USING SOCIAL MEDIA TO STUDY PUBLIC HEALTH

MICHAEL PAUL
JOHNS HOPKINS UNIVERSITY
MAY 29, 2014
WHAT IS PUBLIC HEALTH?

“preventing disease, prolonging life and promoting health”
WHAT IS PUBLIC HEALTH?

“preventing disease, prolonging life and promoting health”
WHAT IS PUBLIC HEALTH?

“preventing disease, prolonging life and promoting health”

• Disease surveillance
• Vaccination
• Behavioral risk factors
• Environmental effects
• Quality and safety
• … and much more
WHAT IS PUBLIC HEALTH?

Population

Surveillance

Intervention

Doctors
TECHNOLOGY AND PUBLIC HEALTH

Surveillance

Population

Doctors

Intervention

social media campaigns apps for health mgmt.
TECHNOLOGY AND PUBLIC HEALTH

Population

Surveillance

electronic records
web news/media
crowdsourcing

Doctors

Intervention
Computational epidemiology is a multidisciplinary field that uses techniques from computer science, mathematics, geographic information science and public health to better understand issues central to epidemiology such as the spread of diseases or the effectiveness of a public health intervention.
DISEASE SURVEILLANCE

Monitoring the world to identify outbreaks and understand prevalence of diseases

Many resources are spent on influenza surveillance.

Important for:

• hospital preparations
• resource allocation
• targeted awareness/prevention
WHY FLU?

• It’s common
  • occurs seasonally

• It’s burdensome
  • takes a toll economically and mentally

• It’s dangerous
  • ~30,000 deaths per year
WHY FLU?

images from http://nyamcenterforhistory.org/tag/spanish-flu/
WHY FLU?
INFLUENZA SURVEILLANCE

What strains of the virus are out there?
When is the flu going to arrive/leave?
Where has the flu broken out?

Surveillance has improved since 1918…
INFLUENZA SURVEILLANCE

Pyramid of influenza patients:

- deaths
- hospitalized / tested positive
- doctor visit (symptomatic)
- symptomatic (but no doctor visit)
- infected but asymptomatic
INFLUENZA SURVEILLANCE

Pyramid of influenza patients:

- deaths
- hospitalized / tested positive
- doctor visit (symptomatic)

Monitored in depth by US government

- Centers for Disease Control and Prevention (CDC)
INFLUENZA SURVEILLANCE

Monitored in depth by US government

- Centers for Disease Control and Prevention (CDC)
How can technology help?
INFLUENZA SURVEILLANCE

How can technology help?

Electronic health records

• Reliable, accurate data
• Easy to access
• Fast to synchronize
• … but not well implemented yet
INFLUENZA SURVEILLANCE

How can technology help?

Internet data

• Search queries

  ![Google search results for "how long does the flu last"]

• Social media

  ![Twitter post: "So Sick :( Flu Is Killing Me!!!!! *Cough Cough*" with timestamp 5:24 PM - 28 May 2014]
INFLUENZA SURVEILLANCE

How can technology help?

Internet data

• Faster than traditional surveillance
  • but less accurate, so there’s a tradeoff
• Very cheap to obtain
• Information outside the hospital system
INFLUENZA SURVEILLANCE

How can technology help?

electronic records (eventually)

hospitalized / tested positive

doctor visit (symptomatic)

symptomatic (but no doctor visit)

infected but asymptomatic

web news/media crowdsourcing
INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

- Symptomatic (but no doctor visit)
- Doctor visit (symptomatic)
- Hospitalized / tested positive
- Deaths

Time
INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

virologic surveillance
- historic gold standard
- slow – relies on lab tests

hospitalized / tested positive

deads

Time
INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

- **syndromic surveillance**
  - look for symptoms without testing
  - faster – newer idea

**Time**
INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

not part of traditional surveillance
• can only get from surveys, self-reports
• potentially fastest indication of infection

symptomatic (but no doctor visit)

Time
INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

**not part of traditional surveillance**
- can only get from surveys, self-reports
- potentially fastest indication of infection

symptomatic (but no doctor visit)

We can get this instantly from social media!
SOCIAL MEDIA

Twitter is often used for research

Users share short “status messages” about what they’re doing and thinking

• 500 million tweets daily
• generally public
• used worldwide

visualization by Miguel Rios
FLU TRACKING
WITH SOCIAL MEDIA

Real-time tracking through Twitter

Use algorithms to aggregate tweets describing flu symptoms

• this gives us daily estimates of flu prevalence
FLU TRACKING
WITH SOCIAL MEDIA

![Graph showing flu tracking with CDC and Twitter data over time.]

- **CDC**
- **Twitter**
FLU TRACKING WITH SOCIAL MEDIA
Forecasting the flu

- where will it be in the future?

Twitter reduces forecasting error up to 5 weeks by 20%
Tracking behavior

How will people behave during a pandemic?

Hard to answer

• can social media provide insights?
FLU TRACKING WITH SOCIAL MEDIA

Novel outbreaks, early detection:

from Salathé et al. NEJM 2013.
What else do people talk about?

A lot, we discovered
PUBLIC HEALTH ON SOCIAL MEDIA

Aches and Pains
PUBLIC HEALTH ON SOCIAL MEDIA

Insomnia
PUBLIC HEALTH ON SOCIAL MEDIA

Allergies
PUBLIC HEALTH ON SOCIAL MEDIA

Allergies

Tweets match the temporal trends of existing surveys:
PUBLIC HEALTH ON SOCIAL MEDIA

Allergies

April 2010
PUBLIC HEALTH ON SOCIAL MEDIA

Social media in China

Several health topics discussed, including:

<table>
<thead>
<tr>
<th>Nutrition</th>
<th>Pregnancy</th>
<th>Skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>食物 (food)</td>
<td>孕期 (pregnancy)</td>
<td>皮肤 (skin)</td>
</tr>
<tr>
<td>(wait)</td>
<td>孕妇 (pregnant woman)</td>
<td>肌肤 (skin)</td>
</tr>
<tr>
<td>水果 (fruit)</td>
<td>妊娠 (gestation)</td>
<td>面膜 (facial mask)</td>
</tr>
<tr>
<td>维生 (vitamins)</td>
<td>孕妇 (pregnant)</td>
<td>可以 (can)</td>
</tr>
<tr>
<td>空腹 (empty stomach)</td>
<td>妊娠 (gestation)</td>
<td>使用 (use)</td>
</tr>
<tr>
<td>营养 (nutrition)</td>
<td>妊娠 (pregnant)</td>
<td>毛孔 (pore)</td>
</tr>
<tr>
<td>健康 (health)</td>
<td>妈妈 (mother)</td>
<td>保湿 (moisture)</td>
</tr>
<tr>
<td>可以 (can)</td>
<td>周 (week)</td>
<td>芦荟 (aloe)</td>
</tr>
<tr>
<td>消化 (digestion)</td>
<td>发育 (develop)</td>
<td>护肤 (skin protection)</td>
</tr>
<tr>
<td>便秘 (constipation)</td>
<td>海带 (kelp)</td>
<td>化妆 (makeup)</td>
</tr>
</tbody>
</table>
PUBLIC HEALTH ON SOCIAL MEDIA

Social media in China

Social media again tracks government data for influenza:

![Graph showing influenza cases tracked by CCDC and Weibo over time.](image)
PUBLIC HEALTH ON SOCIAL MEDIA

Social media in China

We also found high correlations between social media messages and sensor data for air pollution.
HEALTHCARE QUALITY & SAFETY

The web makes it easier to find the right doctor

Top Doctors - Washingtonian Magazine
www.washingtonian.com/healthcare/ The Washingtonian
Find a Healthcare Provider Searchable Healthcare Guide for DC, Maryland and ... The 2014 Top Doctors will be added to this database no later than March 20.

Best Rated Doctors near Washington, DC - RateMDs.com
www.ratemd.com/best-doctors/DC/Washington
Best Rated Doctors near Washington, DC. These Top 10 lists are generated based on a weighting of the doctor's average overall rating score, along with the ...

All Specialties in Washington | US News Doctors
health.usnews.com > Doctors U.S. News & World Report
Doctors in Washington. Many doctors specialize in a particular area of medicine. Please select a ... Best Hospitals badge on your website and ads. Learn about ...

Best doctors Washington, DC - Yelp
www.yelp.com/search?find_desc=best+doctors&find...DC Yelp, Inc.
Reviews on Best doctors in Washington, DC One Medical Group, Andrew T Lee, DO, Lisa Kaufman, MD, One Medical Group, Dr. Dale Matthews, MD, Max ...
We analyzed over 50,000 doctor reviews from 50 states.

The reviews were correlated with existing metrics of provider quality:

- The reviews contain real information.
We also searched Twitter for **medical mistakes**

- Public health concern
- Not well studied (relatively)
- Errors not always reported

Social media contains informal reports of mistakes
HEALTHCARE QUALITY & SAFETY

1,000 tweets analyzed:

<table>
<thead>
<tr>
<th>Error Source</th>
<th>Error Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon</td>
<td>Surgical</td>
<td>18.1%</td>
</tr>
<tr>
<td>Physician</td>
<td>Diagnostic</td>
<td>51.5%</td>
</tr>
<tr>
<td>Nurse</td>
<td>Medication</td>
<td>18.7%</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>Procedure</td>
<td>0.0%</td>
</tr>
<tr>
<td>Other Medical</td>
<td>Infection</td>
<td>4.7%</td>
</tr>
<tr>
<td>Dentist</td>
<td>Communication</td>
<td>0.0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>7.0%</td>
</tr>
</tbody>
</table>

because the doctor screwed up a test .. I had to have ER surgery..was stabbed internally
What is Public Health?

All of these have been studied in social media!

I talked about just a few of them...

- Disease surveillance
- Vaccination
- Behavioral risk factors
- Environmental effects
- Quality and safety
- ... and much more
QUESTIONS?
ACKNOWLEDGMENTS

Slides/ideas:
Mark Dredze (JHU, advisor)
Lyn Finelli (CDC)

Collaborators:
David Broniatowski, Alex Lamb (flu on Twitter)
Byron Wallace, Urmimala Sarkar, Tom Trikalinos (doctor reviews)
Atul Nakhasi, Ralph Passarella, Sarah Bell (patient safety)
Shiliang Wang (Weibo)
Michael Smith (behavioral response)