Accounting for the global under-reporting of Emerging Infections Diseases

Ilana Brito
Massachusetts Institute of Technology
CIESIN, Columbia University

Advances in Geospatial Technologies for Health Conference
Santa Fe, September 13, 2011
Motivation

• Number of EIDs is increasing
• Local diseases can spread globally
• Significant morbidity and mortality
• Devastating effects on economies
• Need to focus surveillance efforts

But how?
• Step 1:

• Where is the highest risk?
Emerging Infectious Disease Dataset

- 335 diseases that have emerged between 1940 and 2003
- Jones et al., Nature 2008

Criteria for inclusion:
- Novel disease in human population
- New geographic locales
- Development of drug-resistance
- Dramatic changes in virulence
What causes new diseases to emerge?

Ecological changes
- Land use changes
- Climate change
- Agriculture
  Intensification

Demographic changes
  Age shifts
  Population density

Bioterrorism

Changes in drug use
  - Personal
  - Agriculture
    - Factory runoff

Changes in immunity
  - HIV status
For zoonotic diseases,
High population density
Mammalian diversity
• Step 1:
  • Where is the highest risk?

• Step 2:
  • Where is surveillance/reporting poor?
Exposure → Novel EID Reported
Exposure
Symptom presentation
Medical visit and examination
Correct diagnosis
Alert authorities
Novel EID Reported

- Cluster of cases
- Access to medical care
- Personal choices
- Medical technologies
- Uniqueness of symptoms
- Medical training
- Communications technology etc.
Exposure

Symptom presentation

Medical visit and examination

Correct diagnosis

Alert authorities

Novel EID Reported

Lag in Reporting
Lag in disease reporting

- Num. of EID outbreaks
- Num. reported
Components of Reporting Lag

**Infrastructure**
- Healthcare spending
- Number of doctors, hospitals
- Communications
- Technology

**Biological**
- Disease etiology
- Disease symptoms
- Pathogenicity
Disease Traits

**More Likely to Report**
- High transmission rate → Large clusters
- High disease severity
- Symptom abnormality

**Less Likely to Report**
- Fast recovery times
- Overlapping symptoms with commonplace diseases
Overlapping symptoms

1. Tuberculosis
2. Adenovirus
3. SARS coronavirus
4. Influenza A
5. Common cold
6. Malaria
7. Rotavirus A
8. Shigella dysenteriae
9. Escherichiae coli
10. Salmonella enteritidis
11. Giardia duodenalis
12. Ebola virus
13. Dengue
14. West Nile virus
# Scoring Symptoms

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headache, Cough, Abdominal pain, Sneezing, Back pain</td>
<td>1</td>
</tr>
<tr>
<td>Fever, Diarrhea, Localized rash, Muscle swelling, Pustules</td>
<td>2</td>
</tr>
<tr>
<td>Cardiac or pulmonary pain, Genital lesions, Abdominal mass, Vomitting, Macules, Blood in urine</td>
<td>3</td>
</tr>
<tr>
<td>Pneumonia, Gastrointestinal obstruction, Eye worm, Hemorrhagic rash, Seizures, Meningitis</td>
<td>4</td>
</tr>
<tr>
<td>Death, Hemorrhagic fever, Paralysis, Coma</td>
<td>5</td>
</tr>
</tbody>
</table>
Applying Survival Analysis to Reporting Lag

\[ h_i(t) = h_0(t)e^{\beta_1 X_{i1} + \ldots + \beta_k X_{ik}} \]
What factors contribute the lag?

- Type of disease (zoonotic, vector-borne, drug-resistance)
- Pathogen type (virus, bacteria etc.)
- Disease severity
- Population Density
- Disease Burden
- Hospital beds
- Health expenditure (as % of GDP)
- Political rights and Civil liberties
- Number of Physicians (per 1000 people)
- Immunizations
- Civilian coverage (birth and death certificates)
- Internet Users (per 100 people)
- Telephone Lines (per 100 people)
- Cell phone Subscriptions (per 100 people)
Model to account for reporting lag

<table>
<thead>
<tr>
<th>Zoonotic disease (n=158)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>( \beta )</td>
<td>s.e.(( \beta ))</td>
<td>p-value</td>
</tr>
<tr>
<td>Disease burden</td>
<td>(-1.9 \times 10^{-7})</td>
<td>(9.3 \times 10^{-8})</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>(all causes/10,000 people)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease severity score</td>
<td>.0054</td>
<td>.002</td>
<td>**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Including drug resistant diseases (n=204)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>( \beta )</td>
<td>s.e.(( \beta ))</td>
<td>p-value</td>
</tr>
<tr>
<td>Disease burden</td>
<td>(-9.6 \times 10^{-8})</td>
<td>(8.0 \times 10^{-8})</td>
<td></td>
</tr>
<tr>
<td>(all causes/10,000 people)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease severity score</td>
<td>.0018</td>
<td>.0019</td>
<td></td>
</tr>
</tbody>
</table>

Healthcare (% GDP) also significant
Conclusions

• Spatial pattern to the lag in disease reporting

• Use the lag to understand which diseases are reported in a more timely fashion.

• Lag is associated with disease types, severity of symptoms, and disease burden.

• Overlapping symptoms make novel disease detection more difficult without additional technology.
2 Final Thoughts:

• Knowing where diseases may emerge and where the gaps in disease reporting are suggests where surveillance efforts should be bolstered.

• Diseases may be observed (now or at any later time) or never observed at all. What is the true burden of EIDs?
Thank you!

Marc Levy
Susana Adamo

EcoHealth Alliance

Tiffany Bogich
Carlos Zambrana
Peter Daszak

Ruth DeFries
Kate Jones
Sebastian Funk

CIESIN

The Earth Institute

ZSL
LIVING CONSERVATION

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

USAID
PREDICT

National Science Foundation