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"
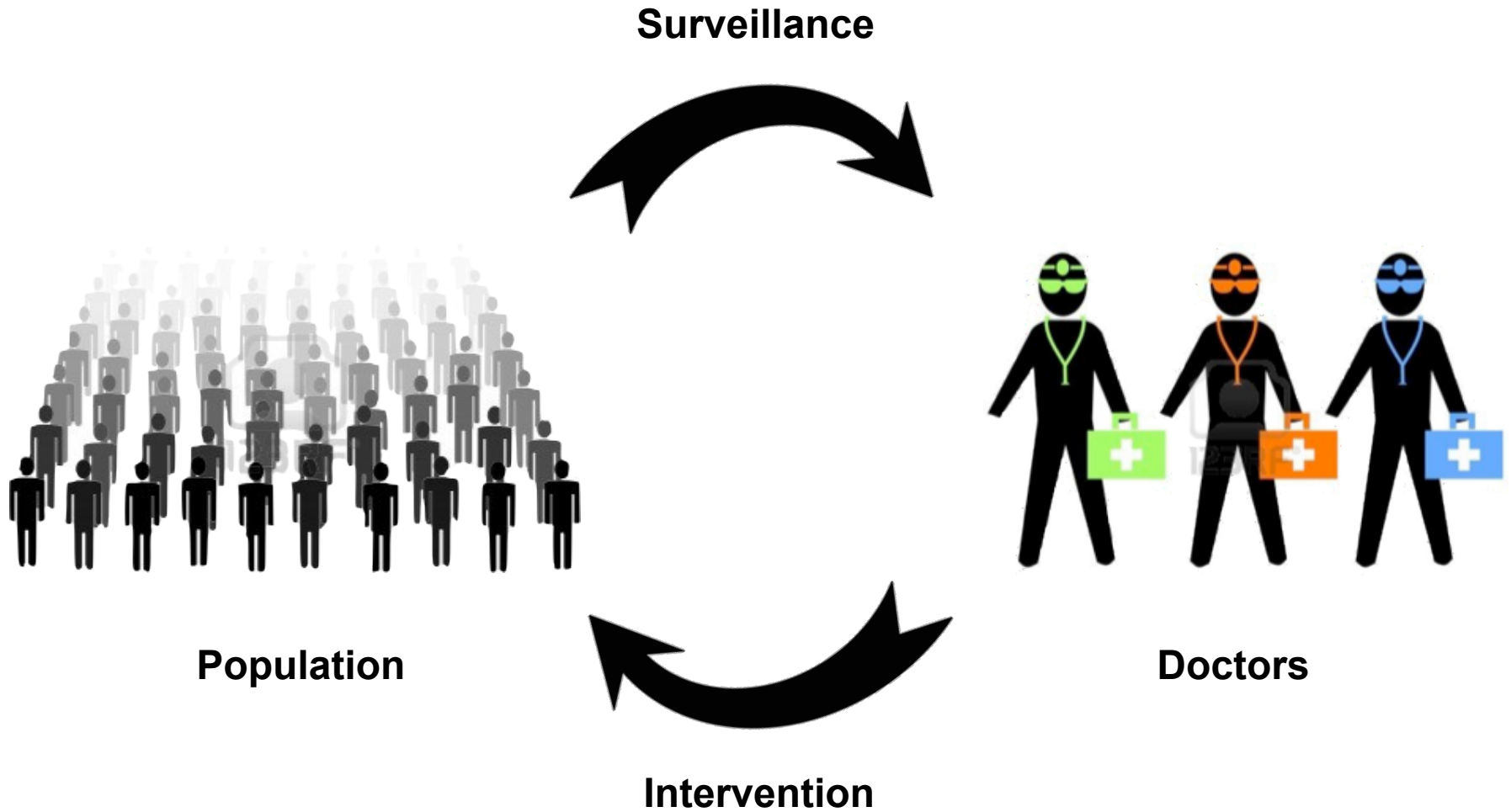

Population   **+**   **Doctors**

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?

Surveillance

Population

Doctors

Intervention

TECHNOLOGY AND PUBLIC HEALTH

Surveillance

Population

Doctors

Intervention

social media campaigns apps for health mgmt.

TECHNOLOGY AND PUBLIC HEALTH

Surveillance

electronic records
web news/media
crowdsourcing

Population

Doctors

Intervention

TECHNOLOGY AND PUBLIC HEALTH

Article | Talk   Read | Edit | View history   Search

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Computational epidemiology

From Wikipedia, the free encyclopedia

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?

'OPEN-FACE' SNEEZERS TO BE ARRESTED

Orders to arrest any one indulging in the 'open face' sneeze or cough in ...

FLU CURFEW TO SOUND FOR CITY SATURDAY NIGHT

Persons Not on Business Expected to Quit the Streets at 9 o'Clock.

The curfew will ring or, rather, blow in Chicago tomorrow night. Promptly at 9 o'clock the whistles ...

'NONESSENTIAL' CROWDS BARRED IN EPIDEMIC WAR

Churches and Saloons Exempt; Conventions, Athletics, Parties Hit.

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?**
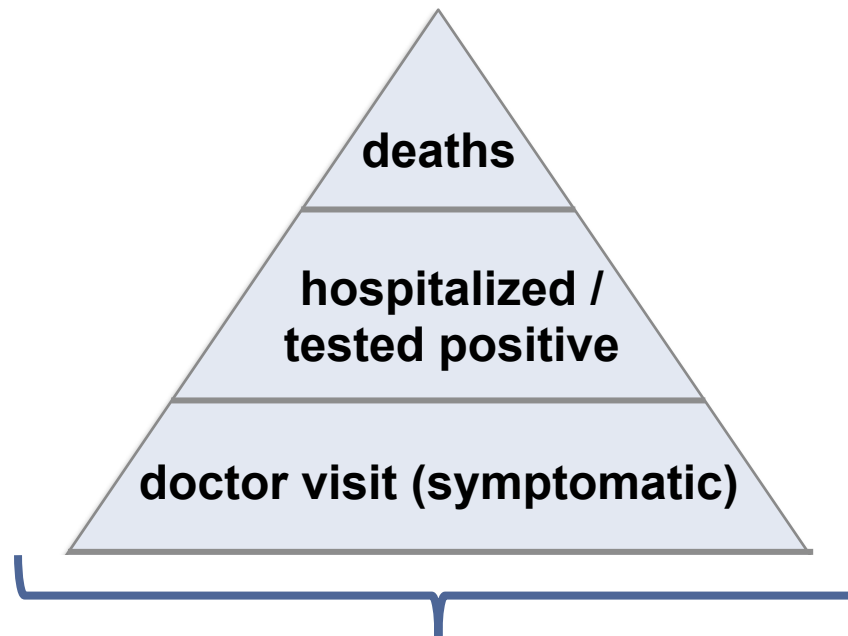
Surveillance has improved since 1918…
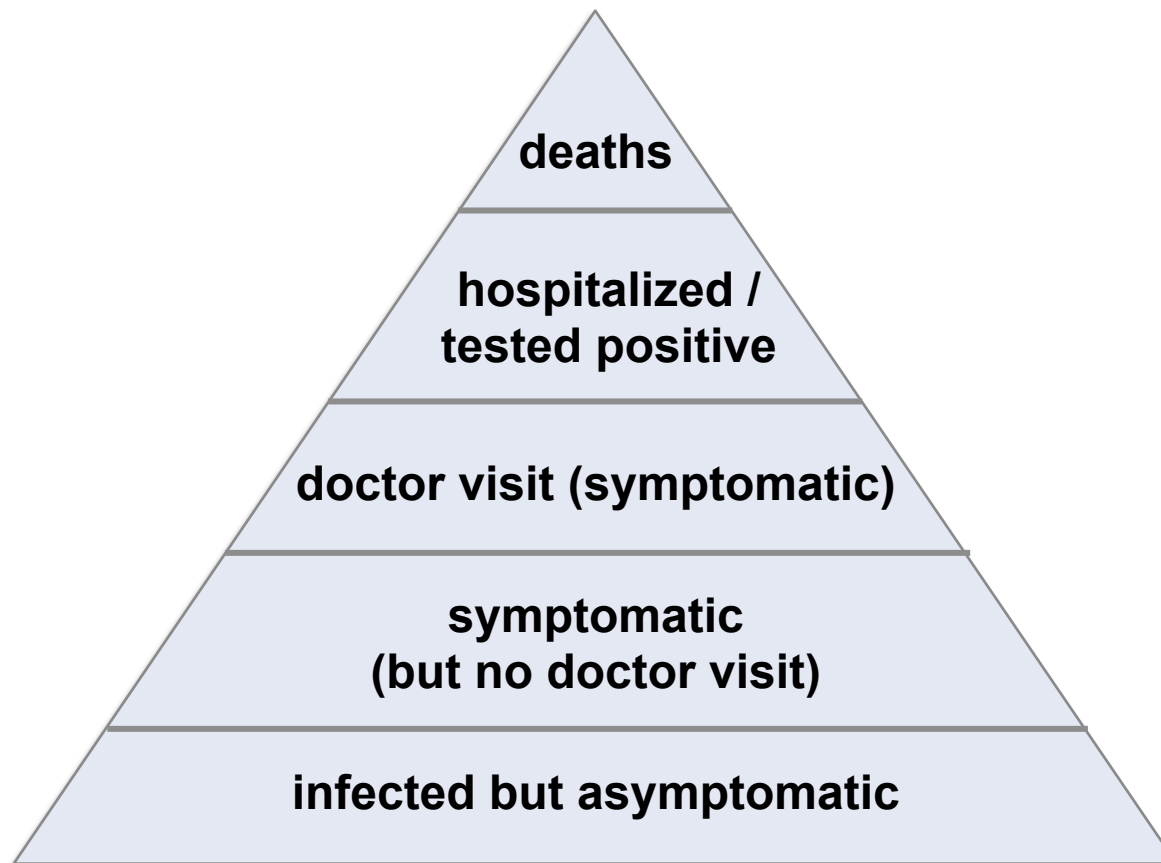
Dover In Grip of Sweeping Epidemic of Influenza

Largest Death Rate For Week In Memory of Oldest Inhabitant-- New Cases Reported Today

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
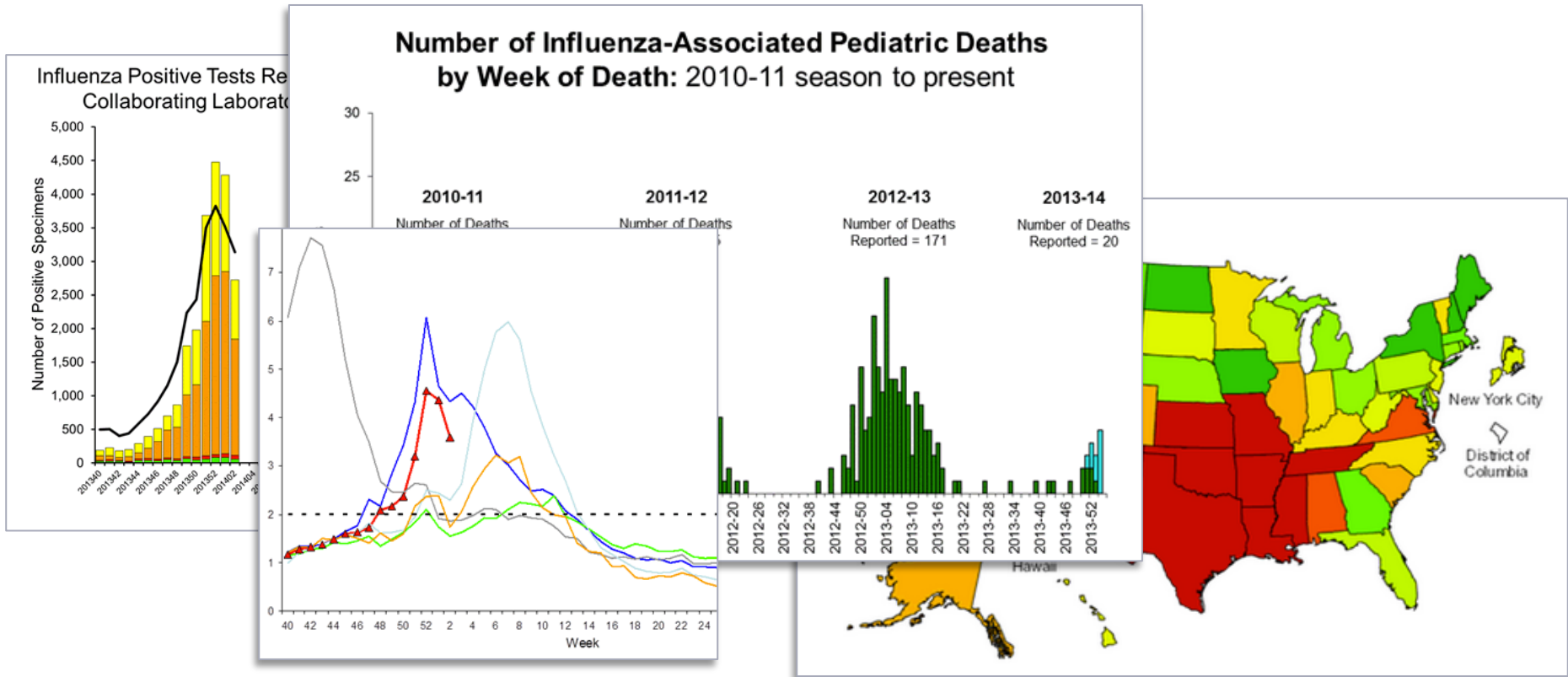
Monitored in depth by US government

- **Centers for Disease Control and Prevention (CDC)**

INFLUENZA SURVEILLANCE

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
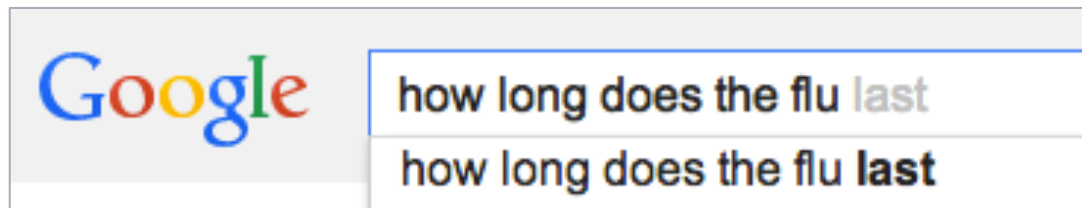
• Social media
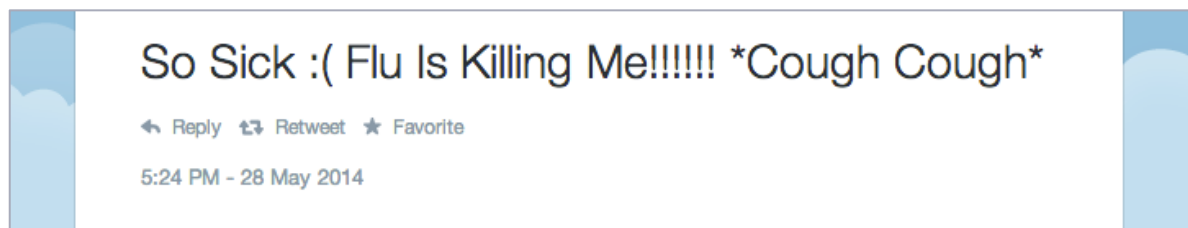
INFLUENZA SURVEILLANCE

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?
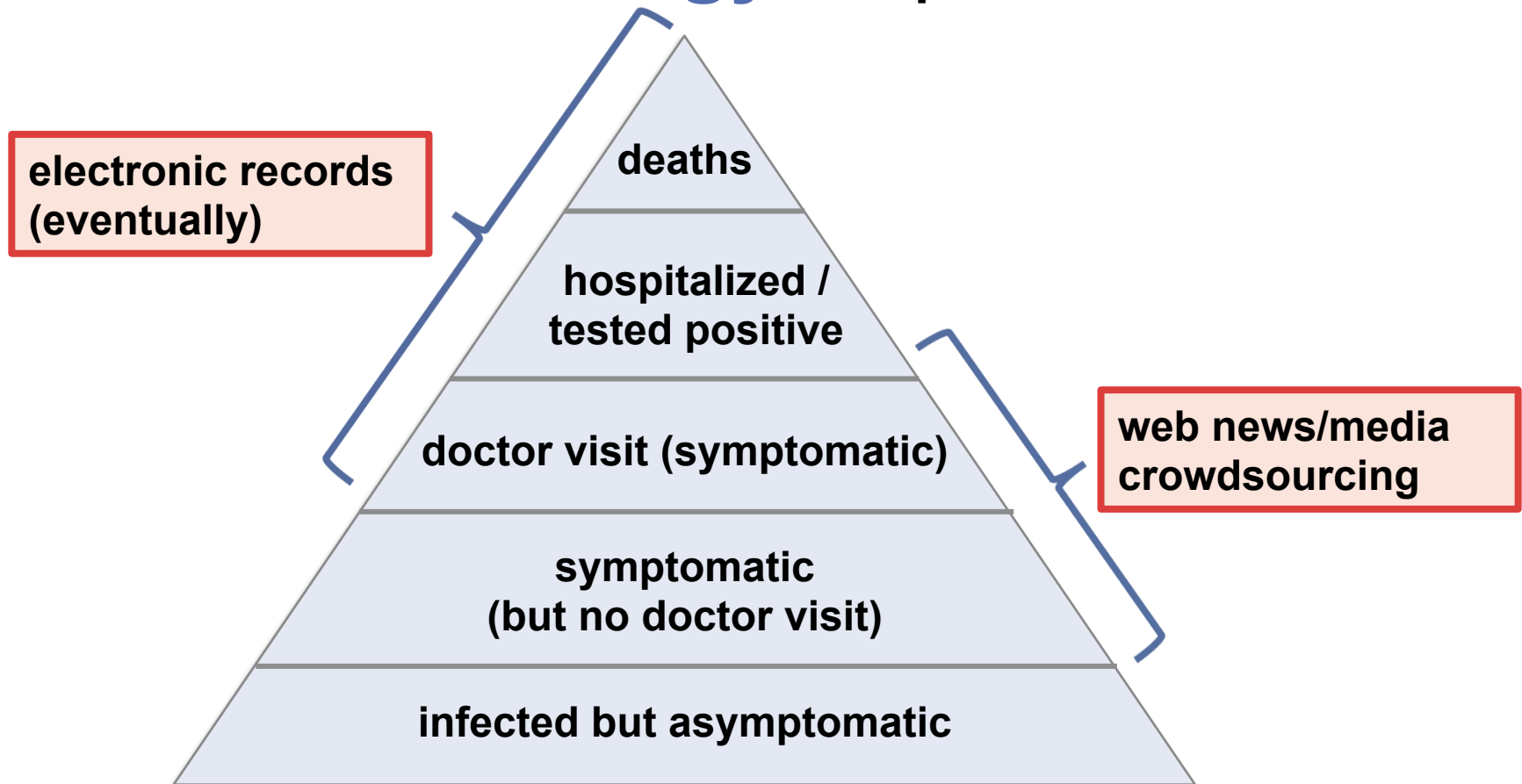
INFLUENZA SURVEILLANCE

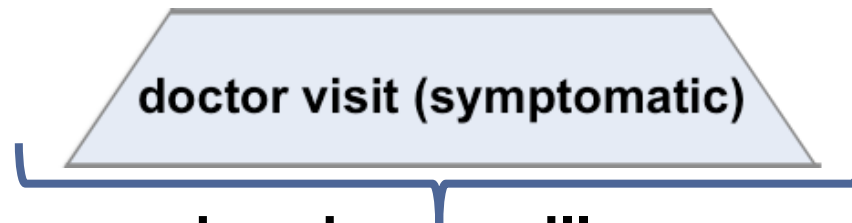**The pyramid is also a timeline:**

INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

deaths

hospitalized /
tested positive
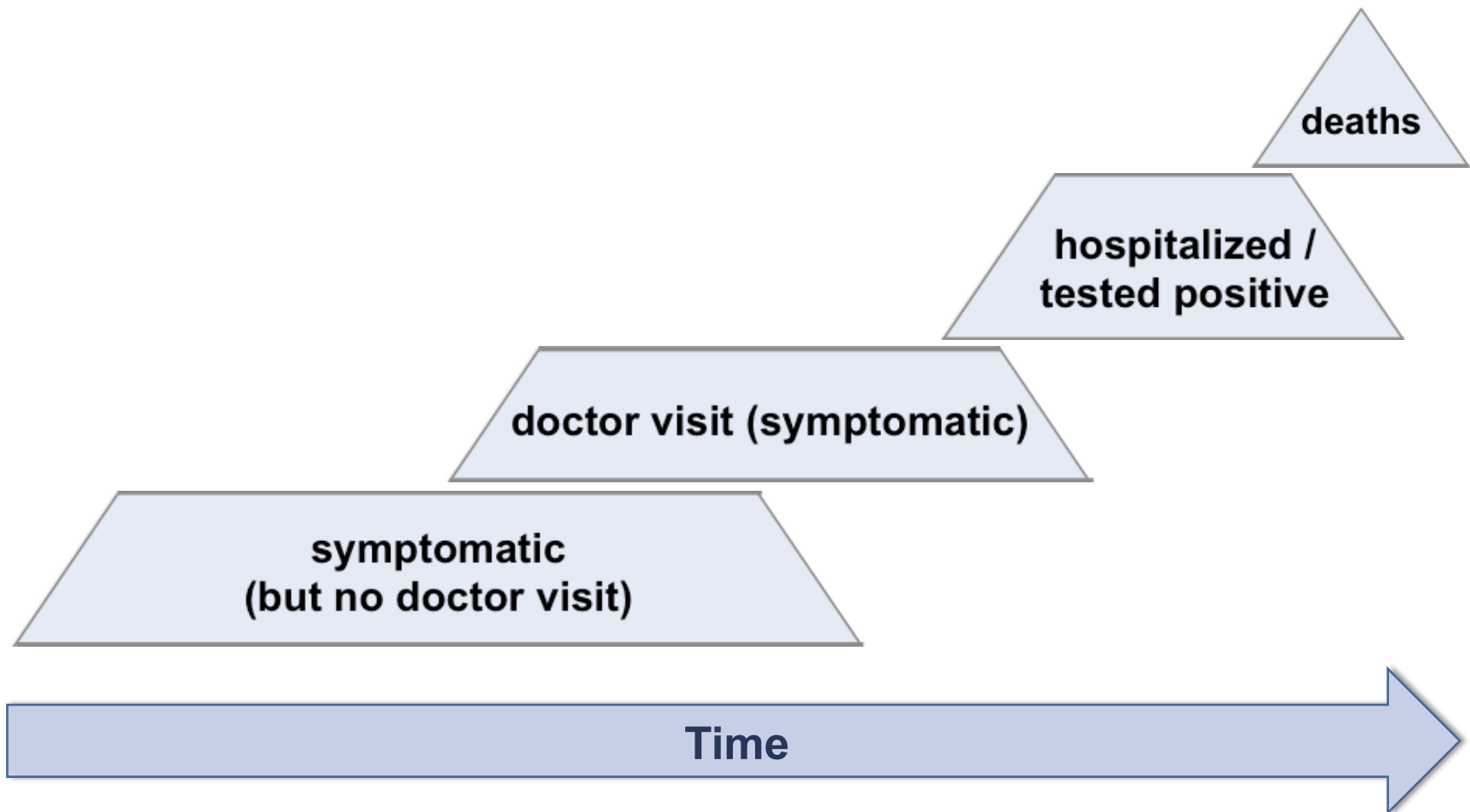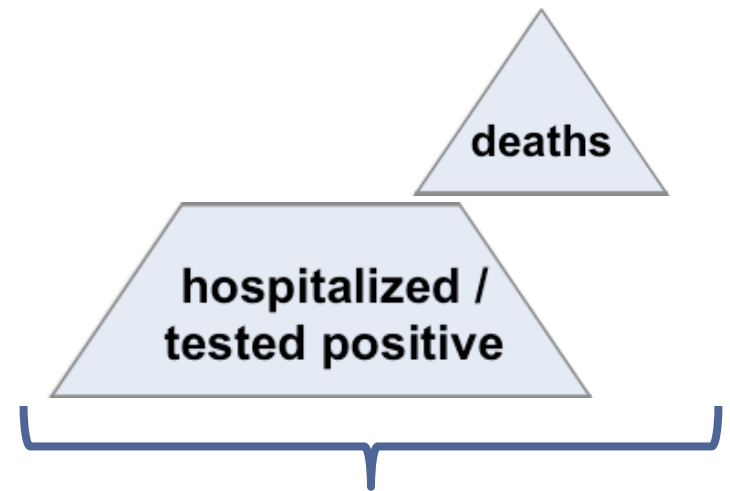
virologic surveillance
- historic gold standard
- slow – relies on lab tests

Time

INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

doctor visit (symptomatic)

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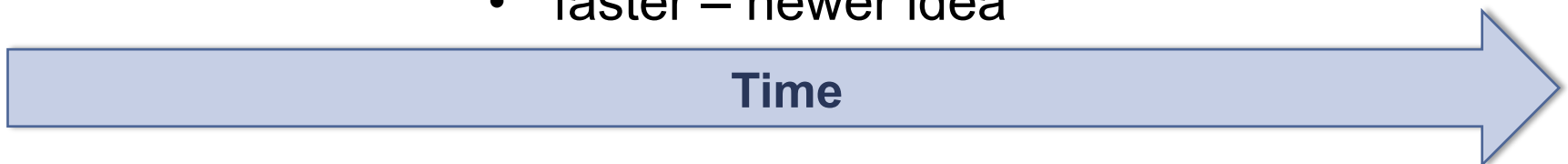
Time

INFLUENZA SURVEILLANCE

The pyramid is also a timeline:

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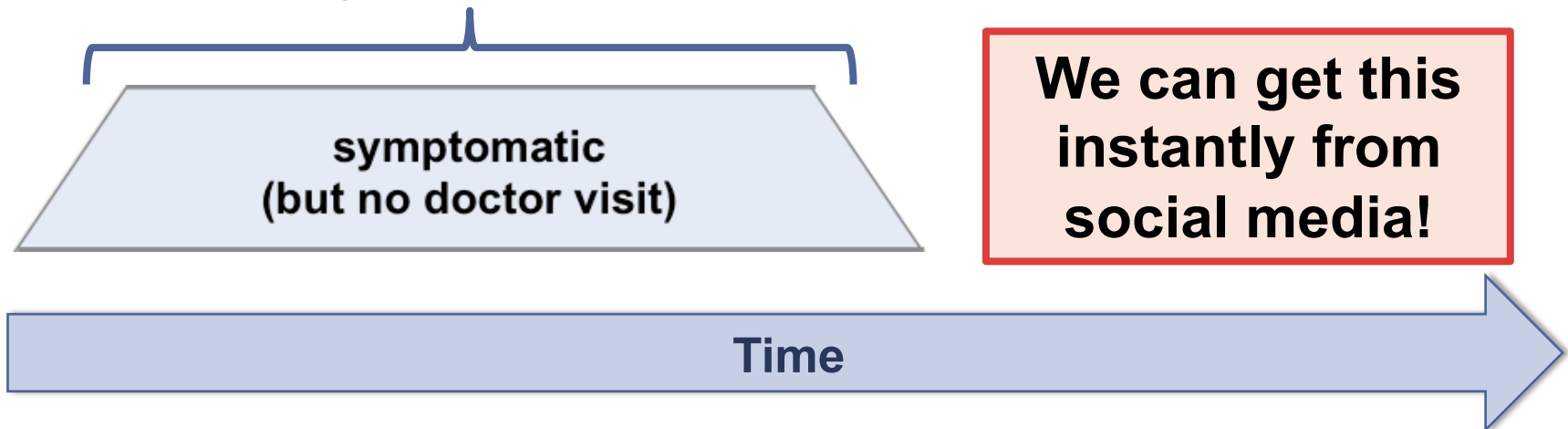- potentially fastest indication of infection

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!
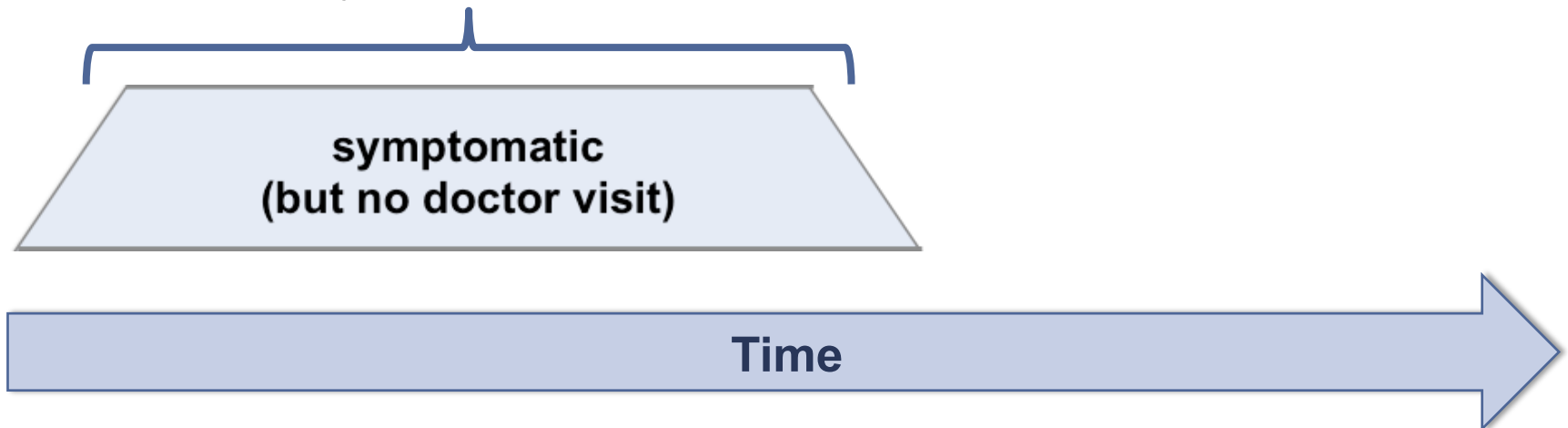
Time

SOCIAL MEDIA

Twitter is often used for research

Users share short "status messages" about what they're doing and thinking

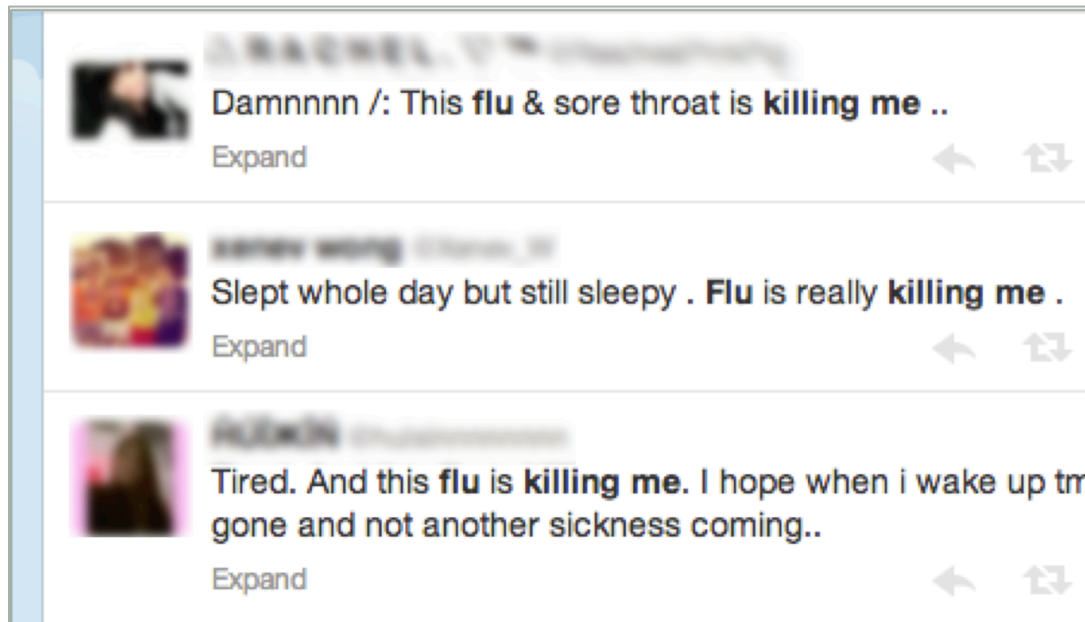- 500 million tweets daily

- generally public

- used worldwide

Support for #France and #England during their #Euro2012 match on June 11, 2012.   #ENGFRA

visualization by Miguel Rios

FLU TRACKING WITH SOCIAL MEDIA

Real-time tracking through Twitter

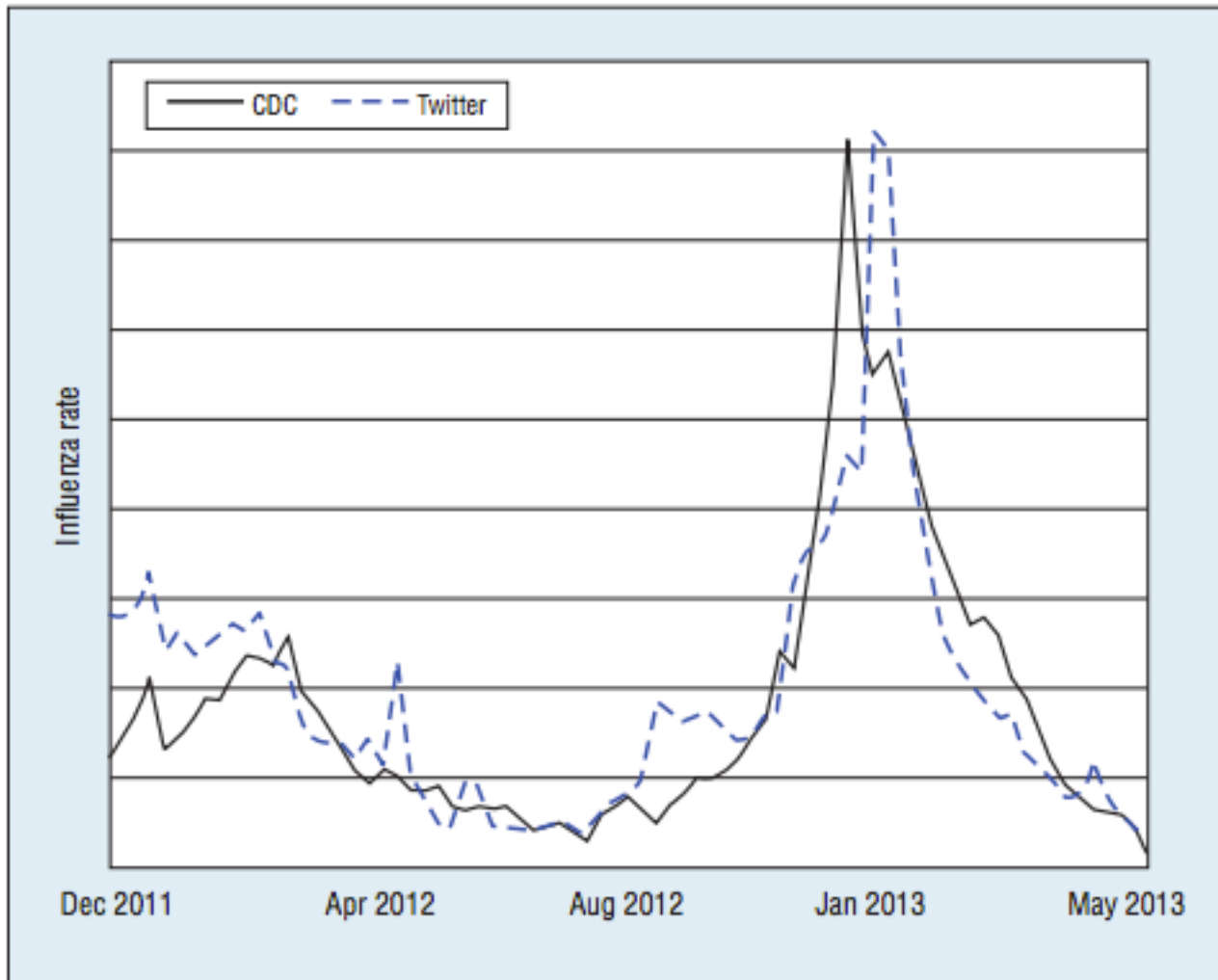Use algorithms to aggregate tweets describing flu symptoms

- **this gives us daily estimates of flu prevalence**

Date	Count
12/30/12	1630
12/31/12	1826
1/1/13	1578
1/2/13	1924
1/3/13	2167
1/4/13	2316
1/5/13	2133
1/6/13	2076
1/7/13	2441
1/8/13	2588
1/9/13	3536
1/10/13	4123
1/11/13	4627
1/12/13	3716
1/13/13	3263
1/14/13	3811

FLU TRACKING WITH SOCIAL MEDIA
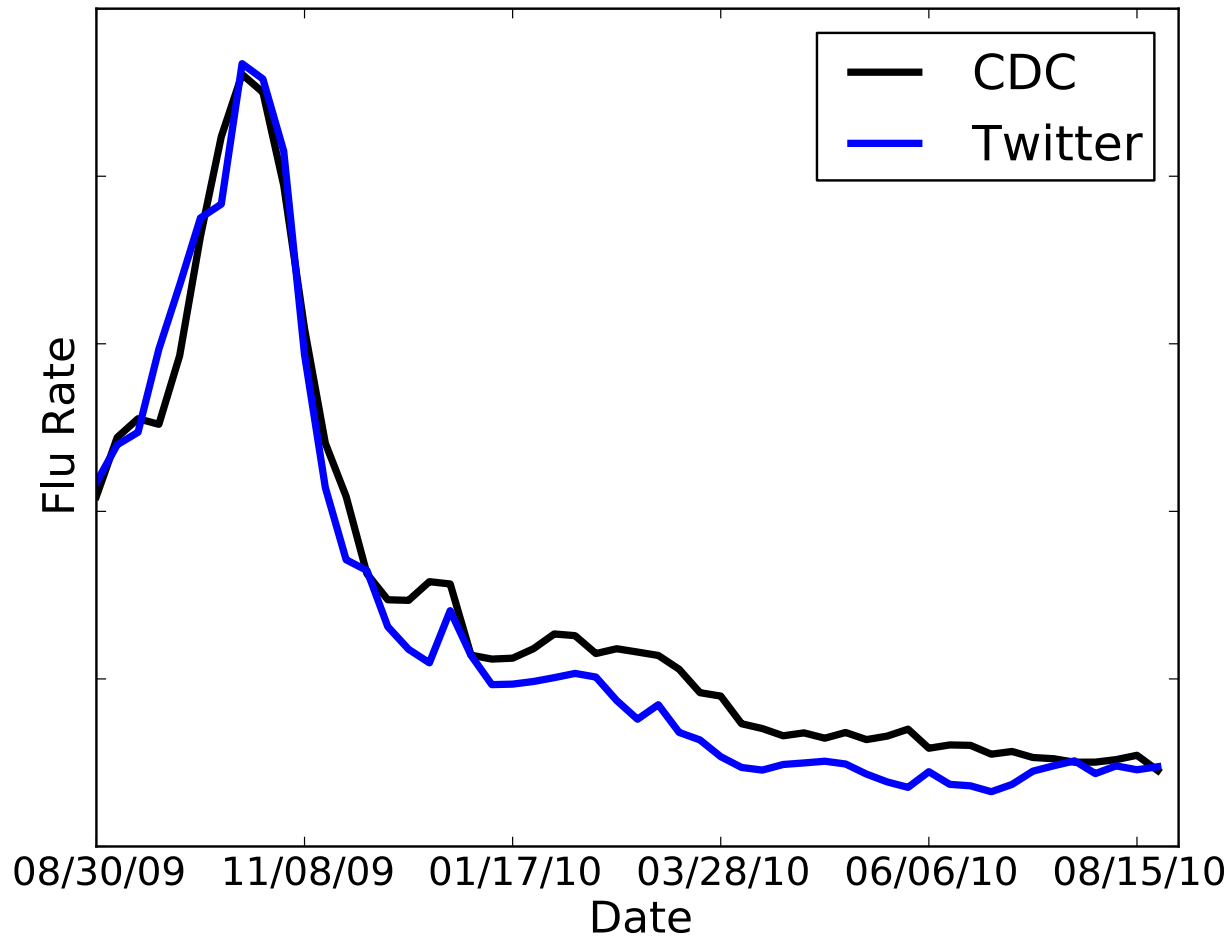
FLU TRACKING WITH SOCIAL MEDIA

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%
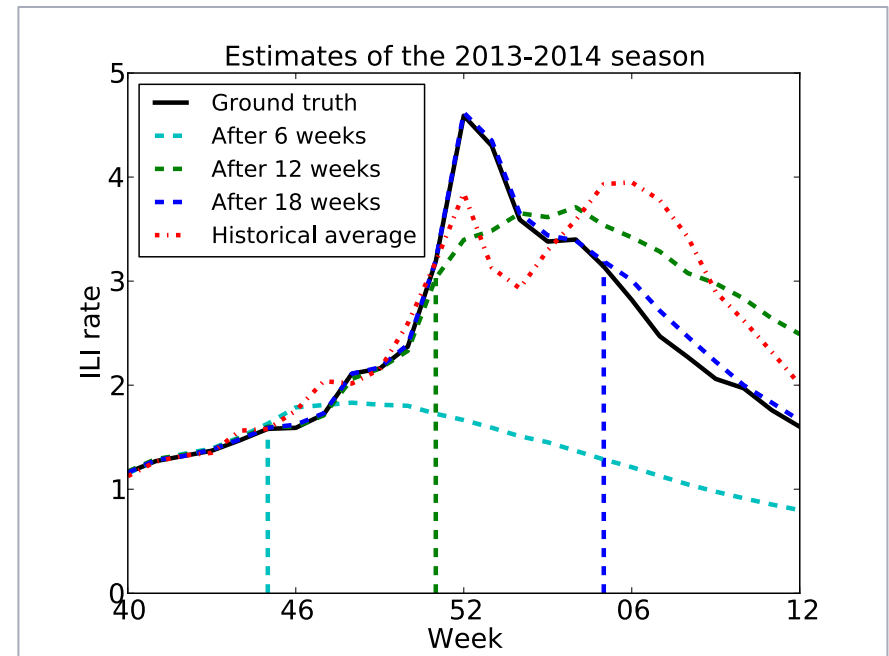
FLU TRACKING WITH SOCIAL MEDIA

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:
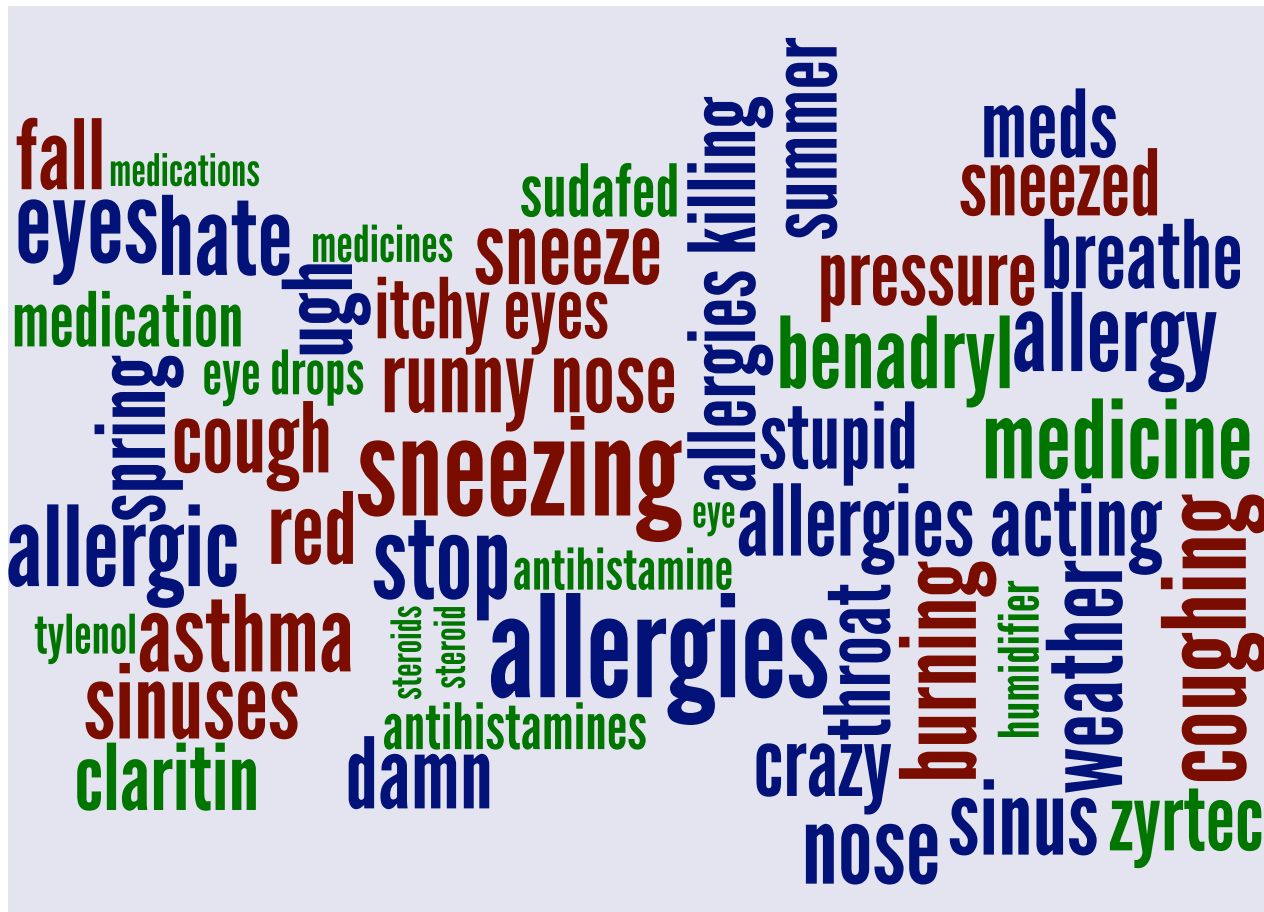

from Salathé et al. NEJM 2013.

BEYOND FLU

What else do people talk about?

A lot, we discovered

PUBLIC HEALTH ON SOCIAL MEDIA

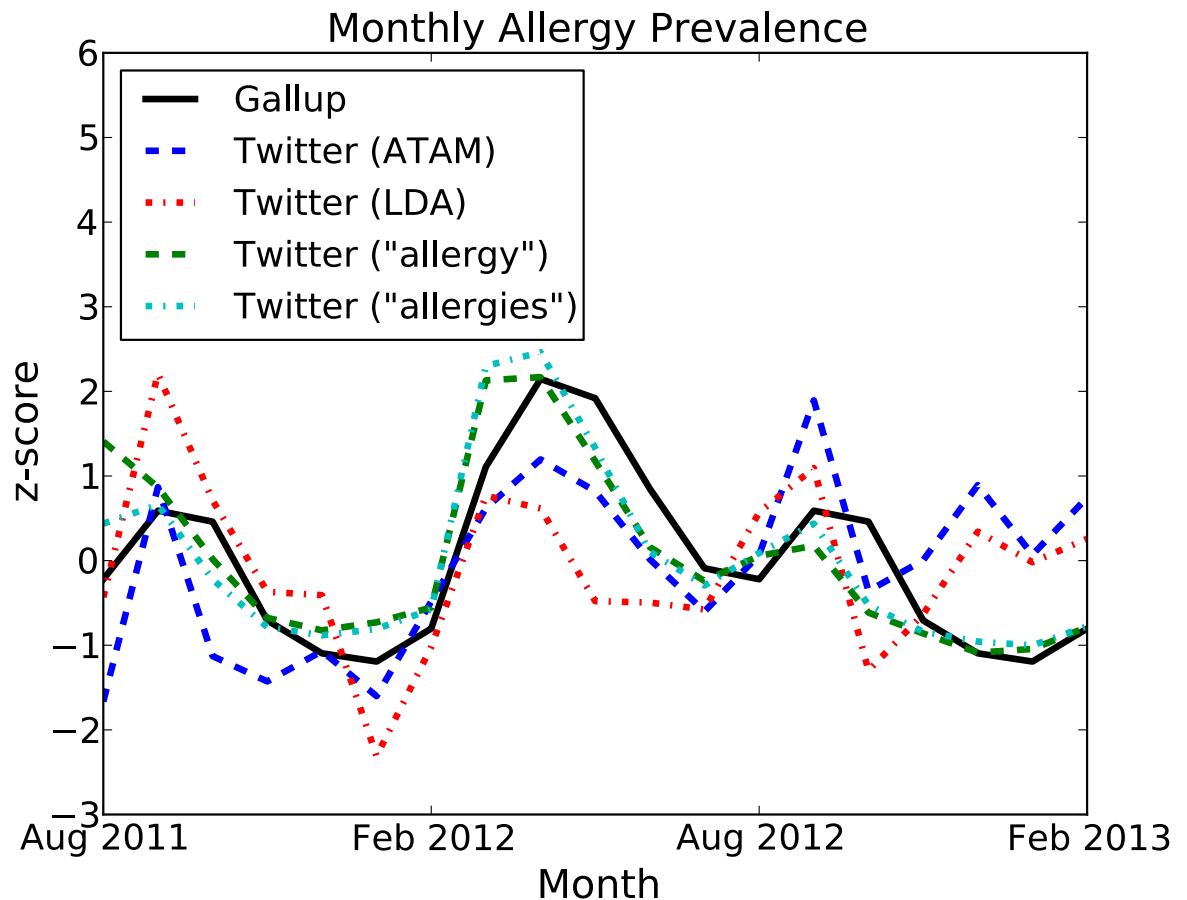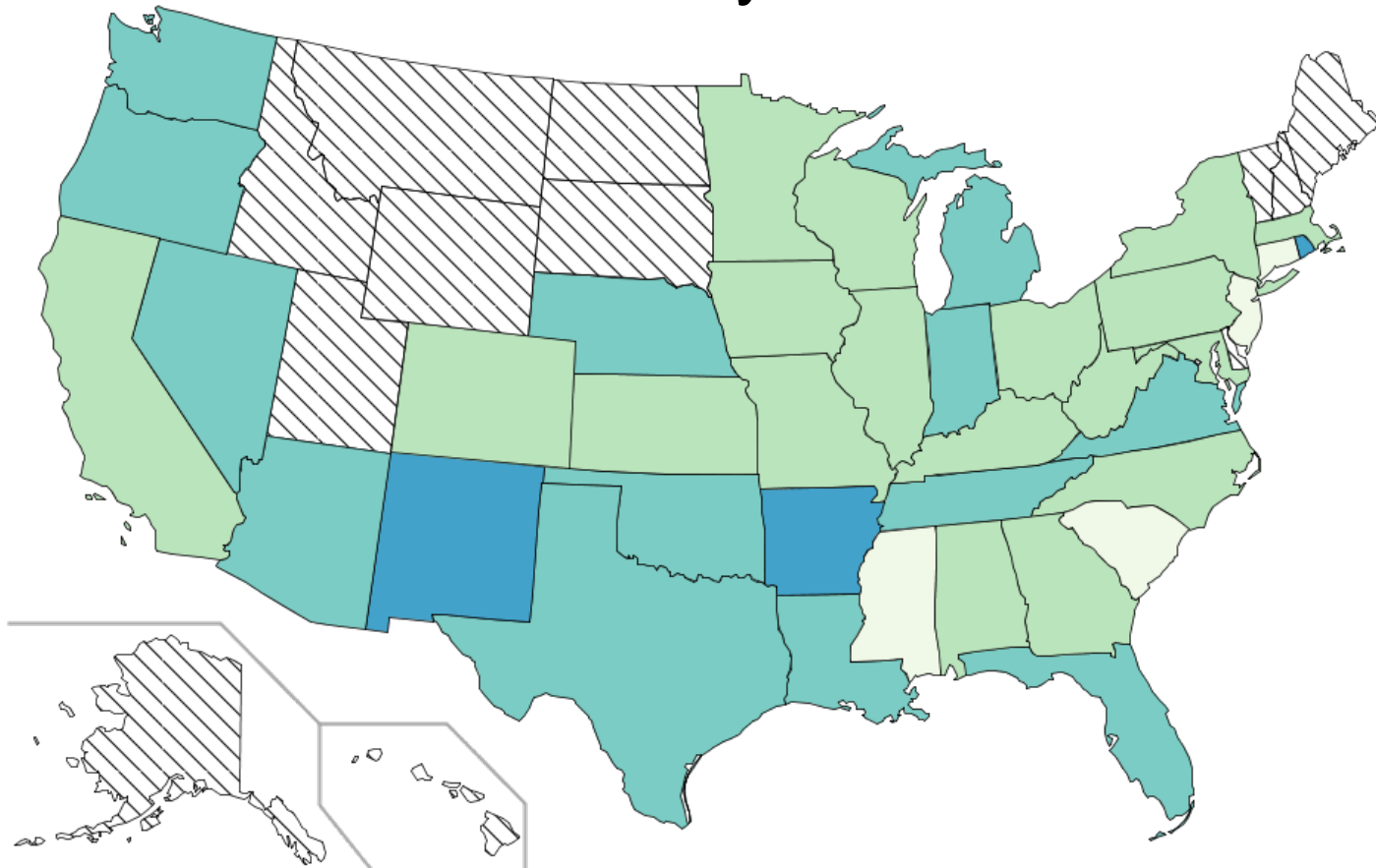## Aches and Pains

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:

Monthly Allergy Prevalence

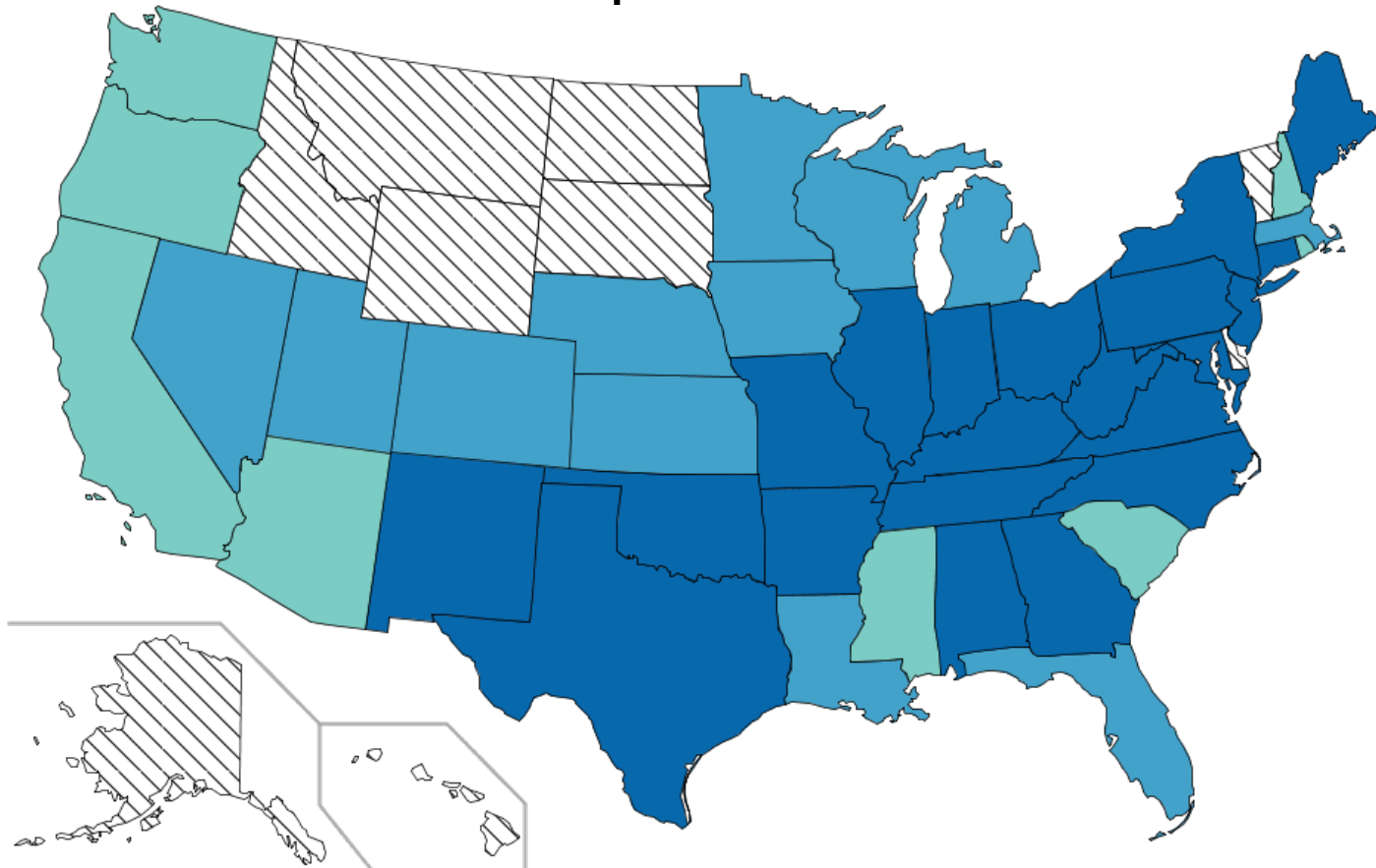# PUBLIC HEALTH ON SOCIAL MEDIA

Allergies

February 2010

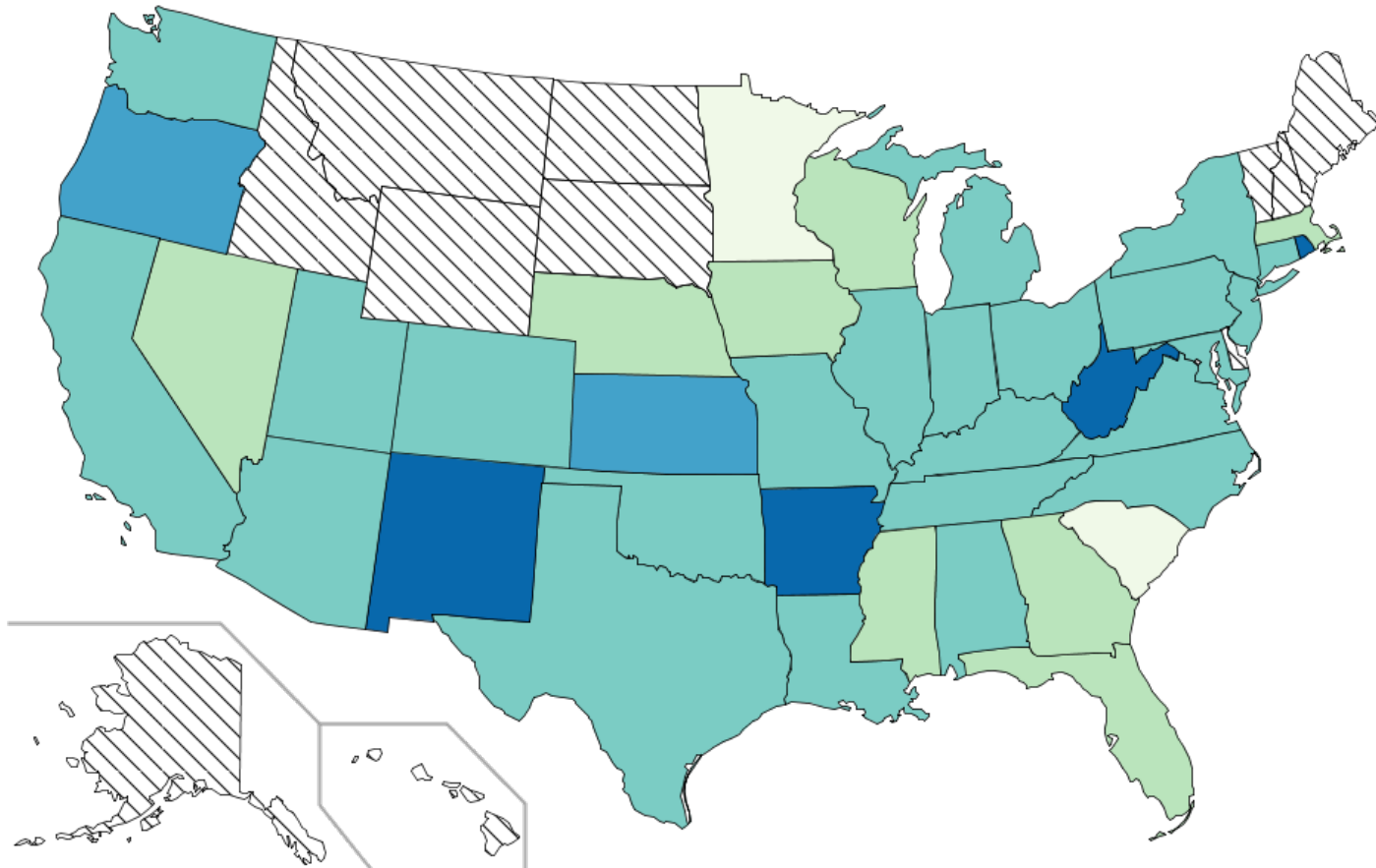# PUBLIC HEALTH ON SOCIAL MEDIA

Allergies

April 2010

PUBLIC HEALTH ON SOCIAL MEDIA

Allergies

June 2010

PUBLIC HEALTH ON SOCIAL MEDIA

Social media in China

Several health topics discussed, including:

Nutrition		Pregnancy		Skin	
食物	(food)	孕期	(pregnancy)	皮肤	(skin)
等	(wait)	孕妇	(pregnant woman)	肌肤	(skin)
水果	(fruit)	胎儿	(fetus)	面膜	(facial mask)
维生	(vitamins)	妊娠	(gestation)	可以	(can)
空腹	(empty stomach)	怀孕	(pregnant)	使用	(use)
营养	(nutrition)	妈妈	(mother)	毛孔	(pore)
健康	(health)	周	(week)	保湿	(moisture)
可以	(can)	发育	(develop)	芦荟	(aloe)
消化	(digestion)	海带	(kelp)	护肤	(skin protection)
便秘	(constipation)	影响	(influence)	化妆	(makeup)

PUBLIC HEALTH ON SOCIAL MEDIA

Social media in China

Social media again tracks government data for **influenza:**
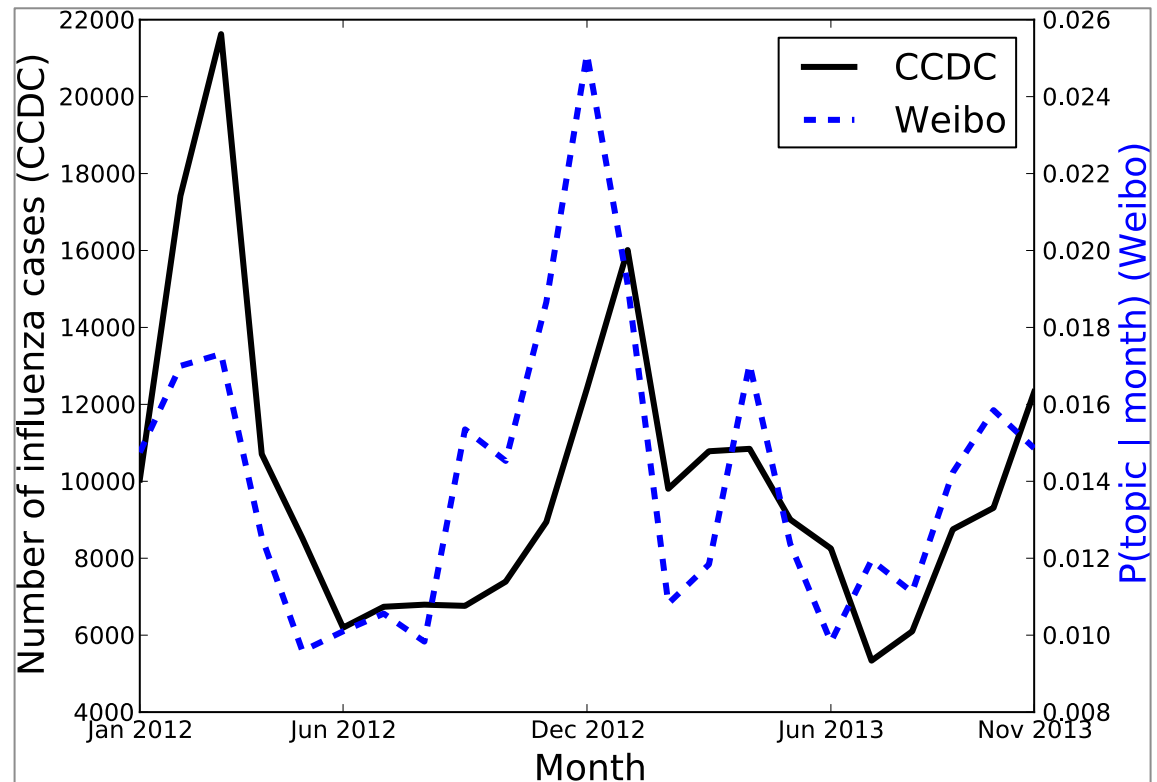
PUBLIC HEALTH ON SOCIAL MEDIA

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

HEALTHCARE QUALITY & SAFETY

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

HEALTHCARE QUALITY & SAFETY

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

CNN Health

Click
stori
med
happ
hosp
not
inter

Medical mistakes: Patients' stories

<< < 1 2 3 4 5 6 7 8 9 10 11 12 13 14

HEALTHCARE QUALITY & SAFETY

1,000 tweets analyzed:

> because the doctor screwed up a test .. I had to have ER surgery..was stabbed internally
>
> ↩ Reply ⇄ Retweet ★ Favorite

Error Source		Error Type	
Surgeon	18.1%	Surgical	7.2%
Physician	51.5%	Diagnostic	15.0%
Nurse	18.7%	Medication	22.2%
Pharmacist	0.0%	Procedure	36.5%
Other Medical	4.7%	Infection	0.0%
Dentist	0.0%	Communication	9.6%
Unknown	7.0%	Unknown	9.6%

WHAT IS PUBLIC HEALTH?

- **Disease surveillance**
- **Vaccination**
- **Behavioral risk factors**
- **Environmental effects**
- **Quality and safety**
- **… and much more**

All of these have been studied in social media!

I talked about just a few of them…

QUESTIONS?

ACKNOWLEDGMENTS