Designing multi-objective surveillance systems

Samuel V. Scarpino
Omidyar Fellow
Santa Fe Institute

scarpino@santafe.edu
@svscarpino
scarpino.github.io
Where will the next pandemic emerge?
Where will the next **influenza** pandemic emerge?
What factors are important for influenza?

Air travel

Density of pig farms
Worldwide pig density

Pigs density map matching FAOSTAT 2005 (modelled)

AGRICULTURE AND CONSUMER PROTECTION DEPARTMENT
Animal Production and Health Division

Number per square km

- <1
- 1–5
- 5–10
- 10–20
- 20–50
- 50–100
- 100–250
- >250
- Water
- Unsuitable for ruminat

Source: Gridded Livestock of the World
Worldwide pig density

Source: Gridded Livestock of the World
and it happened again with Ebola
and it happened again with Ebola

Table I. — Age distribution of persons positive for either Lassa (LAS), Ebola (EBO) or Marburg (MAR) virus antibodies.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Nb tested</th>
<th>LAS-positive (prevalence %)</th>
<th>EBO-positive (prevalence %)</th>
<th>MAR-positive (prevalence %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>49</td>
<td>5 (10 %)</td>
<td>2 (4 %)</td>
<td>0</td>
</tr>
<tr>
<td>10-19</td>
<td>68</td>
<td>11 (16 %)</td>
<td>5 (7 %)</td>
<td>0</td>
</tr>
<tr>
<td>20-29</td>
<td>108</td>
<td>21 (19 %)</td>
<td>6 (6 %)</td>
<td>1</td>
</tr>
<tr>
<td>30-39</td>
<td>94</td>
<td>16 (17 %)</td>
<td>5 (5 %)</td>
<td>1</td>
</tr>
<tr>
<td>40-59</td>
<td>88</td>
<td>9 (10 %)</td>
<td>6 (7 %)</td>
<td>1</td>
</tr>
<tr>
<td>60 plus</td>
<td>26</td>
<td>5 (16 %)</td>
<td>2 (8 %)</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>433</td>
<td>67 (16 %)</td>
<td>26 (6 %)</td>
<td>5 (1 %)</td>
</tr>
</tbody>
</table>
Disease surveillance stands at a crossroads
Four Step Program

1. Formalize the public health objectives

2. Specify candidate data sources & acquire historical data

3. Simulate data where missing

4. Select the most informative data sources and validate!

Scarpino, Dimitrov, Meyers 2012
1. Surveillance goal - influenza in Texas
2. Data - ILINet, Hospitalizations, & G.F.T.
3. Simulate data

1. Obtain a hospitalization time series.

2. Integrate over published estimates for the hospitalization rate of influenza.

3. Account for noise.

   \[ ILI = C1 \times PerfectInformation + C0 + N(0, \sigma^2) \]

4. Account for imperfect reporting.
3. Simulate data - provider behavior
3. Simulate data - empirical reporting rates

![Graphs showing simulated data for reporting rates.](image-url)
4. Select providers

\[ R^2(Hosp, S, \hat{\xi}) = \]

- **Hosp** = goal time series (state-wide hospitalizations)
- **S** = subset of selected providers
- **\( \hat{\xi} \)** = observational noise and reporting
- **\( \xi \)** = mock provider pool
4. Select providers

\[ R^2(Hosp, S, \hat{\xi}) = \frac{\text{Var}(Hosp) - \text{Var}\left( Hosp - \sum_{i \in S} \alpha_i P_i(\hat{\xi}) \right)}{\text{Var}(Hosp)} \]

- \( Hosp \) = goal time series (state-wide hospitalizations)
- \( S \) = subset of selected providers
- \( \hat{\xi} \) = observational noise and reporting
- \( \xi \) = observational noise and reporting
- \( \alpha_i \) = regression coefficients
- \( P \) = mock provider pool
4. Select providers - account for provider behavior

\[ R^2(Hosp, S, \hat{\xi}) = \frac{\text{Var}(Hosp) - \text{Var}
\left( Hosp - \sum_{i \in S} \alpha_i P_i(\hat{\xi}) \right)}{\text{Var}(Hosp)} \]

\[ \max_{S \subseteq P} E_{\xi} \left[ R^2(Hosp, S, \hat{\xi}) \right] \]
An optimized network is more informative

Scarpino et al. 2012

Providers

Performance ($R^2$)

RSquare
Cover
CDC - Random
CDC - Greedy

Scarpino et al. 2012
An optimized network is more efficient

Obs. ILINet – 82 Providers – $R^2 \sim 0.6$

Rsquare – 38 Providers – $R^2 \sim 0.6$

Scarpino et al. 2012
But what about pandemic influenza?

Early Detection

Situational Awareness
The RSquare network doesn’t detect early cases
But a mixed network can!
And we can benchmark ILINet
And the results hold for situational awareness.
Dengue in Puerto Rico

![Graph showing dengue virus counts from 1991 to 2005. The x-axis represents years from 1991 to 2005, and the y-axis represents dengue virus counts ranging from 0 to 800. There are several peaks in the graph, indicating periods of increased dengue virus counts.]
Greedy selection providers

R^2 = 0.61

R^2 = 0.89
Island-wide incidence

![Graph showing Island-wide incidence with performance (out-of-sample $R^2$) on the y-axis and number of clinics on the x-axis. Legend includes Island (Island-wide), Regional (H.S.R.), Serotypes (Serotypes), Multi-objective (Island-wide), Multi-objective (H.S.R.), and Multi-objective (Serotypes).]
Multi-Objective Surveillance

![Performance vs. Number of Clinics](image)

- **Island (Island-wide)**
- **Regional (H.S.R.)**
- **Serotypes (Serotypes)**
- **Multi-objective (Island-wide)**
- **Multi-objective (H.S.R.)**
- **Multi-objective (Serotypes)**

Performance (out-of-sample $R^2$)

Number of Clinics
Regional surveillance

![Graph showing performance of regional surveillance with varying numbers of clinics.

Performance (out-of-sample $R^2$) is plotted against the number of clinics.

Lines represent different objectives and regions:
- Blue: Island (Island-wide)
- Grey: Regional (H.S.R.)
- Orange: Serotypes (Serotypes)
- Dashed blue: Multi-objective (Island-wide)
- Dashed grey: Multi-objective (H.S.R.)
- Dashed orange: Multi-objective (Serotypes)
Serotype surveillance

![Graph showing performance (out-of-sample $R^2$) vs. number of clinics for different categories: Island (Island-wide), Regional (H.S.R.), Serotypes (Serotypes), Multi-objective (Island-wide), Multi-objective (H.S.R.), Multi-objective (Serotypes).]
Performance without redundancy

- **Proportion of Cases Covered**
- **Number of Clinics**

Legend:
- **Green** - Multi-objective
- **Orange** - Serotype
- **Gray** - Regional
- **Blue** - Island
Performance without redundancy

Proportion of Cases Covered

Number of Clinics

- Green: Multi-objective
- Orange: Serotype
- Gray: Regional
- Purple: Island
Alternative design algorithms

Performance ($\hat{R}^2$: Island-wide incidence)
Alternative design algorithms

$\hat{R}^2$: Island-wide incidence

Performance

<table>
<thead>
<tr>
<th>Island</th>
<th>Multi-objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Alternative design algorithms

\( \hat{R}^2 \): Island-wide incidence
Alternative design algorithms

Performance ($\hat{R}^2$: Island-wide incidence)

- Island
- Multi-objective
- Diversity
- Regional
Alternative design algorithms

\[ R^2: \text{Island-wide incidence} \]

Performance

- Island
- Multi-objective Diversity
- Regional
- Serotype
Alternative design algorithms

Performance ($R^2$: Island-wide incidence)

- Island
- Multi-objective
- Diversity
- Regional
- Serotype
- Volume
Alternative design algorithms

Performance

\( R^2: \) Island-wide incidence

- Island
- Multi-objective
- Diversity
- Regional
- Serotype
- Volume
- Population
Similar pattern for the other objectives

Performance

(\hat{R}^2: Island-wide incidence)

- Island
- Multi-objective
- Diversity
- Regional
- Serotype
- Volume
- Population

(\hat{R}^2: Serotype-specific incidence)

- Serotype
- Multi-objective
- Diversity
- Island
- Regional
- Volume
- Population

(\hat{R}^2: H.S.R. incidence)

- Regional
- Island
- Diversity
- Serotype
- Volume
- Population
Hidden validation

A.) Island-wide incidence

- **Observed**
- **Validation estimate**
Chikungunya Surveillance
Texas Arbovirus Risk Map

Welcome to the Texas Arbovirus Risk Map. Click on a category below to explore data associated with mosquito-borne diseases in Texas.

- Risk Model
- Dengue and Chikungunya risk maps for Texas. There are two distinct types of risk: Import Risk, and Sustained Transmission Risk.
- Import Risk
- Log Import Risk
- Sustained Transmission Risk

- Suitability
- Environmental
- Socio-Economic
- Transportation
Four Step Program

1. Formalize the public health objectives

2. Specify candidate data sources & acquire historical data

3. Simulate data where missing

4. Select the most informative data sources and validate!

Scarpino, Dimitrov, Meyers 2012
Acknowledgments

Ned Dimitrov
Lauren Ancel Meyers
Michael Johanssson
Bruce Clements
Lyn Finelli
Alison Galvani

NIH MIDAS
DTRA
The Santa Fe Institute
The Omidyar Group
Questions?

Samuel V. Scarpino
Omidyar Fellow
Santa Fe Institute

scarpino@santafe.edu
@svscarpino
scarpino.github.io