Command, Control, and Interoperability
Center for Advanced Data Analysis

Information Theoretic Tools and Mathematical Epidemiology In Biosurveillance

A.A. Yakubu (Howard)
N.H. Fefferman (Rutgers)
A. Nkwanta (Morgan State)
W. Powell (Princeton)
D. Hartley (Georgetown, FAZD)
H. Gaff (Old Dominion, FAZD)

Tiny Killers

"When we think of the major threats to our national security, the first to come to mind are nuclear proliferation, rogue states and global terrorism. But another kind of threat lurks beyond our shores, one from nature, not humans---an avian flu pandemic."

President Barack Obama

Emerging and Re-emerging Infectious Diseases

The key to predicting the possibility of a new epidemic is to understand if a particular virus carried by certain animals can mutate, get transmitted, and finally spread among humans.

Question:

What are the mutations that must take place to enact such a scenario?

Mathematics of Infectious Disease

Interdisciplinary Research

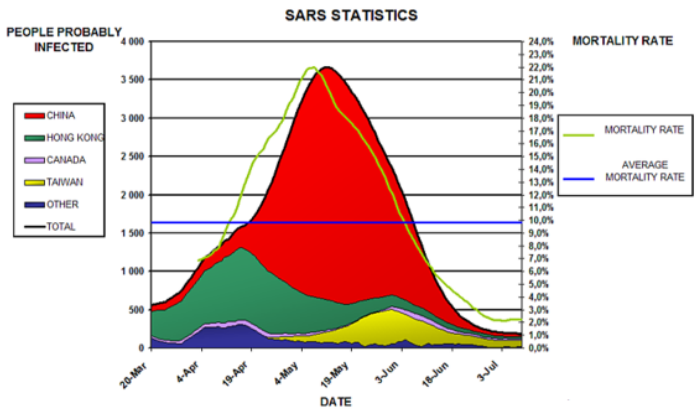
Command, Control, and Interoperability Center for Advanced Data Analysis

History of Mathematical Epidemiology

- Daniel Bernoulli (1760). Defense of the practice of inoculation against smallpox
- P. D. En’ko (1873-1894). Compartmental models
- A. G. McKendrick and W. O. Kermack (1900-1935). R_0
- R. A. Ross (1900-1935). Malaria

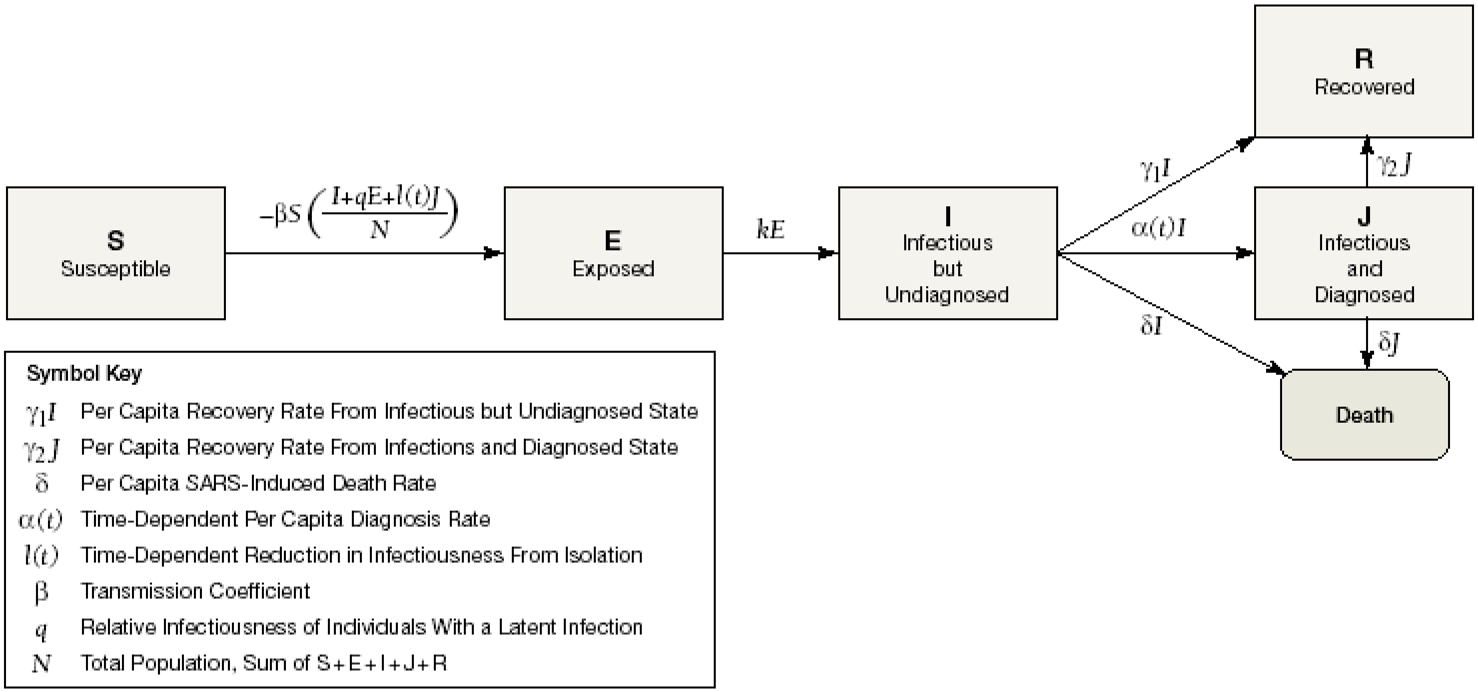
Predictive Power of Models: The case of SARS in Toronto

- A Mathematical model of SARS showed how Isolation and Quarantine measures could reduce the size of a SARS outbreak by a factor of 1000. The mathematical results agreed with actual observations in the greater Toronto area (JAMA, 2003 and J. Theor. Biol. 2003).



SARS Model: An example

Figure. Box Diagram Illustrating a Mathematical Model of Outcomes of a SARS Epidemic



Symbol Key

$\gamma_1 I$	Per Capita Recovery Rate From Infectious but Undiagnosed State
$\gamma_2 J$	Per Capita Recovery Rate From Infections and Diagnosed State
δ	Per Capita SARS-Induced Death Rate
$\alpha(t)$	Time-Dependent Per Capita Diagnosis Rate
$l(t)$	Time-Dependent Reduction in Infectiousness From Isolation
β	Transmission Coefficient
q	Relative Infectiousness of Individuals With a Latent Infection
N	Total Population, Sum of S+E+I+J+R

R_0

- A fundamental concept of mathematical epidemiology is that a threshold R_0 (basic reproduction number) can be identified.
- R_0 estimates the average number of secondary infections generated by a typical infectious individual with a given infection.

$$R_0 > 1$$

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis

$$R_0 < 1$$

Model Importance

- Models can provide understanding of the underlying mechanisms that influence disease outbreaks and suggest surveillance techniques.
- Models can help to determine best surveillance techniques as well as how they might affect short- and long-term disease dynamics.
- Models offer ways of integrating population level knowledge based on previous disease outbreaks with available individuals and population data to predict the outcomes of several alternate scenarios
- Models can be used as educational and research tools for simulations and explorations
- Mathematical models are capable of predicting outbreaks when real data is incomplete or inaccurate

Early Detection Of Disease Outbreaks Is Crucial For Public Safety

Failure of biosurveillance increases disease incidence and mortality

True with all infectious diseases including

- natural exposure from zoonotic infections
- purposeful acts of bioterrorism
  - ➢ Smallpox
  - ➢ Avian influenza
  - ➢ Rift valley fever
  - ➢ Brucellosis
  - ➢ Tularemia
  - ➢ Anthrax

Image from <www.alpharubicon.com/basicnbc/basicnbc.htm>

Image from <microbes.historique.net/anthracis.html>

Goal of Biosurveillance: early and accurate detection of outbreaks from multi-source, multi-scale, health-related data

Current methods are statistical: analyze existing data sets and set acceptance thresholds for “normal”

- Require historical data
- Usually require manual data manipulation to be effective

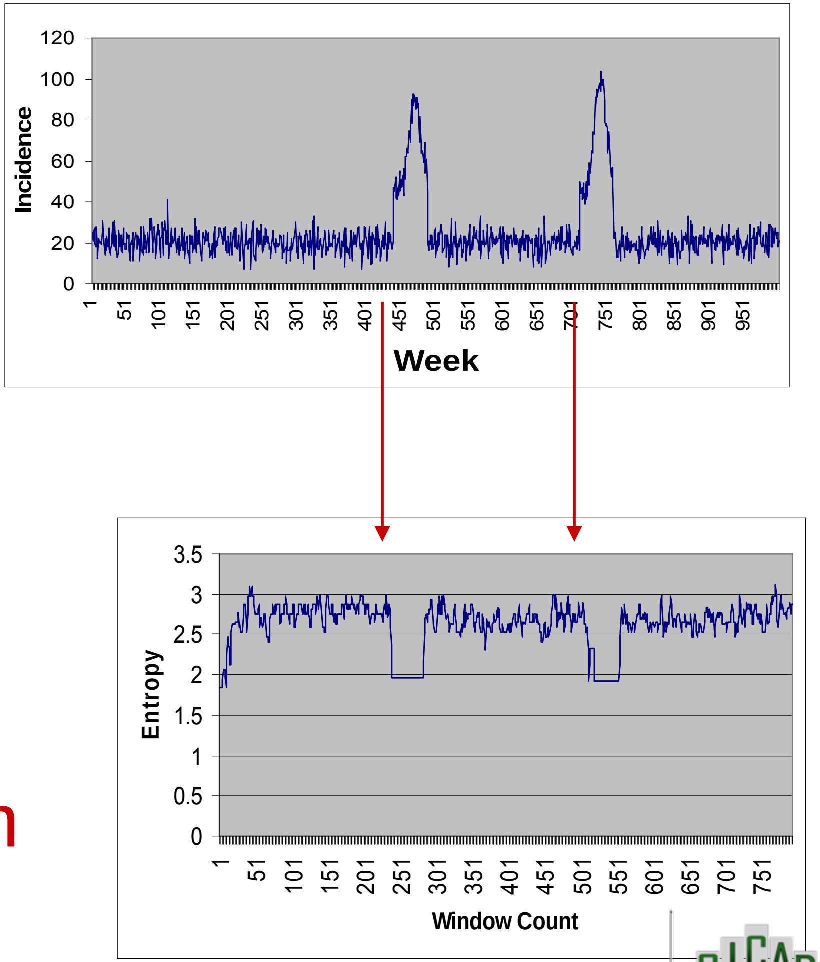
New idea: Explore potential of information theoretic measures

- Quantify ‘information’ in a signal, not values of the signal
- May not require historical data to be effective in detecting transitions from “normal” to “abnormal”

Entropy and Biosurveillance

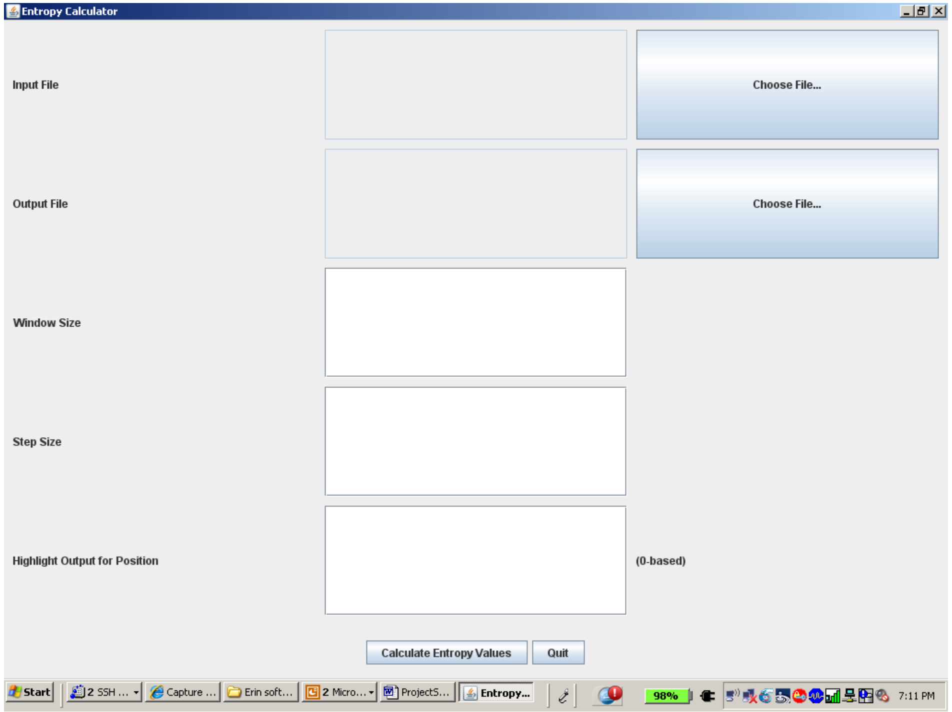
Early work done as MSI summer teams showed concept can work

Lots of work is still needed to produce a general algorithm for practical use



Recently completed software to support research

- Initial measure from 2007 teams – based on Shannon entropy
- New entropy measure – based on LZW entropy



In the process of computational experimentation to discover best general algorithms

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis

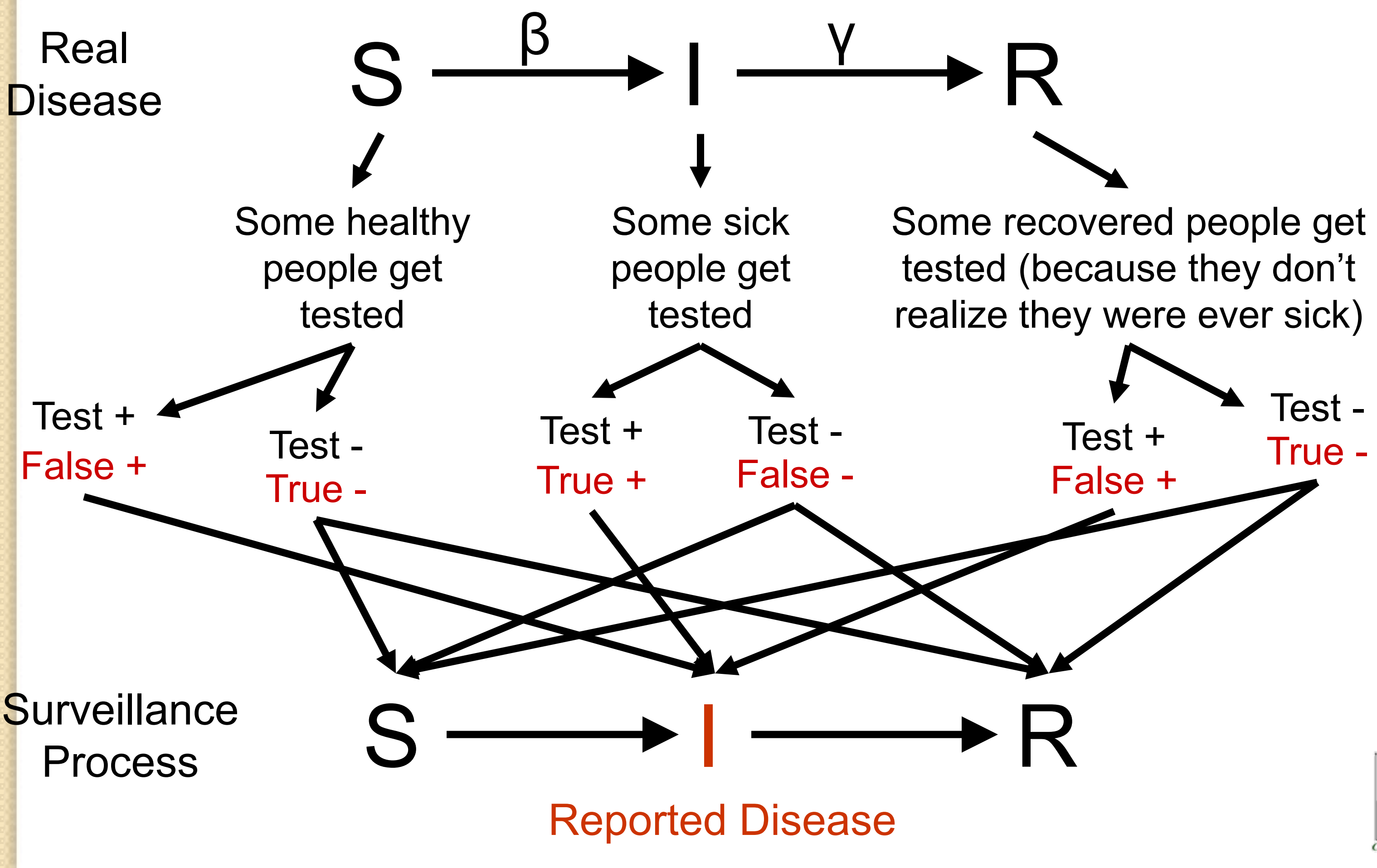
Collaboration with FAZD

1. Work with animal/wildlife outbreaks
    - Different spatial/time scales of surveillance
    - Different types of measured outcomes as system input
    - Different time windows for effective control
2. Detect changes in disease vector populations
    - Presence vs incidence
    - Density vs frequency, etc.

Human behavior, Clinical diagnostic accuracy, and Biosurveillance

- Laboratory confirmation is the primary method of disease determination and confirmation. Only once a disease has been confirmed by laboratory test is it reported, and therefore, from a surveillance perspective, detectable. Unfortunately, for many of the diseases of greatest concern, laboratory tests are not 100% accurate (cf. Bonini et al. 2002).
- Human factor

Surveillance vs. Disease



Real Disease

S $\xrightarrow{\beta}$ I $\xrightarrow{\gamma}$ R

Some healthy people get tested

Some sick people get tested

Some recovered people get tested (because they don’t realize they were ever sick)

Test +
False +

Test -
True -

Test +
True +

Test -
False -

Test +
False +

Test -
True -

Surveillance Process

S → I → R

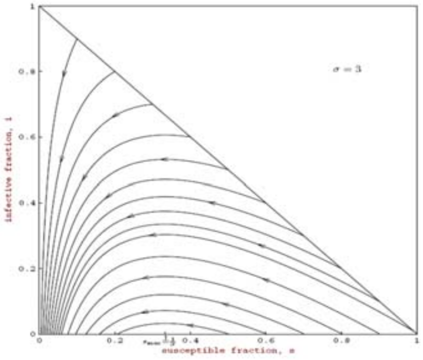
Reported Disease

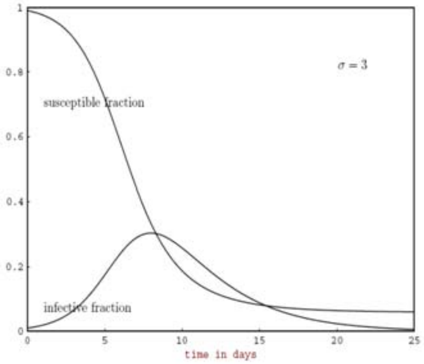
CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis

Classic SIR Epidemic Model

$$\begin{aligned} dS/dt &= -\beta S I \\ dI/dt &= \beta S I - \gamma I \\ dR/dt &= \gamma I \end{aligned}$$

$$S(0) = S_0 \geq 0,\ I(0) = I_0 \geq 0,\ R(0) = R_0 \geq 0.$$



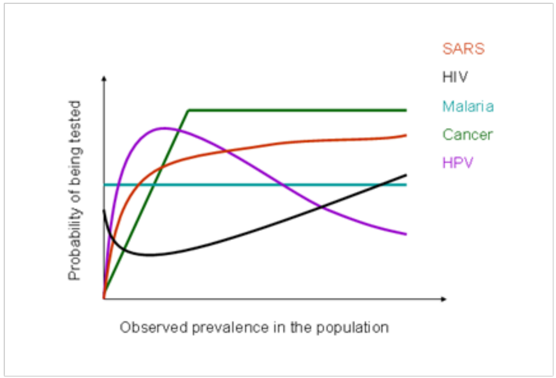


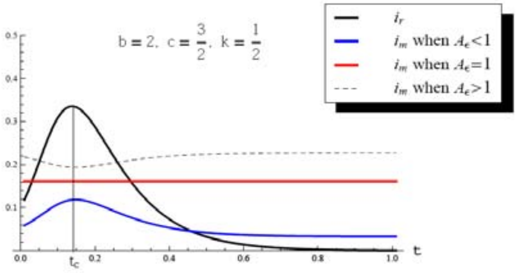
Question

How does implicit assumption of constant behavior and sensitivity and specificity in testing for disease incidence impacts the accuracy of estimated transmission parameters?

SURVEILLANCE S-I-R MODEL (Nina Fefferman, Nianpeng Li, etc)

$$\begin{aligned} S_m &= S_r(1-\alpha) + S_r\alpha(1-\epsilon_1) + I_r(1-\delta) + I_r\delta\epsilon_2 \\ &\quad + R_r(1-\lambda) + R_r\lambda(1-\epsilon_1) \\ I_m &= S_r\alpha\epsilon_1 + I_r\delta(1-\epsilon_2) + R_r\lambda\epsilon_1 \end{aligned}$$





Examples of HIV types: Graphs of i_m for different ϵ's.

Related Research Topics

- Application of entropy to biosurveillance and bioterrorism data.
- Algorithms for calculating and monitoring changes in entropy.
- Entropy aided detection of beginnings of outbreak scenarios.
- Tying in Infectious Disease Models

CCICADA

Command, Control, and Interoperability Center for Advanced Data Analysis

Thank You!