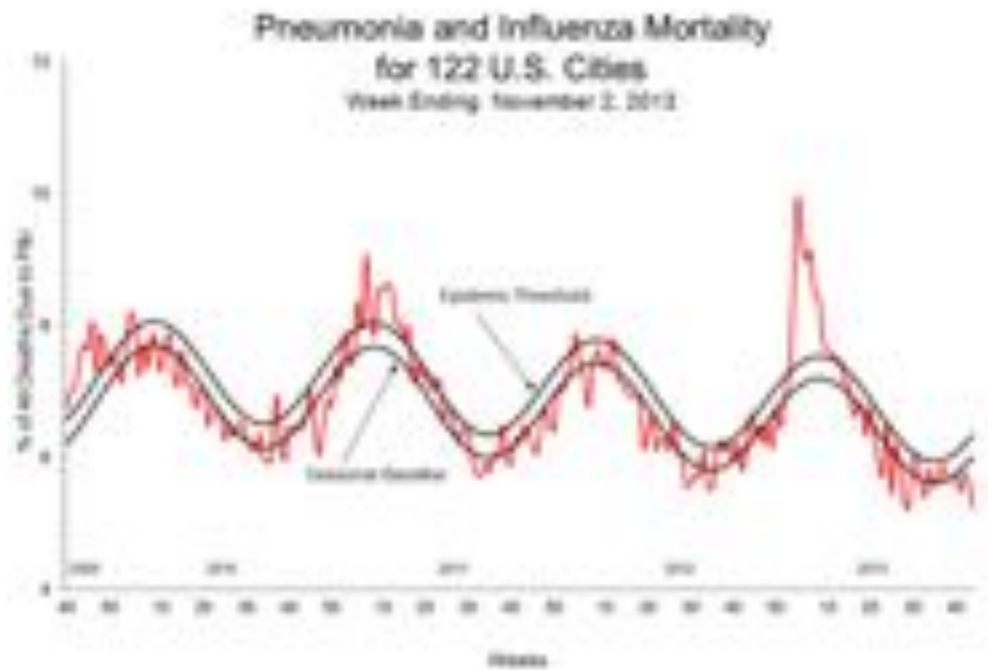
SIMULATION ENHANCED SURVEILLANCE

Case Study

Using Structured Agent-Based Simulations Combined with Updated Surveillance Data to Forecast Future ILI Activity

Problem: Forecasting Future Depends on Now

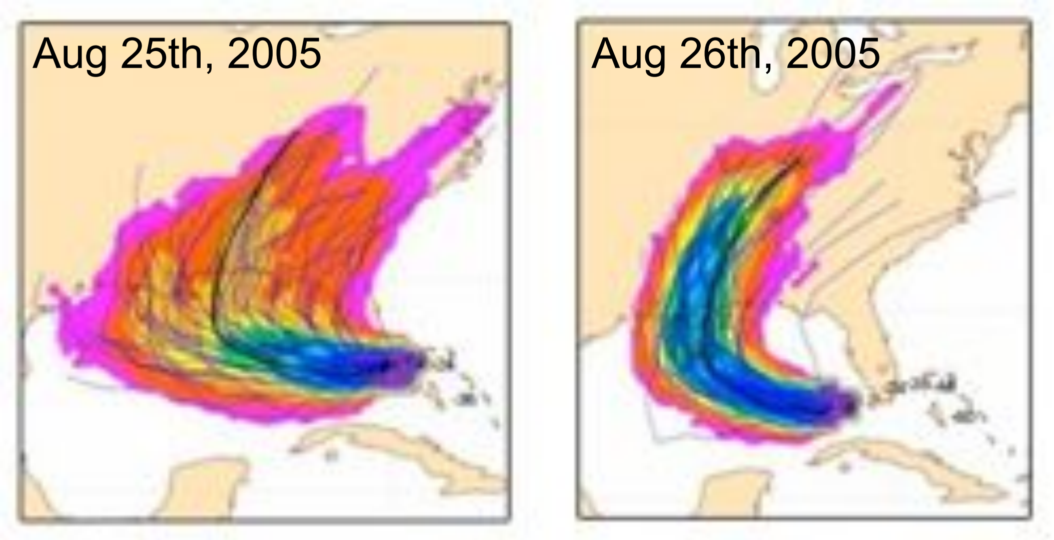
- "Now" is rarely well understood (and subject to revision when it becomes "Then")



http://www.cdc.gov/flu/weekly/weeklyarchives2013-2014/bigpi44.htm

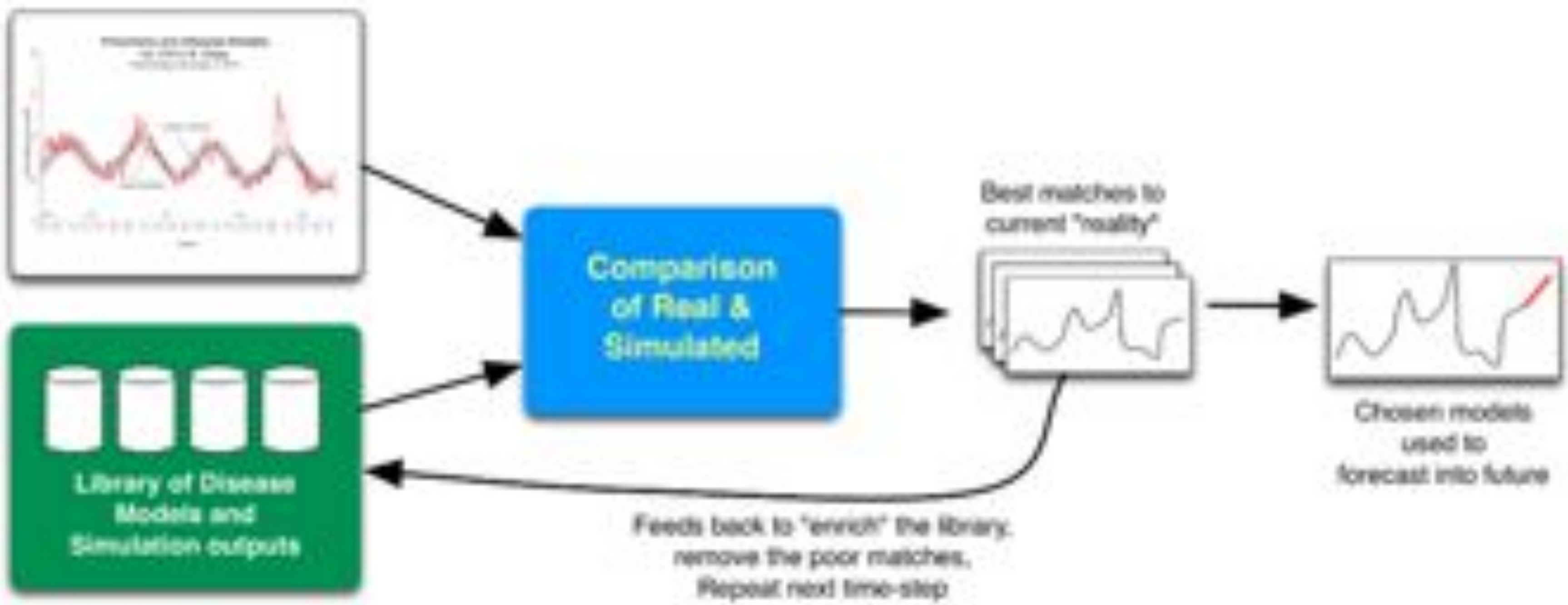
Surveillance Forecasting with Simulation

- Inspired by weather forecasting
  - Real-world observations are compared to simulated results, best matching simulation used for forecasting
  - More advanced techniques, use ensembles of models and specific realizations of parameters sets
    - Simulations with parameter sets that continue to match are kept in the ensemble, those that don't match are discarded



Forecasting and Nowcasting

- Iterative loop allows simulations one to hone in on the best match to reality and better understand "now" as well as be more accurate about the future



Impact for Infectious Disease Surveillance

- Using simulations with deeper structures (like school and workplace attendance) in this loop, allows the simulation to adapt more appropriately with reality
  - Challenge: Situational awareness of the mitigation strategies

- Future Work: Adopting this technique in a synthetic information environment allows the inclusion of disparate data sources
  - Social media (Twitter, Google Search trends, news, blogs)
  - Unify multiple surveillance streams (types and locations)

QUESTIONS??

FUNDING

DTRA CNIMS Contract HDTRA1-11-D-0016-0001,
Defense Threat Reduction Agency Comprehensive National Incident Management System

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 2U01GM070694-09 The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Thanks to:
Entire group at the Network Dynamics and Simulation Science Laboratory

Ken Kleinman and Allyson Abrams
for their help with the in silico Surveillance Evaluation Study