Launching the next generation of digital disease surveillance tools

Mauricio Santillana

Faculty member, CHIP, Boston Children's Hospital Informatics Program
Associate, Harvard Institute for Computational and Applied Sciences
Instructor, Harvard Medical School

HARVARD School of Engineering and Applied Sciences

HARVARD MEDICAL SCHOOL

Boston Children's Hospital

Beyond Google searches…

What are doctors searching for?

flu near you
do you have it in you?

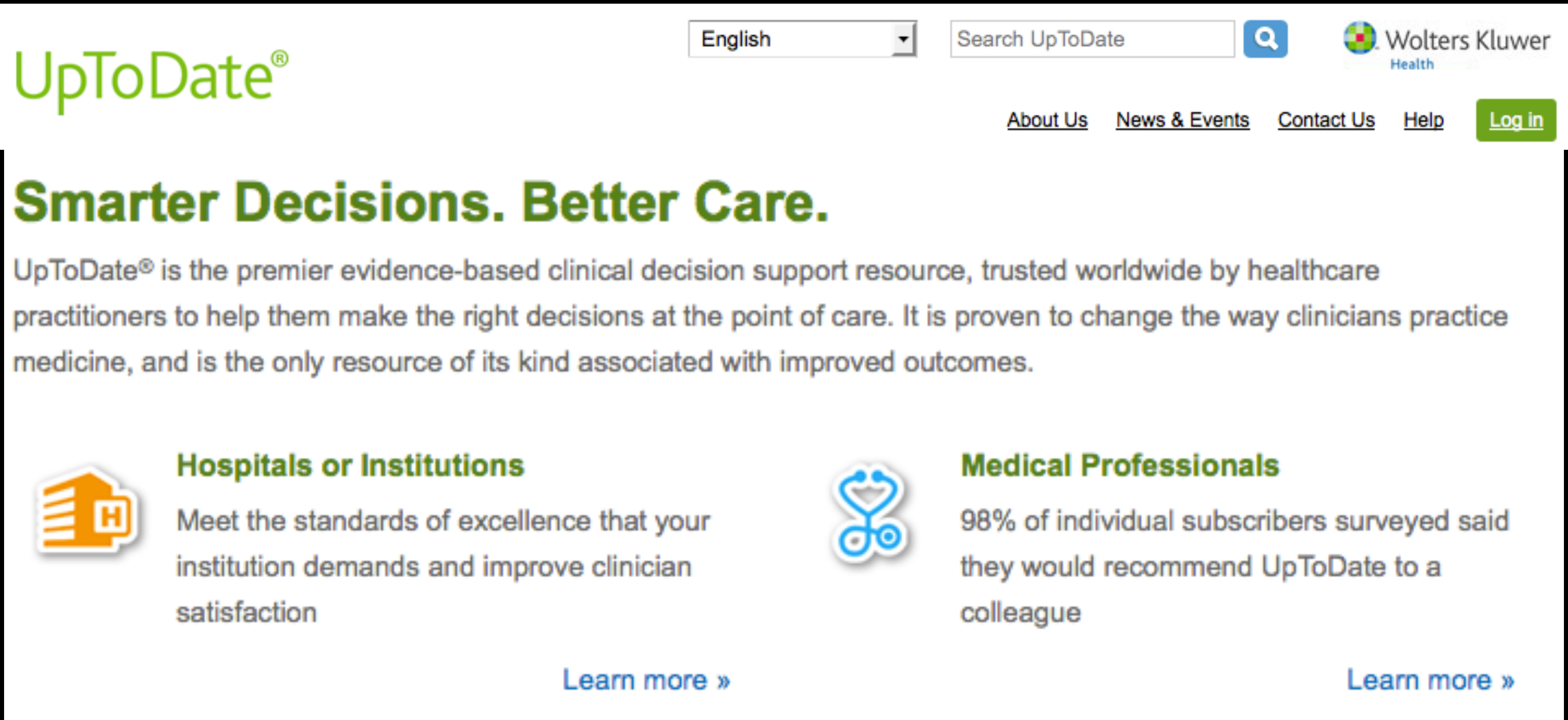
What are people tweeting? What are they reporting on crowd-sourced disease surveillance apps?

Can we use Electronic Health Records (EHR) to track disease incidence? What lab tests or medications are doctors prescribing?

What if we could get access to internet **searches** of **medical doctors** while in their practices (anonymously)?

Work with John Brownstein (HMS), Elaine Nsoesie (BCH, HMS), Sumiko Mekaru (BCH), David Scales (HMS)

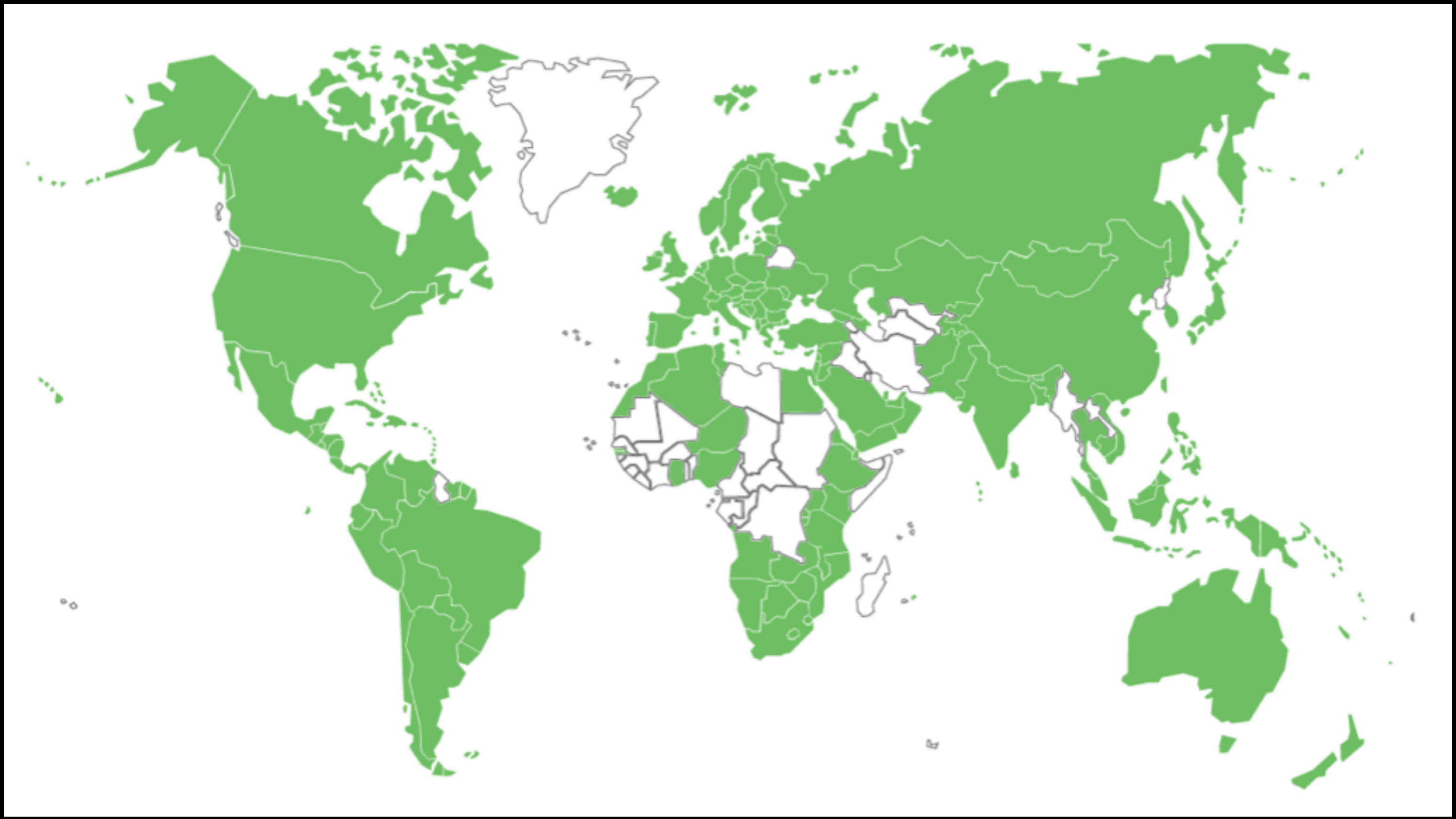
UpToDate is a clinician's data base



Good reputation

"The data suggests the use of computerized tools such as UpToDate enable better decisions, better outcomes and better care." — Ashish Jha, M.D., M.P.H., Harvard, and Study Author

Widely used by clinicians:
700,000 clinicians in 158 countries, almost 90% of academic medical centers in the United States



OXFORD JOURNALS

Clinical Infectious Diseases

Using Clinicians' Search Query Data to Monitor Influenza Epidemics

Mauricio Santillana,[1,2] Elaine O. Nsoesie,[2,3] Sumiko R. Mekaru,[2] David Scales,[2,4] and John S. Brownstein[2,3,5]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, [2]Children's Hospital Informatics Program, Boston Children's Hospital, [3]Department of Pediatrics, Harvard Medical School, Boston, and [4]Department of Internal Medicine, Cambridge Health Alliance, Massachusetts; and [5]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

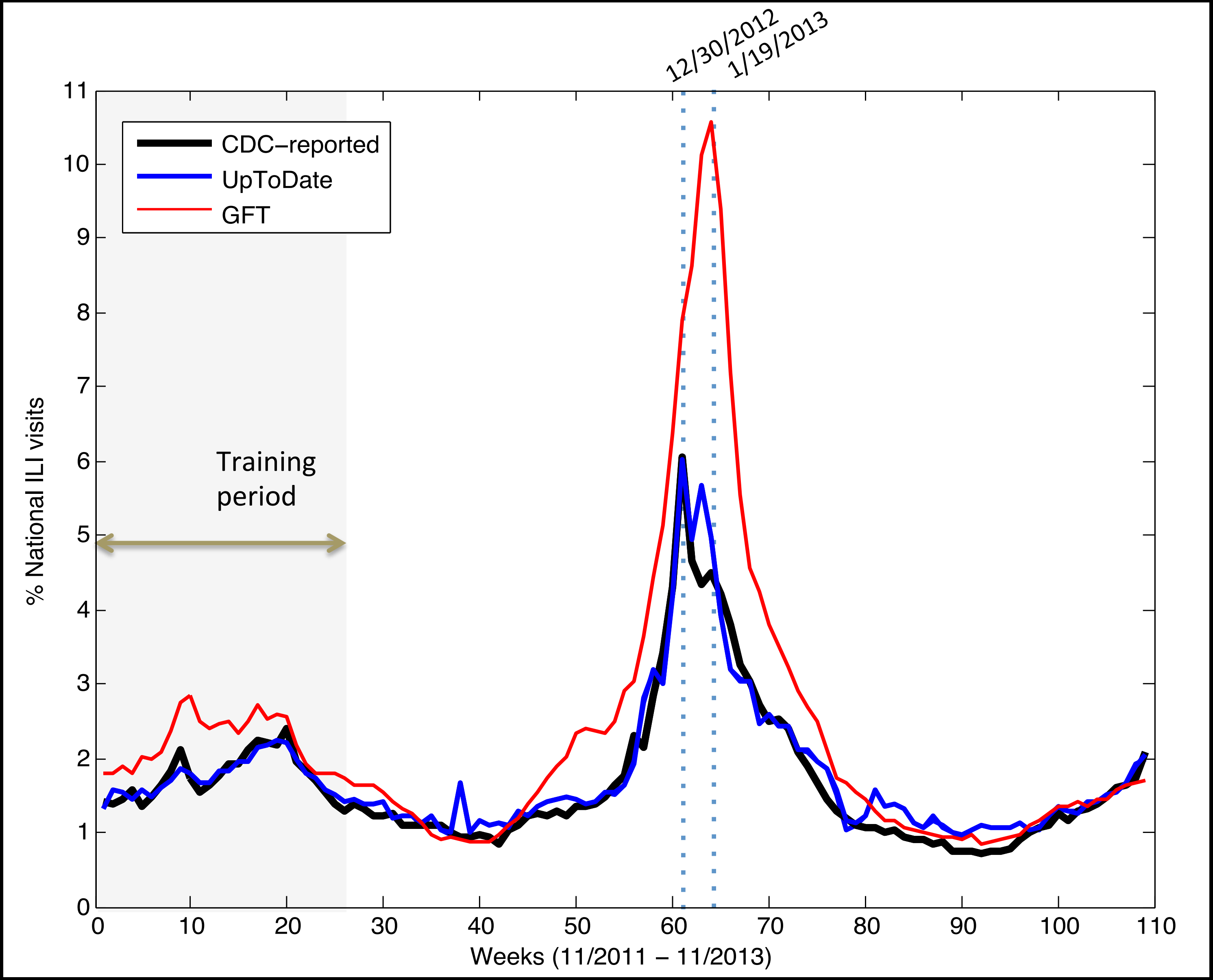
Search query information from a clinician's database, UpToDate, is shown to predict influenza epidemics in the United States in a timely manner. Our results show that digital disease surveillance tools based on experts' databases may be able to provide an alternative, reliable, and stable signal for accurate predictions of influenza outbreaks.

Keywords. digital disease detection; Internet-based disease surveillance; prediction of influenza.

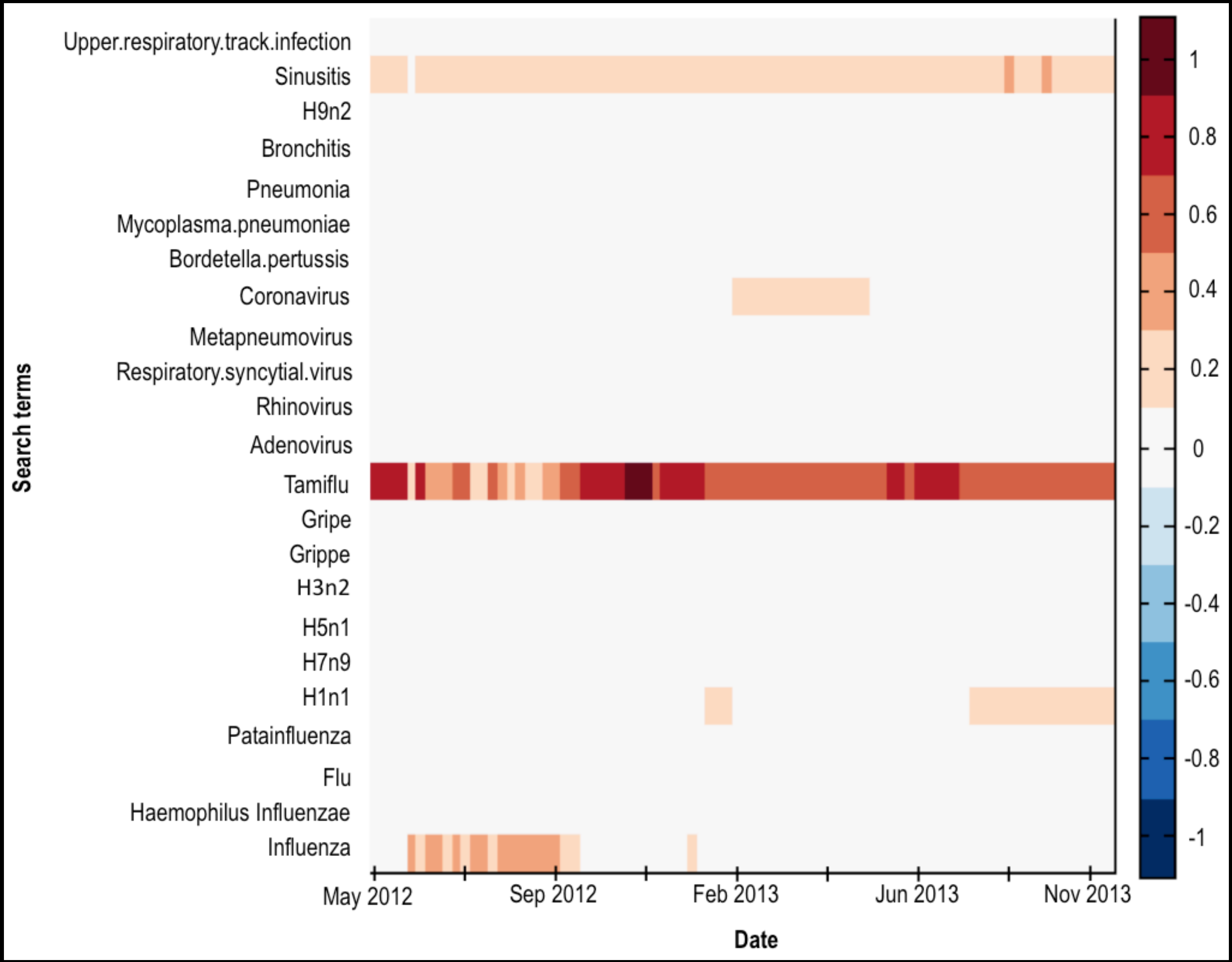
validated traditional surveillance systems and have the potential to provide timely epidemiologic intelligence to inform prevention messaging and healthcare facility staffing decisions.

The potential for the public's search activity to be influenced by anxiety, fears, and rumors raises concerns regarding reliability [10–13]. Although recent revisions to GFT have shown that these concerns can be partially mitigated [13–15], shifting Internet-based surveillance from the entire public to subject-matter experts may maintain timeliness while generating a more reliable and stable signal requiring much less data. A recent small retrospective study using data on queries to a Finnish primary care guidelines database demonstrated, for example, that disease-specific queries for Lyme disease, tularemia, and other infectious diseases correlated well with concurrent confirmed cases [16].

Here, we show that UpToDate (www.uptodate.com), a physician-authored clinical decision support Internet resource that is used by 700 000 clinicians in 158 countries and almost 90% of academic medical centers in the United States, can be used for syndromic surveillance of influenza. Specifically, we use UpToDate's search query activity related to ILI to design a timely sentinel of influenza incidence in the United States.



Santillana et al. Clinical Infectious Diseases, 2014 (Published online)

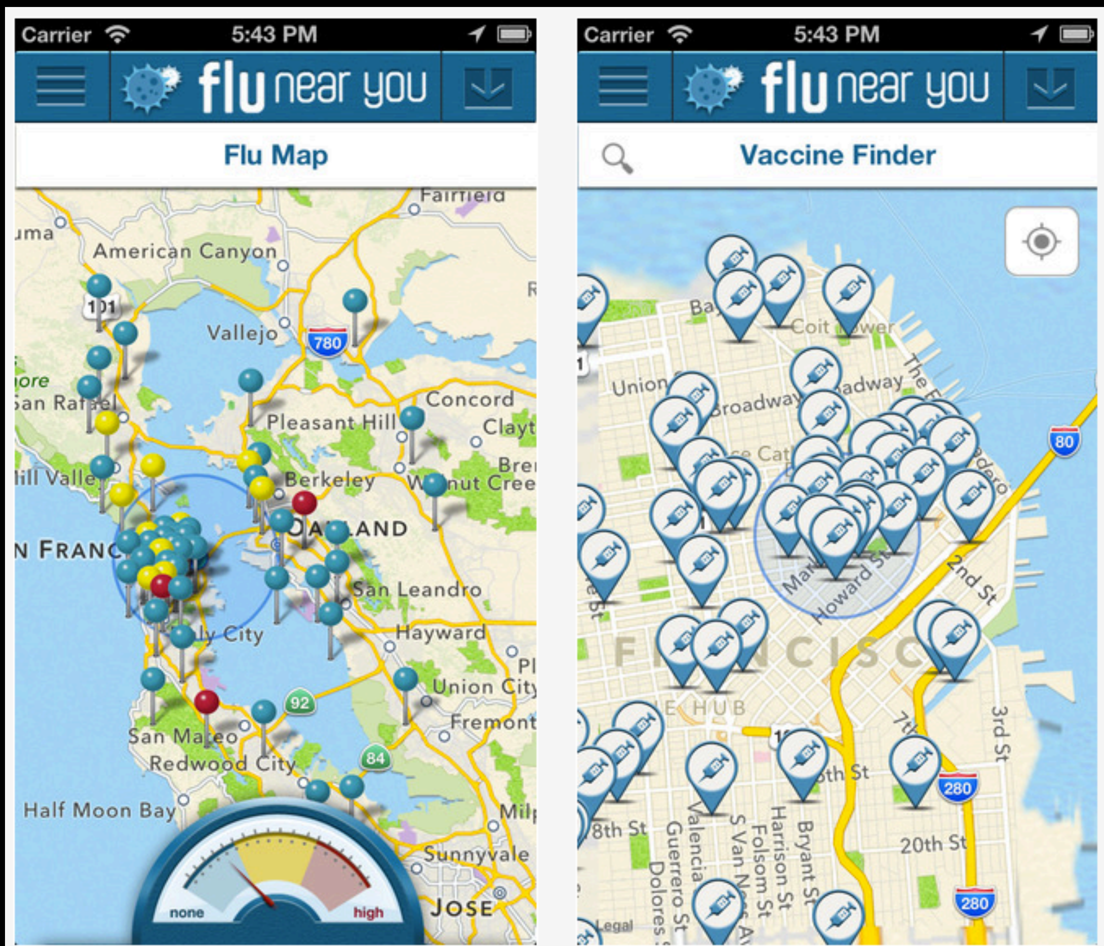


Santillana et al. Clinical Infectious Diseases, 2014 (Published online)

What if we could ask the general public if they are sick?

Working with Andre Nguyen (Harvard SEAS)

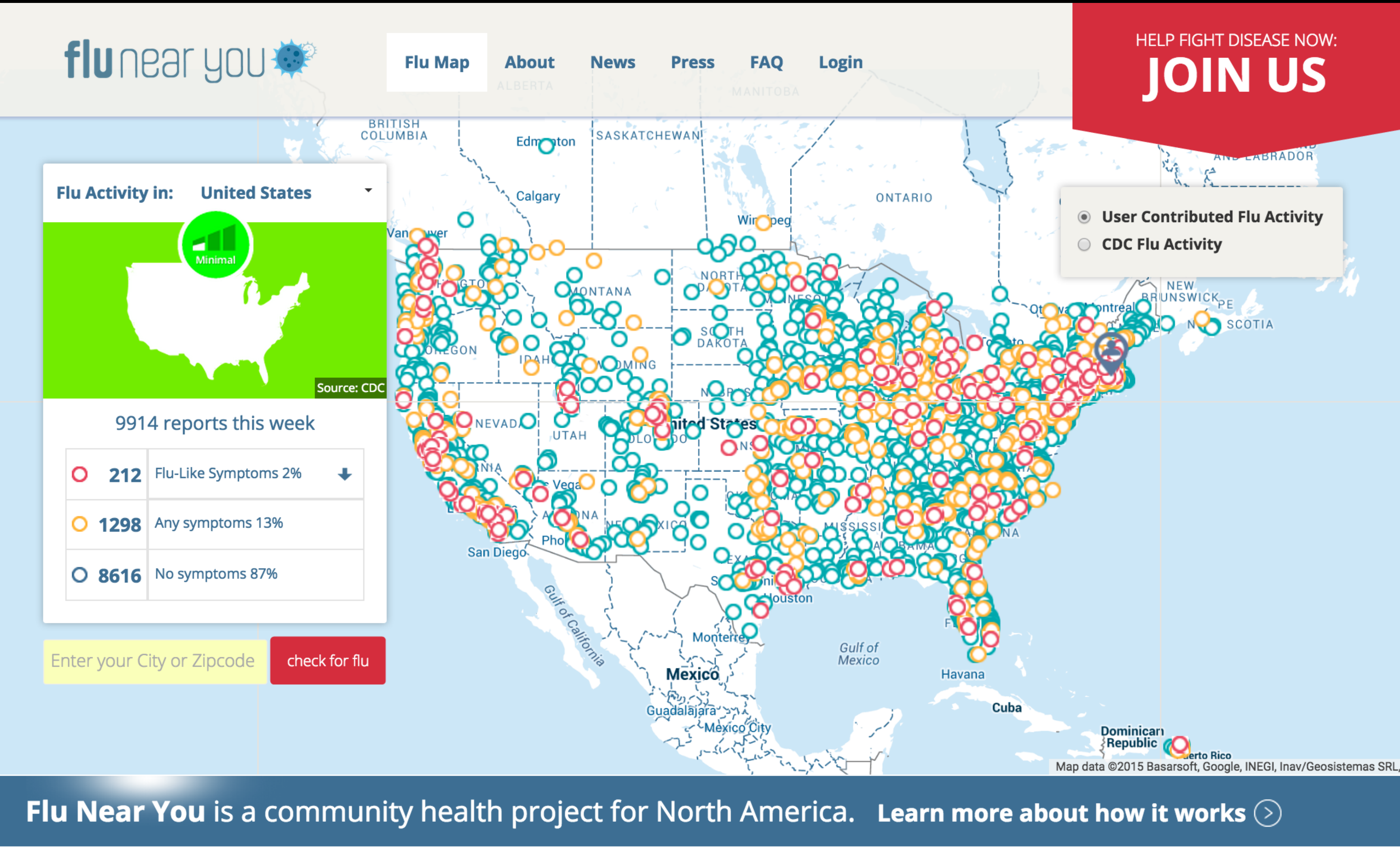
flu near you
do you have it in you?



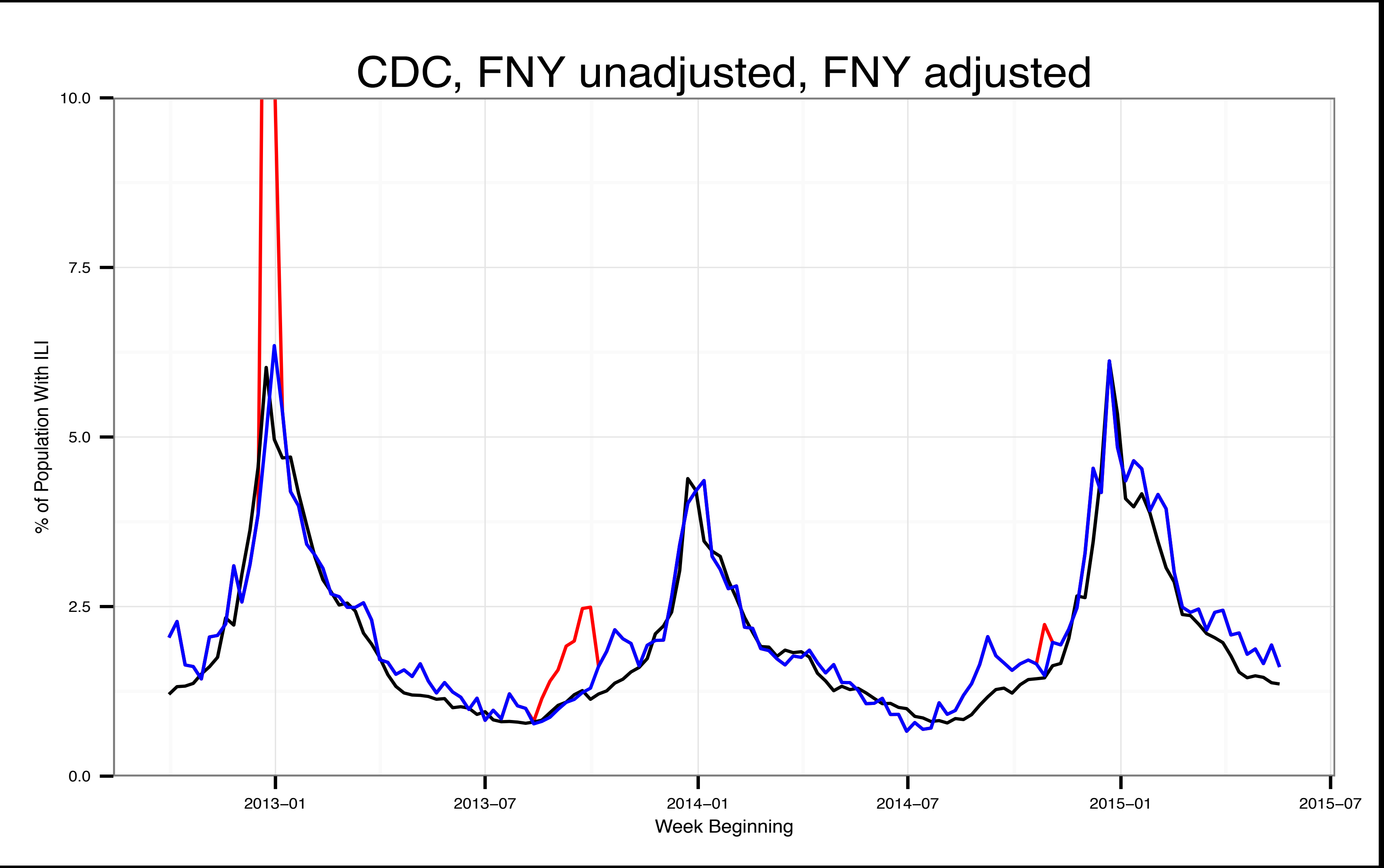
Launched in 2011

Working with Andre Nguyen (Harvard SEAS), Rumi Chunara, John Brownstein

https://flunearyou.org/



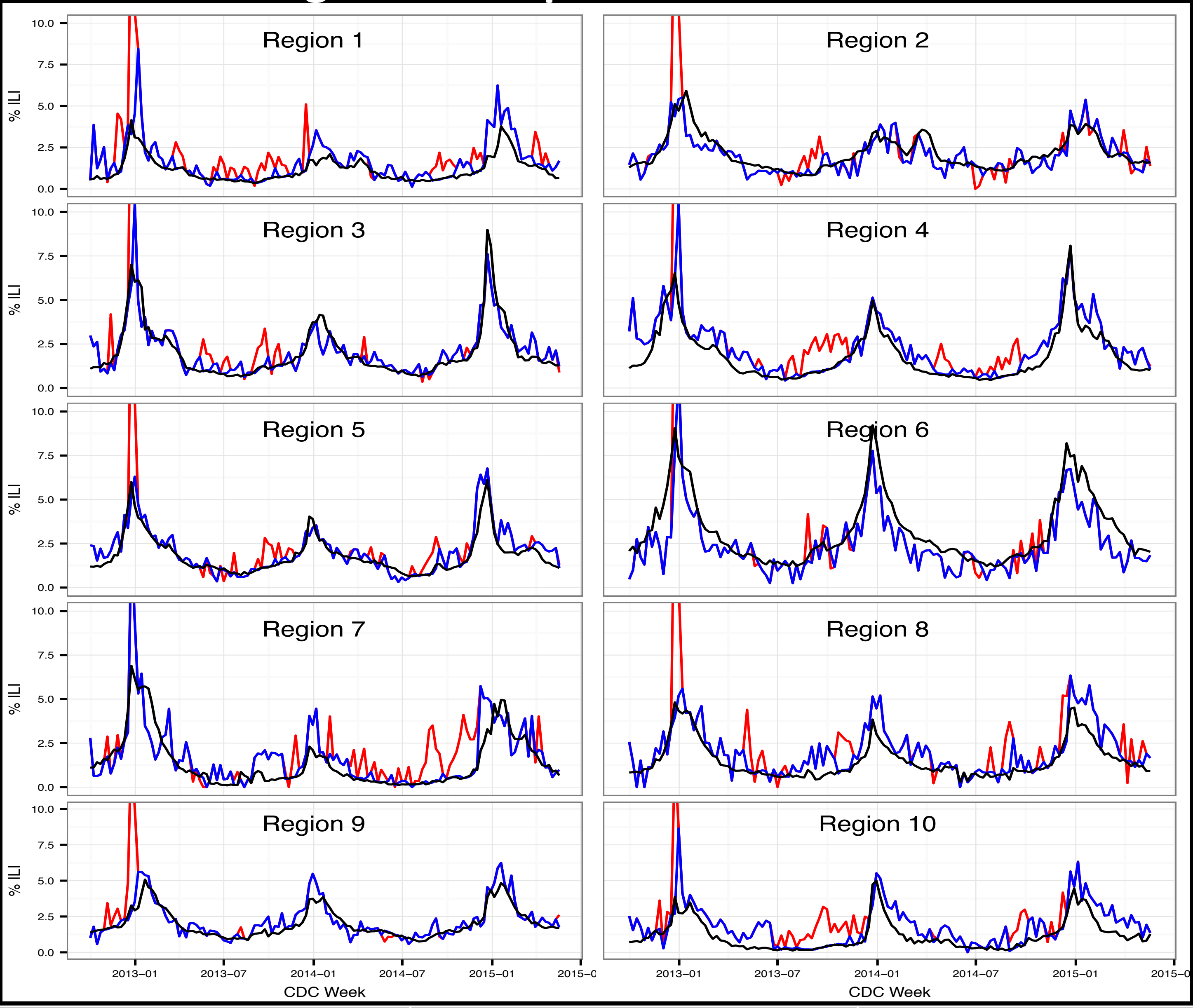
National comparison with CDC



Plots produced by Kristin Baltrusaitis

Time Series	Correlation	RMSE
FNY – raw	0.808	1.16
FNY – CDC adjusted	0.956	0.384

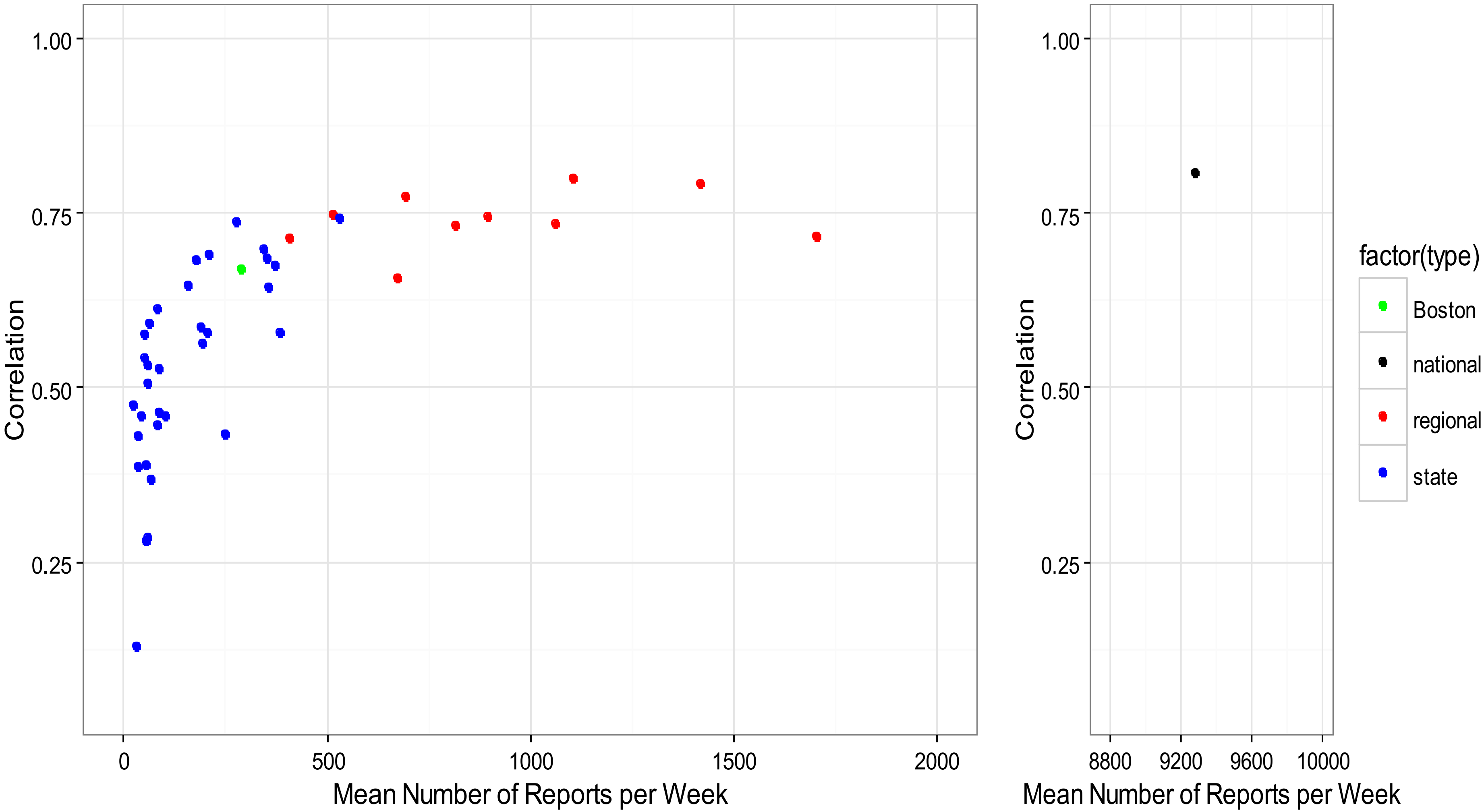
Regional comparisons with CDC



Plots produced by Kristin Baltrusaitis

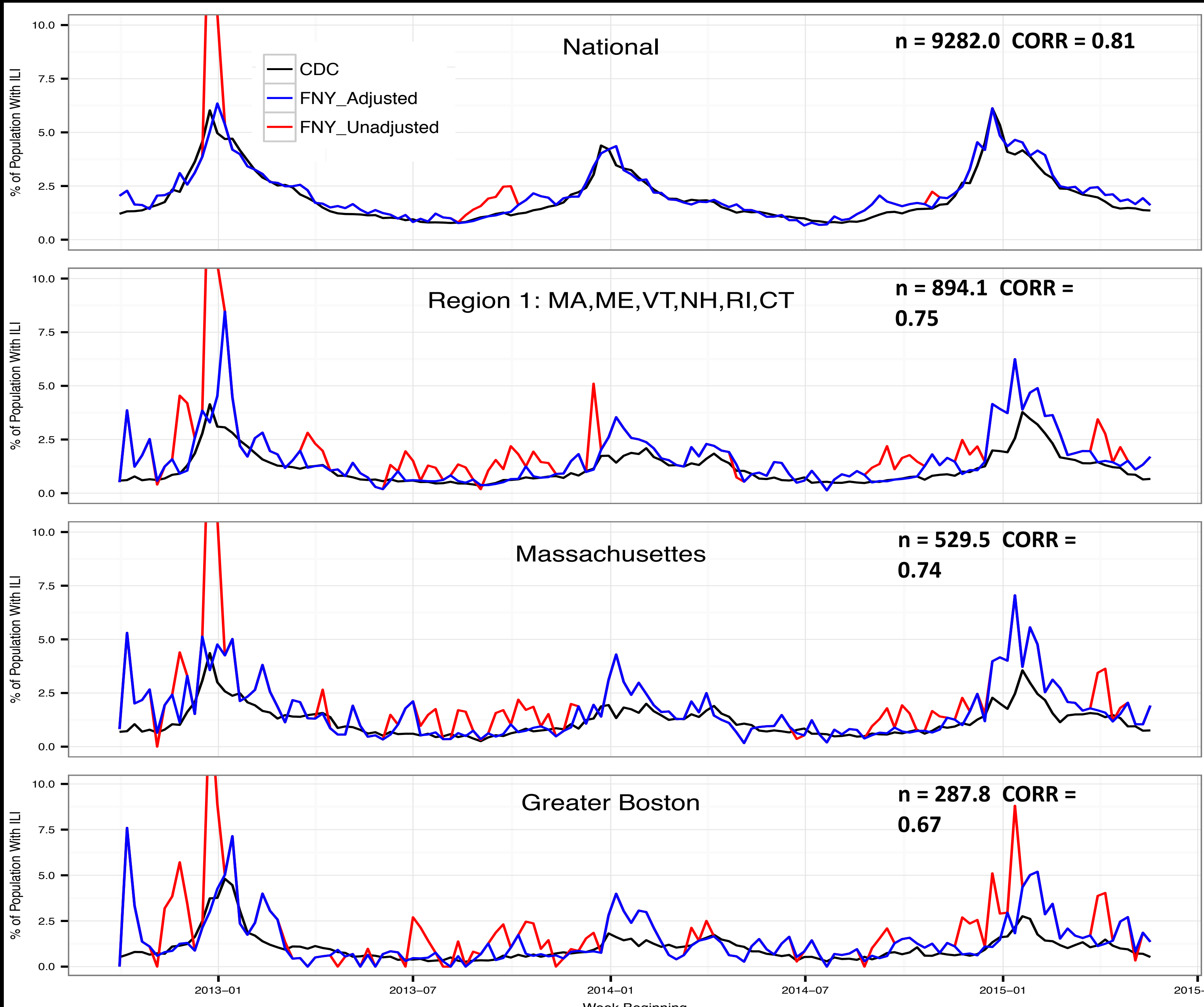
Time Series	Correlation Range	RMSE Range
FNY – raw	0.66 – 0.80	1.33 – 1.77
FNY – CDC adjusted	0.80 – 0.90	0.60 – 1.26

Correlation of FNY with flu information from CDC and Boston's Health Department



Plots produced by Kristin Baltrusaitis

Correlation of FNY with CDC. Multiple Geographic Scales



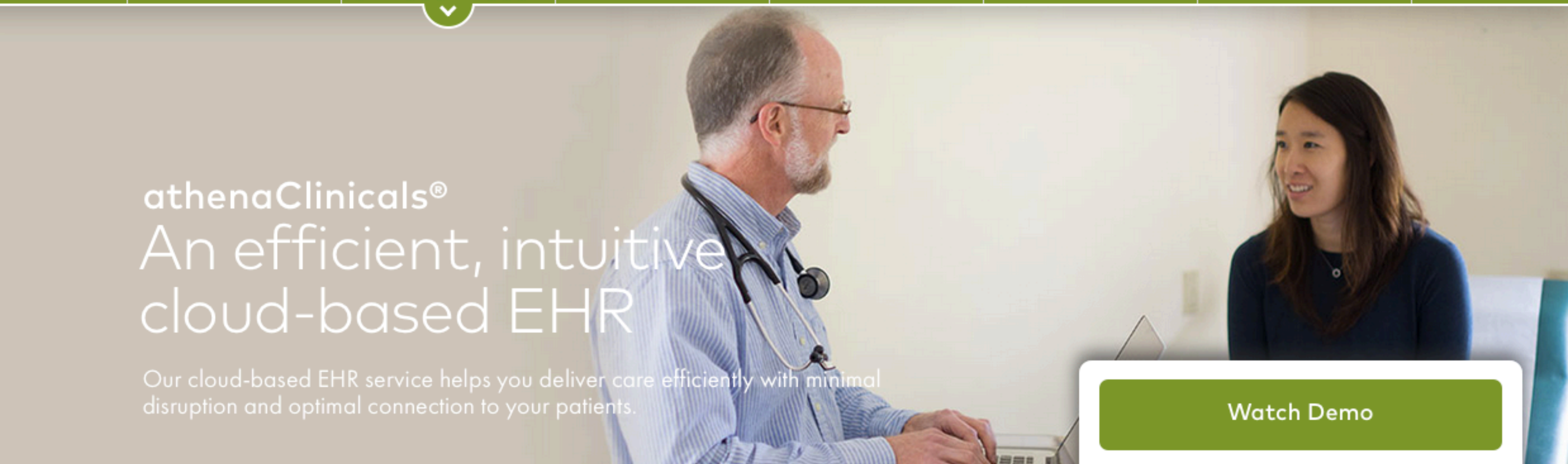
What if we had real-time access to **electronic health records**?

Work with: Andre Nguyen (Harvard SEAS), Tamara Louie (HSPH)
John Brownstein (HMS), Iyue Sung (athenahealth)

athenahealth

Products & Services | Industry Solutions | Results & Insights | About athenahealth

Services Overview | Electronic Health Records (EHR) | Medical Billing & Practice Management | Patient Engagement | Order Transmission | Epocrates®

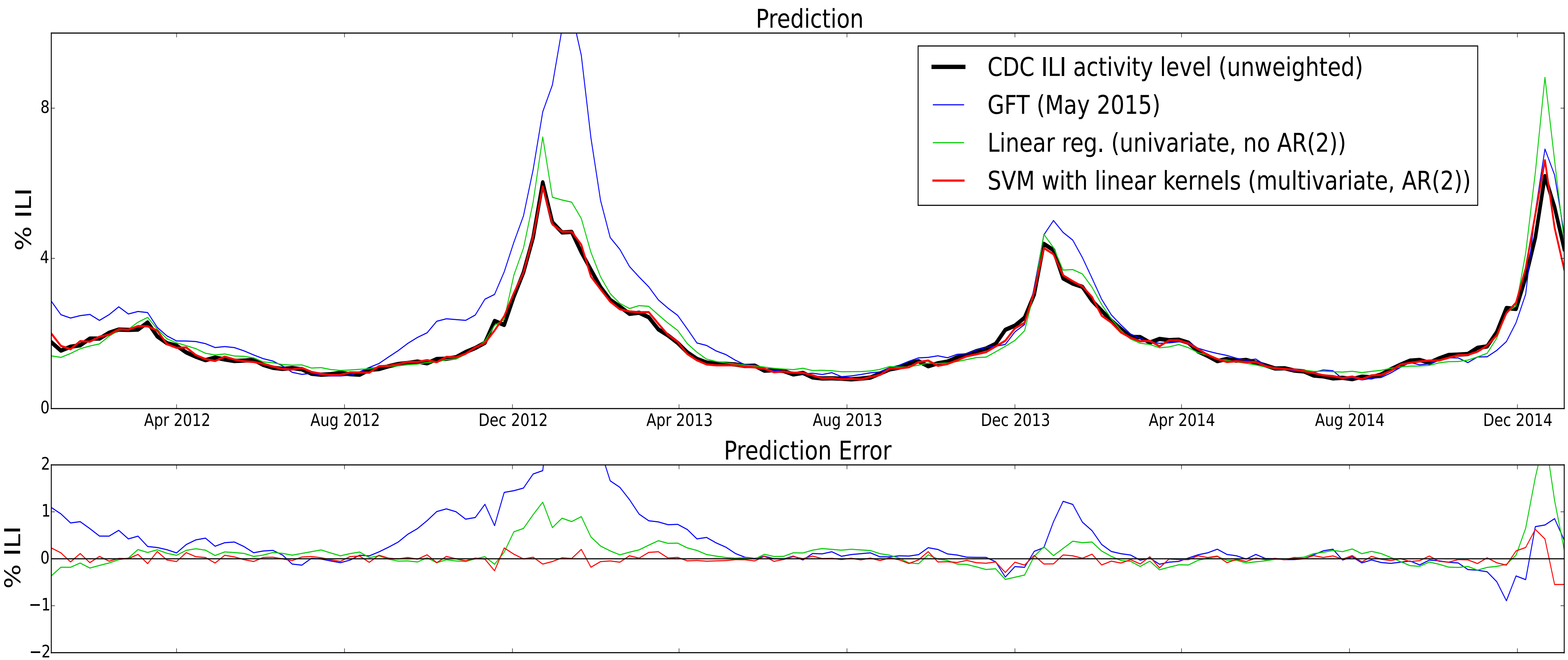


athenaClinicals®

An efficient, intuitive cloud-based EHR

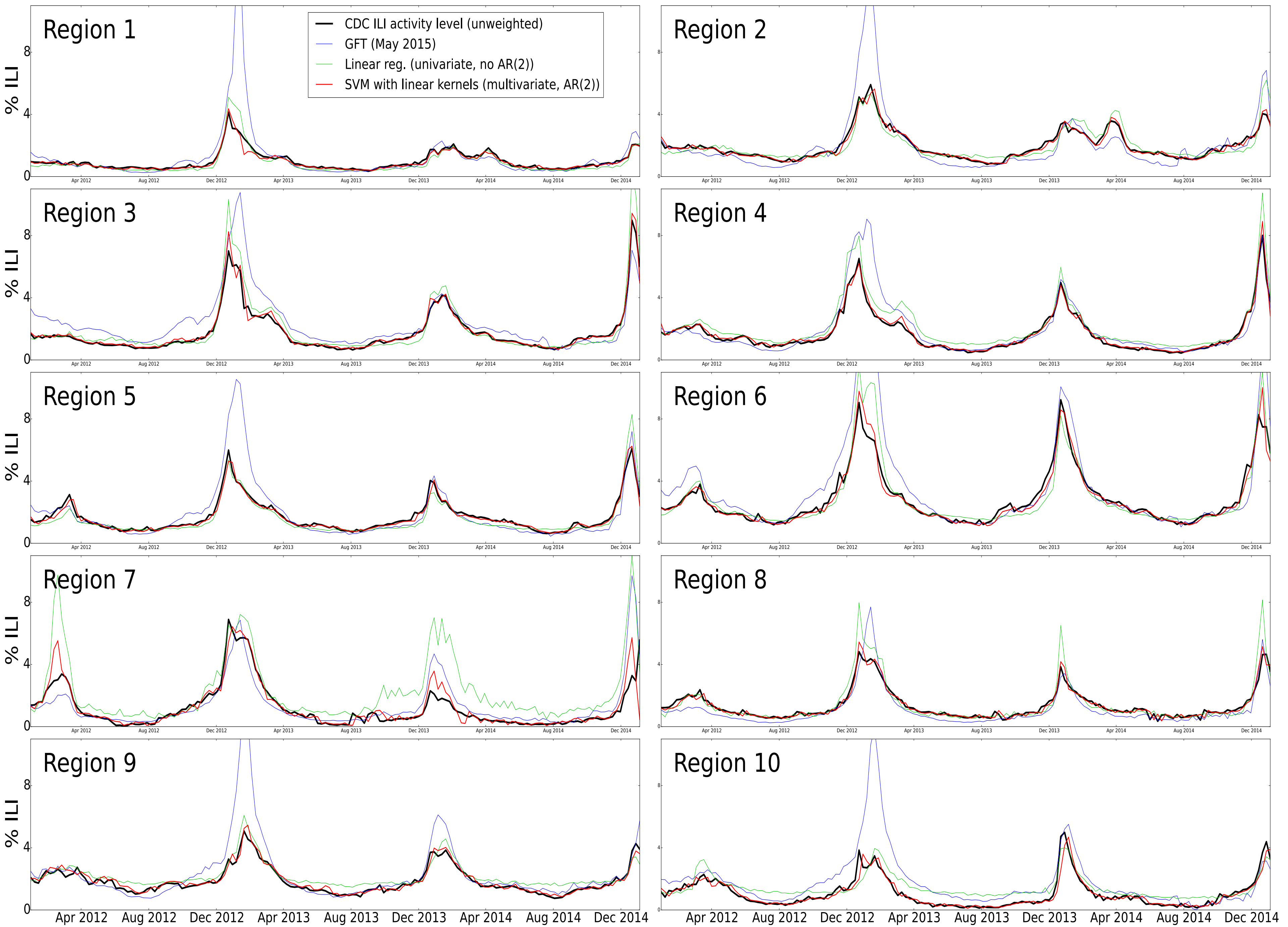
Our cloud-based EHR service helps you deliver care efficiently with minimal disruption and optimal connection to your patients.

Watch Demo



Santillana et al. 2016 Scientific Reports, 25732

We can predict flu in finer geographic scales with amazing accuracy!



Santillana et al. 2016 Scientific Reports, 25732

Overall					
Data Utilized	**Inclusion of AR Terms?**	**RMSE**	**Rel. RMSE (%)**	**Correlation**	**Algorithm**
National					
GFT	--	1.03	32.91%	0.913	--
% ILI	No	0.37	12.50%	0.981	Linear
% ILI, % viral, % flu visits	Yes	0.12	4.92%	0.994	SVM (linear)
Region 1					
GFT	--	1.33	55.16%	0.786	--
% ILI	No	0.28	21.87%	0.960	Linear
% ILI, % viral, % flu visits	Yes	0.14	11.87%	0.973	SVM (linear)
Region 2					
GFT	--	1.15	34.35%	0.907	--
% ILI	No	0.44	21.09%	0.903	Linear
% ILI, % viral, % flu visits	Yes	0.19	8.39%	0.980	SVM (linear)
Region 3					
GFT	--	1.06	56.18%	0.860	--
% ILI	No	0.59	22.15%	0.981	Linear
% ILI, % viral, % flu visits	Yes	0.25	9.87%	0.987	SVM (linear)

Santillana et al. 2016 Scientific Reports, 25732

Refining the spatial resolution…

Tracking Flu using twitter
(Daily analysis in NYC)

Work with R. Nagar, Q. Yuan, C. Freifeld, A. Nojima, R. Chunara, and J. S. Brownstein

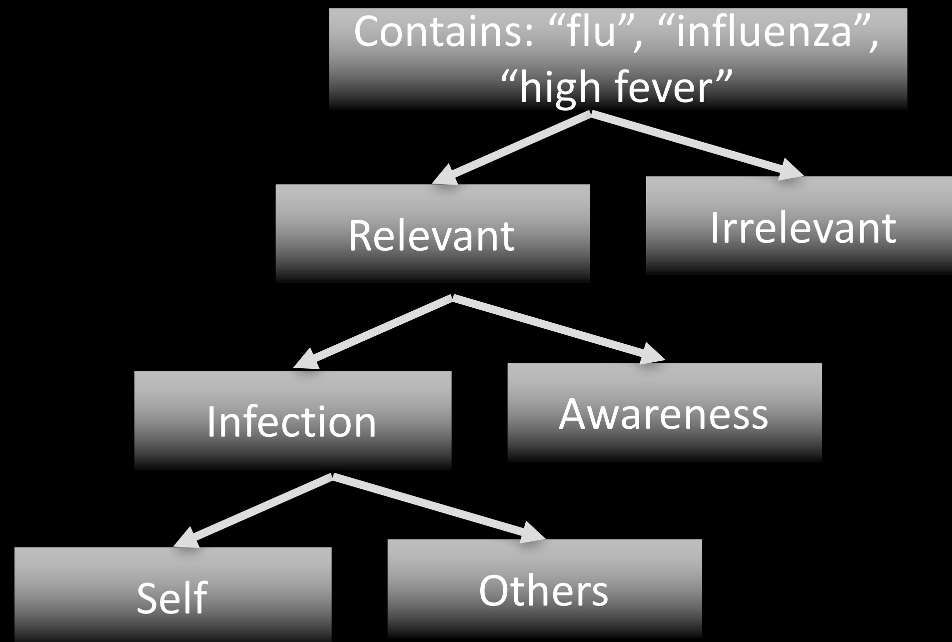
Natural Language Processing
(Using geo-located tweets)

1. Identified tweets containing “flu”, “influenza”, “gripe”, “high fever”
2. Classified tweets in categories

Table 1. Examples of Classified Tweets

Label	Example Tweets
Irrelevant	The flu shot prevents hangovers, so going all out w/5 fingers of...wait, it's "no drinking" that prevents hangovers? #toolate
	Romney: I would not put no flu zones over Syria. Military is not necessary in the conflict #debate2012
RISH	This flu is kicking my butt,2nd day off work. Hopefully I can win the battle because I'm losing sick days & that might hurt my pockets later
	Maldito flu sueltame!!! ðŸ˜«ðŸ˜¡
RISM	Uhhh I think I might be getting the flu /:
	Creo que me va a dar Gripe :'(
RISL	Finally getting over a miserae flu.
	@boyXsupreme ik ik it's awful. the past two weeks darci and I have had the flu but thank god we're done with it. get lots of rest + tlc ðŸ’•
RIOH	When @CiaraAnnex3 has the flu... Love you but stay the hell away <3
	Running on no sleep my poor daughter has the flu
RIOM	@giannarussso YOU PROB GOTS THE FLU!!, ariannas has it
	@YanieseRivera damn girl, do you have the flu?
RIOL	On day 6, son's #flu is gone. He threw open side door and screamed to the outside, "FREEDOM"! Then shoveled snow. I am miserable on day 3.
	@_AlexAlford she's good too, fortunately she never actually got the flu. just a fever for a day or two
RASH	I'd rather get the flu than get the flu shot, #JustSayin. No needles for me.
	The flu is an epidemic here and I volunteer at a preschool twice a week. If I don't get he flu it will be a miracle.
RASM	I survived more than a week in NYC without contracting the flu! Let's hope the plane ride home won't break me. Trying to stay healthy here.
	So glad to be back in NYC, but stay away from me you Flu filled city.
RASL	Ah yea... flu shot acquired! (@ Duane Reade) http://t.co/RSR2Bl11
	Just got a flu shot and Tdap booster at @onemedical â€” if youâ€™ll be in close contact with an infant, consider taking these vaccines.
RAOH	Flu infections sweep America hospitalizing thousands and leaving 18 children dead of complications, ... http://t.co/Yf9eQikm
	19,000 flu cases across NY this week. I'm not leaving my house.
RAOM	Wash your hand. America sick girl. #influenza #SICKENING #cleanup http://t.co/MKdASaV9
	The latest figures from the CDC show that flu cases are still rising in the west. Listen to our newscast: http://t.co/0kCiyoP8
RAOL	#patient advice. #flu vaccine not only protects you but your community Too. Less outbreaks. U may survive the flu, but sicker people may not
	â€œ@grubstreetny: Hereâ€™s How New York Chefs Beat the Flu http://t.co/bV3pEZ2Câ€□ uuuuu for real?

Categories



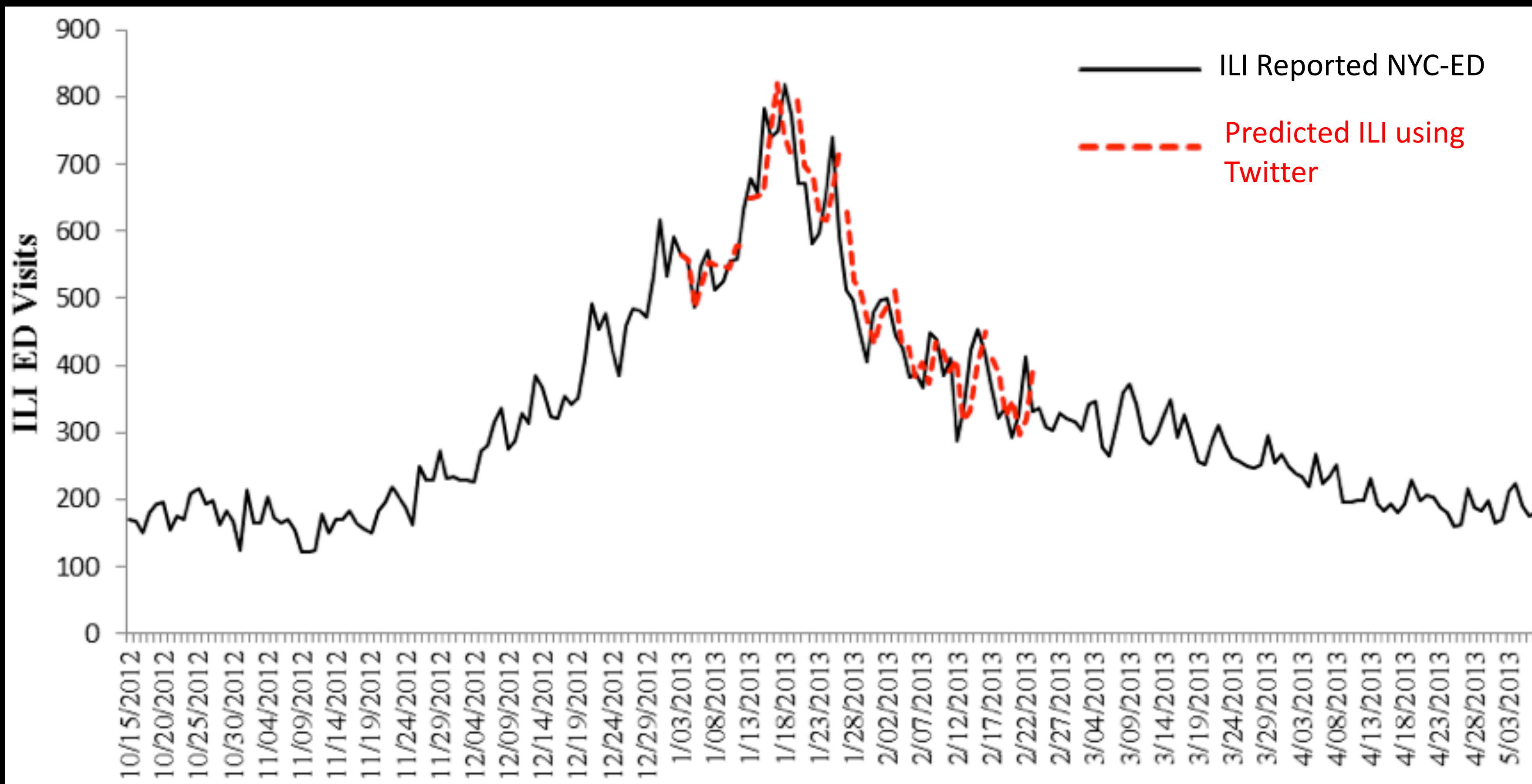
First experiment: was done by hand…

Nagar et al. (2014) Journal of Medical Internet Research. In press

Table 1. Examples of Classified Tweets

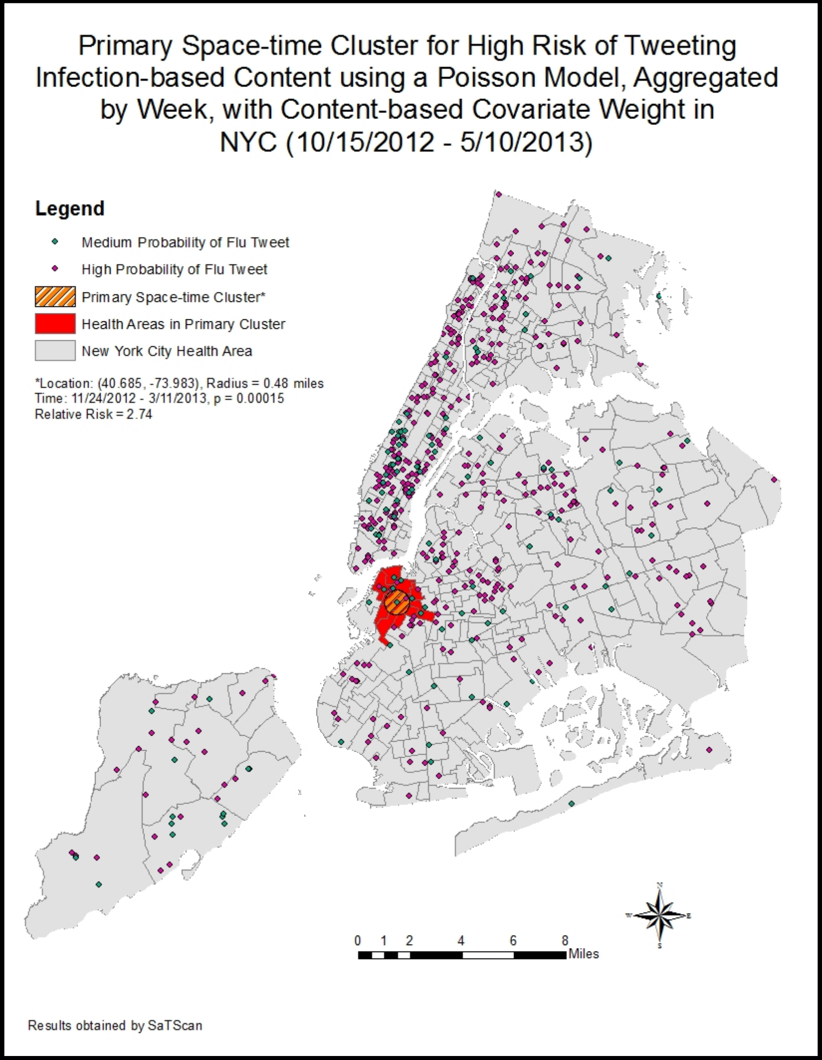
Label	Example Tweets
Irrelevant	The flu shot prevents hangovers, so going all out w/5 fingers of...wait, it's "no drinking" that prevents hangovers? #toolate
	Romney: I would not put no flu zones over Syria. Military is not necessary in the conflict #debate2012
RISH	This flu is kicking my butt,2nd day off work. Hopefully I can win the battle because I'm losing sick days & that might hurt my pockets later
	Maldito flu sueltame!!! ðŸ˜«ðŸ˜¡
RISM	Uhhh I think I might be getting the flu /:
	Creo que me va a dar Gripe :'(
RISL	Finally getting over a miserae flu.
	@boyXsupreme ik ik it's awful. the past two weeks darci and I have had the flu but thank god we're done with it. get lots of rest + tlc ðŸ’•
RIOH	When @CiaraAnnex3 has the flu... Love you but stay the hell away <3
	Running on no sleep my poor daughter has the flu
RIOM	@giannarussso YOU PROB GOTS THE FLU!!, ariannas has it
	@YanieseRivera damn girl, do you have the flu?
RIOL	On day 6, son's #flu is gone. He threw open side door and screamed to the outside, "FREEDOM"! Then shoveled snow. I am miserable on day 3.
	@_AlexAlford she's good too, fortunately she never actually got the flu. just a fever for a day or two
RASH	I'd rather get the flu than get the flu shot, #JustSayin. No needles for me.
	The flu is an epidemic here and I volunteer at a preschool twice a week. If I don't get he flu it will be a miracle.
RASM	I survived more than a week in NYC without contracting the flu! Let's hope the plane ride home won't break me. Trying to stay healthy here.
	So glad to be back in NYC, but stay away from me you Flu filled city.
RASL	Ah yea... flu shot acquired! (@ Duane Reade) http://t.co/RSR2Bl11
	Just got a flu shot and Tdap booster at @onemedical â€” if youâ€™ll be in close contact with an infant, consider taking these vaccines.
RAOH	Flu infections sweep America hospitalizing thousands and leaving 18 children dead of complications, ... http://t.co/Yf9eQikm
	19,000 flu cases across NY this week. I'm not leaving my house.
RAOM	Wash your hand. America sick girl. #influenza #SICKENING #cleanup http://t.co/MKdASaV9
	The latest figures from the CDC show that flu cases are still rising in the west. Listen to our newscast: http://t.co/0kCiyoP8
RAOL	#patient advice. #flu vaccine not only protects you but your community Too. Less outbreaks. U may survive the flu, but sicker people may not

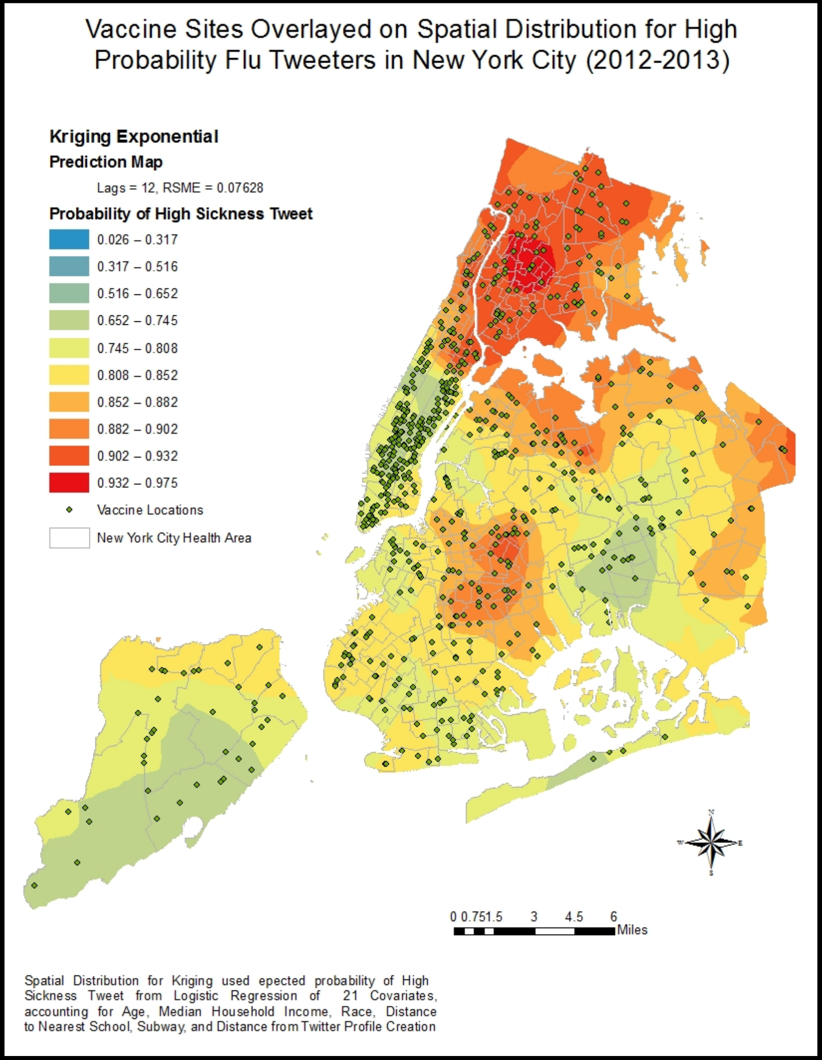
Daily ILI visits (as reported by the NYC emergency department) compared to predicted ILI using twitter data



Nagar et al. (2014) Journal of Medical Internet Research. In press

Spatial Analysis





Nagar et al. (2014) Journal of Medical Internet Research. In press

We now have multiple independent ways to estimate flu activity nationally and regionally.

What do we do with **more and more** flu predictors?

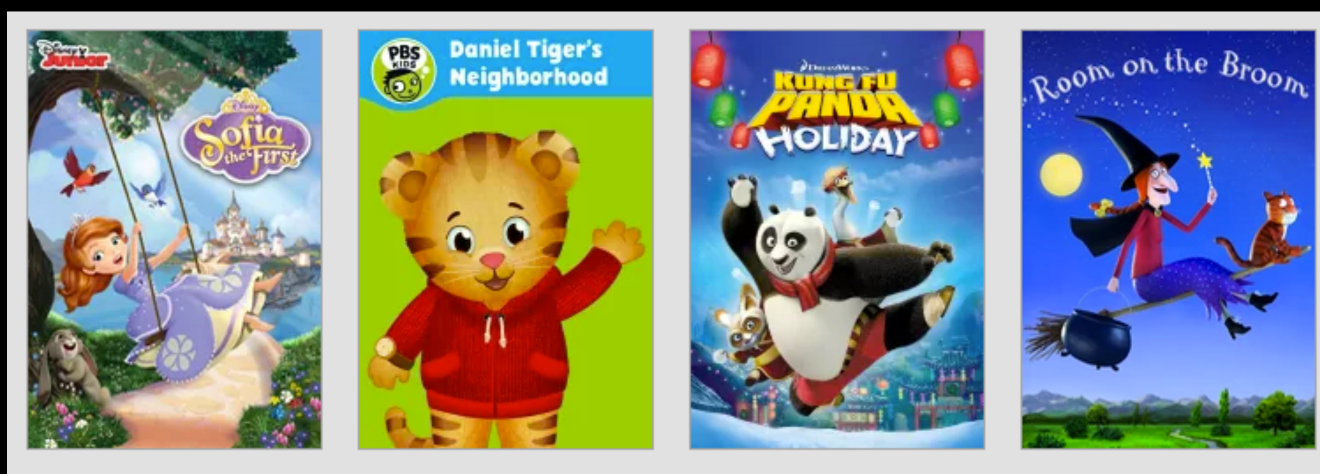
Data ↗ **Information** ↗

Knowledge and Prediction ?

Movie selection prediction

(millions of users)

NETFLIX Browse Personalize KiDS

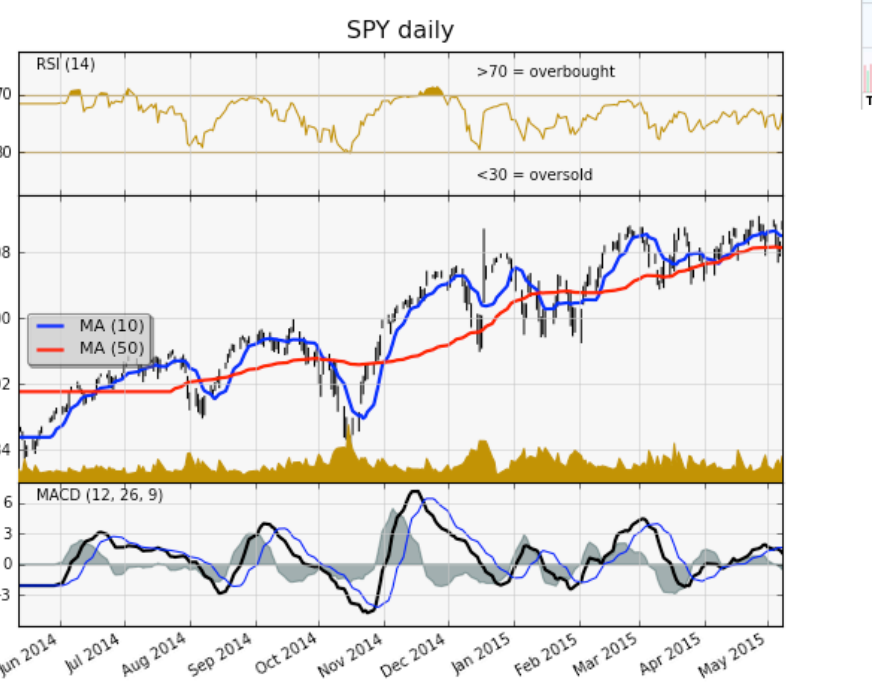


Factors: **time** of the day you are watching, **day**, **location**, **previous movie** you watched

Stock market price prediction
(multiple actors)

1. Fundamental Analysis

2. Time-series analysis



News

Recent Articles

Stocks Waffle Amid Greece Bailout Uncertainty
at TheStreet Mon May 11 2015

5 Rocket Stocks That Look Ready for Blastoff
at TheStreet Mon May 11 2015

Stock Market Update from Briefing.com
Briefing.com Mon May 11 2015

Tech Stocks from Briefing.com
Briefing.com Mon May 11 2015

All 10 S&P Sectors Are in the Red
at Bloomberg Mon May 11 2015

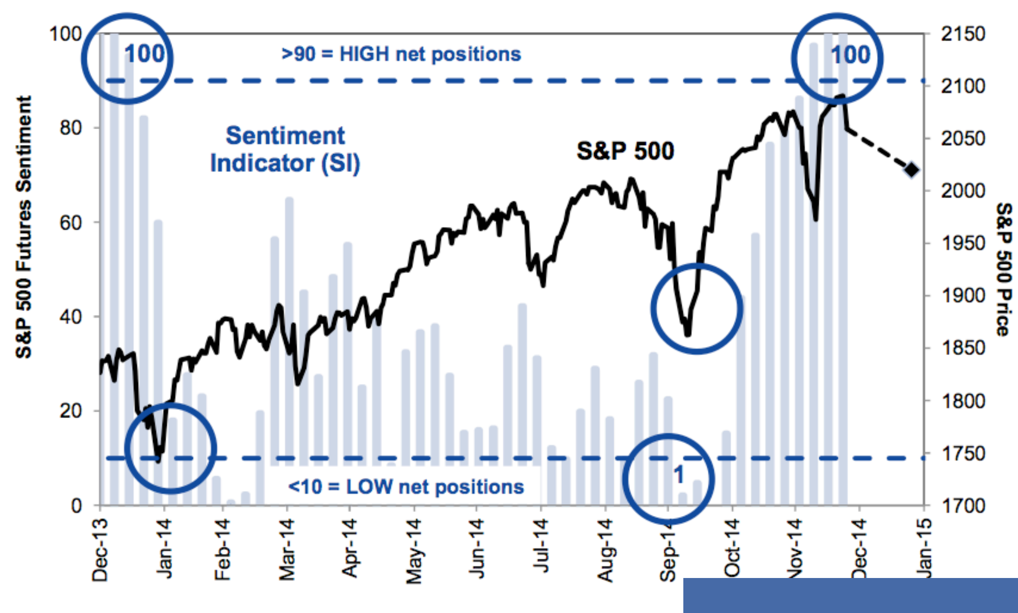
U.S. Equities Fluctuate After Rally Leaves S&P 500 Near Record
Bloomberg Mon May 11 2015

Stocks narrowly mixed as Street digests jobs; Cisco jumps 1.5%
at CNBC Mon May 11 2015

Stock Market Signals From Social Media

Multiple predictors using social media

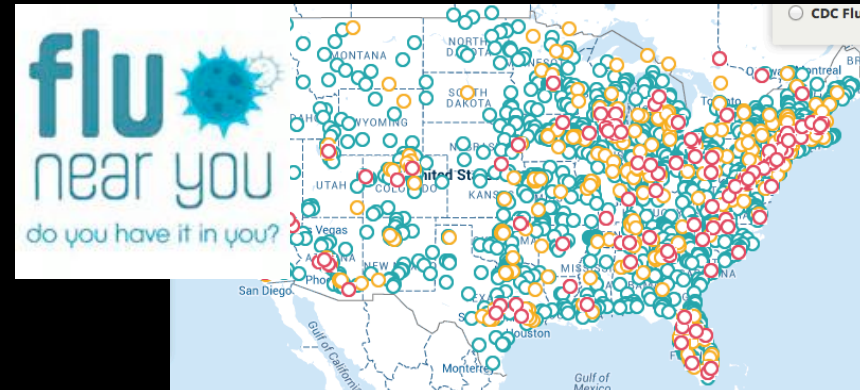
Exhibit 1: GS US equity Sentiment Indicator suggests near-term downside risk to S&P 500
as of December 31, 2014

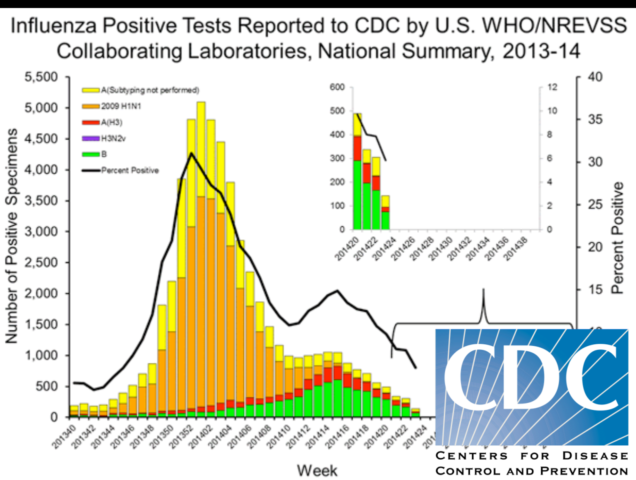


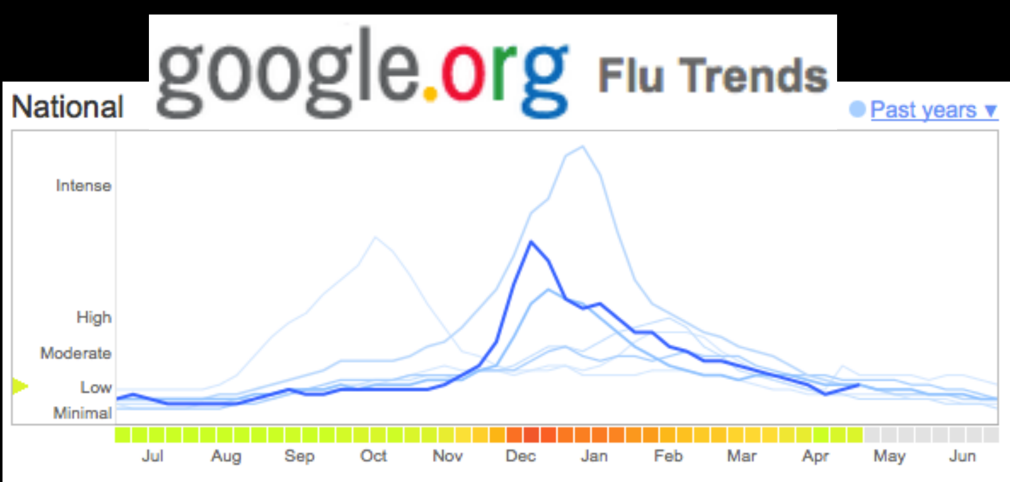
Source: Haver, CFTC, FactSet, and Goldman Sachs Global Investment Research.

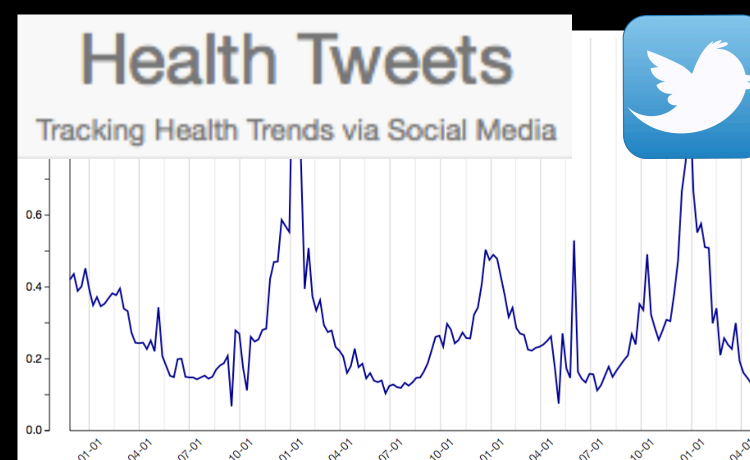
Digital Disease Detection

(millions of citizens worldwide)










GUESS
THE
NUMBER
OF
JELLY BEANS

So, what do we do with multiple predictors?

One solution: Pick the best performing one!

OR

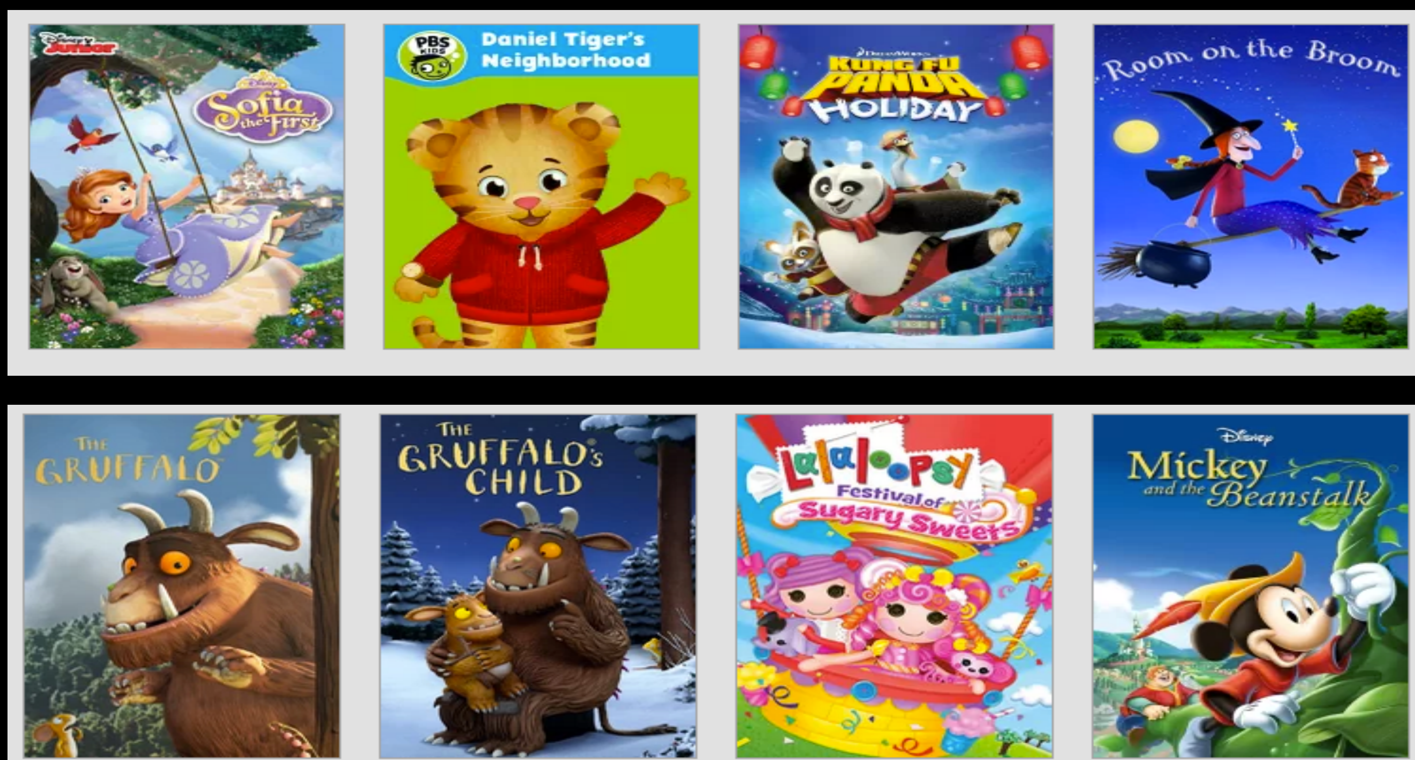
Combine information using a **voting system** or **ensemble** method! (Machine Learning)

Movie selection

NETFLIX Browse Personalize KiDS

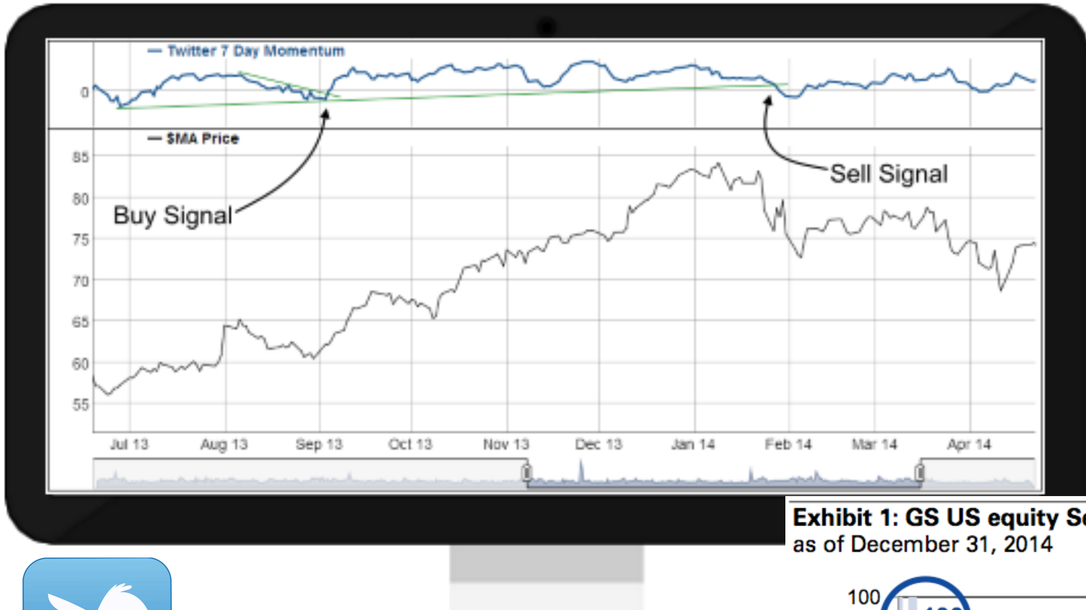
Different algorithms find optimal suggestions

Combine predictions using voting system (ensemble)



Factors: **time** of the day you are watching, **day**, **location**, **previous movie** you watched

Stock Market Signals From Social Media

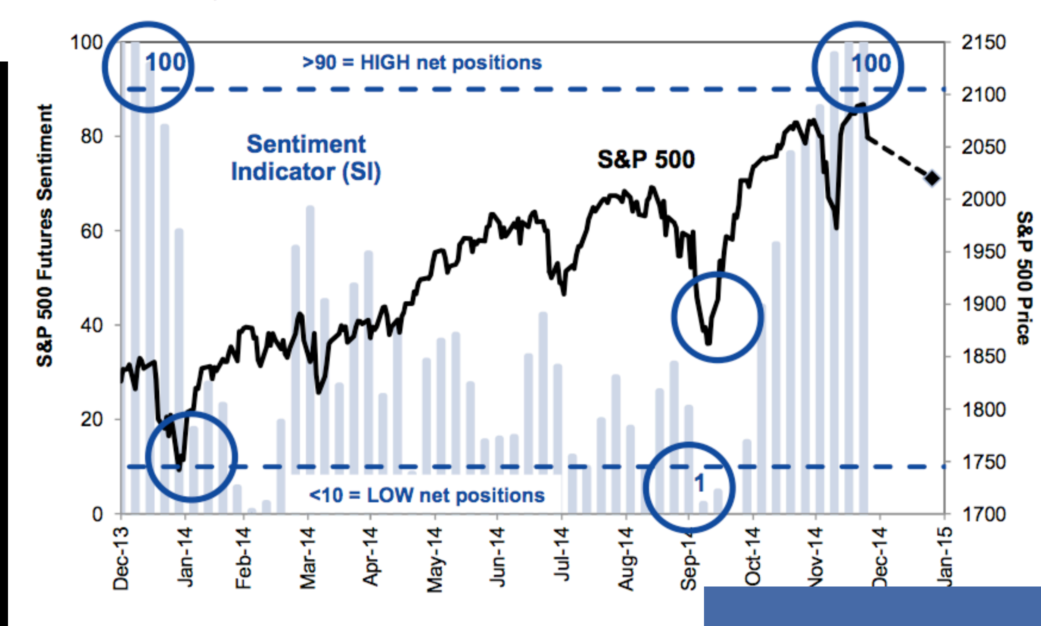


Combine information from all predictors and produce an accurate and robust prediction

Combine:

1. Fundamental analysis
2. Time-series analysis
3. Internet-based analysis

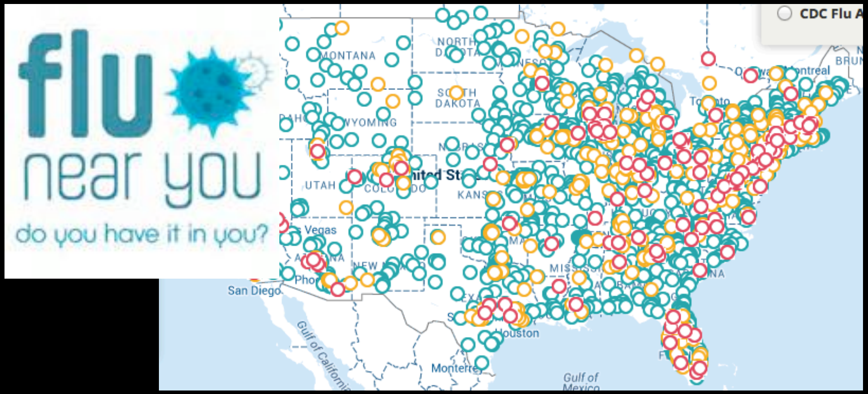
Exhibit 1: GS US equity Sentiment Indicator suggests near-term downside risk to S&P 500
as of December 31, 2014

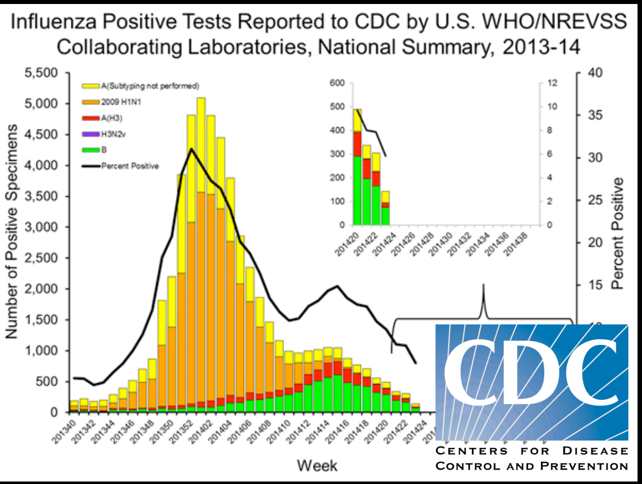


Source: Haver, CFTC, FactSet, and Goldman Sachs Global Investment Research.

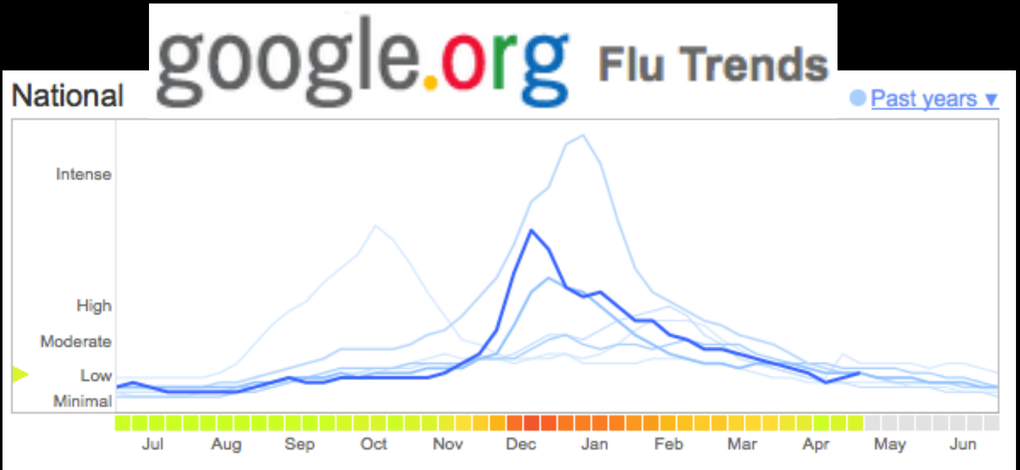
Breiman L. *Stacked regressions*. Machine Learning, 24, 49-64 (1996)

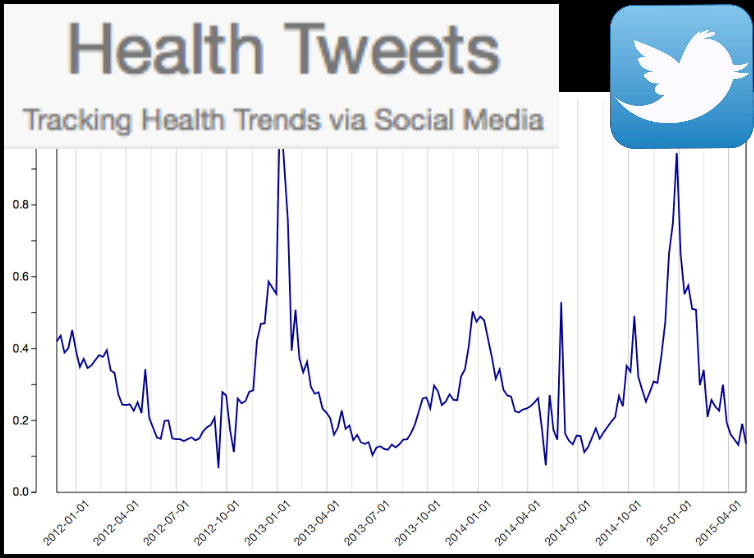
Digital Disease Detection
(US case study)





athenahealth





Performance of individual data sources

	CORR	RMSE (%ILI)	Rel RMSE (%)	RMAE (%)	Hit Rate
FNY	0.948	0.385	15.9	39.3	65.9
ATH	0.977	0.351	14.1	36.7	**77.7**
GT	0.978	0.245	13.3	42.9	65.9
GFT	0.980	0.333	12.3	35.3	75.3
TWT	0.937	0.414	15.1	50.1	62.4
CDC Baseline	0.930	0.501	18.2	46.7	68.2
CDC Virology	0.923	-	-	-	69.4

Santillana et al. PLoS Computational Biology, 2015

Performance ensemble

	CORR	RMSE (%ILI)	Rel RMSE (%)	RMAE (%)	Hit Rate
FNY	0.948	0.385	15.9	39.3	65.9
ATH	0.977	0.351	14.1	36.7	**77.7**
GT	0.978	0.245	13.3	42.9	65.9
GFT	0.980	0.333	12.3	35.3	75.3
TWT	0.937	0.414	15.1	50.1	62.4
CDC Baseline	0.930	0.501	18.2	46.7	68.2
CDC Virology	0.923	-	-	-	69.4
SVM (RBF)	**0.989**	**0.176**	**8.27**	**23.6**	69.4

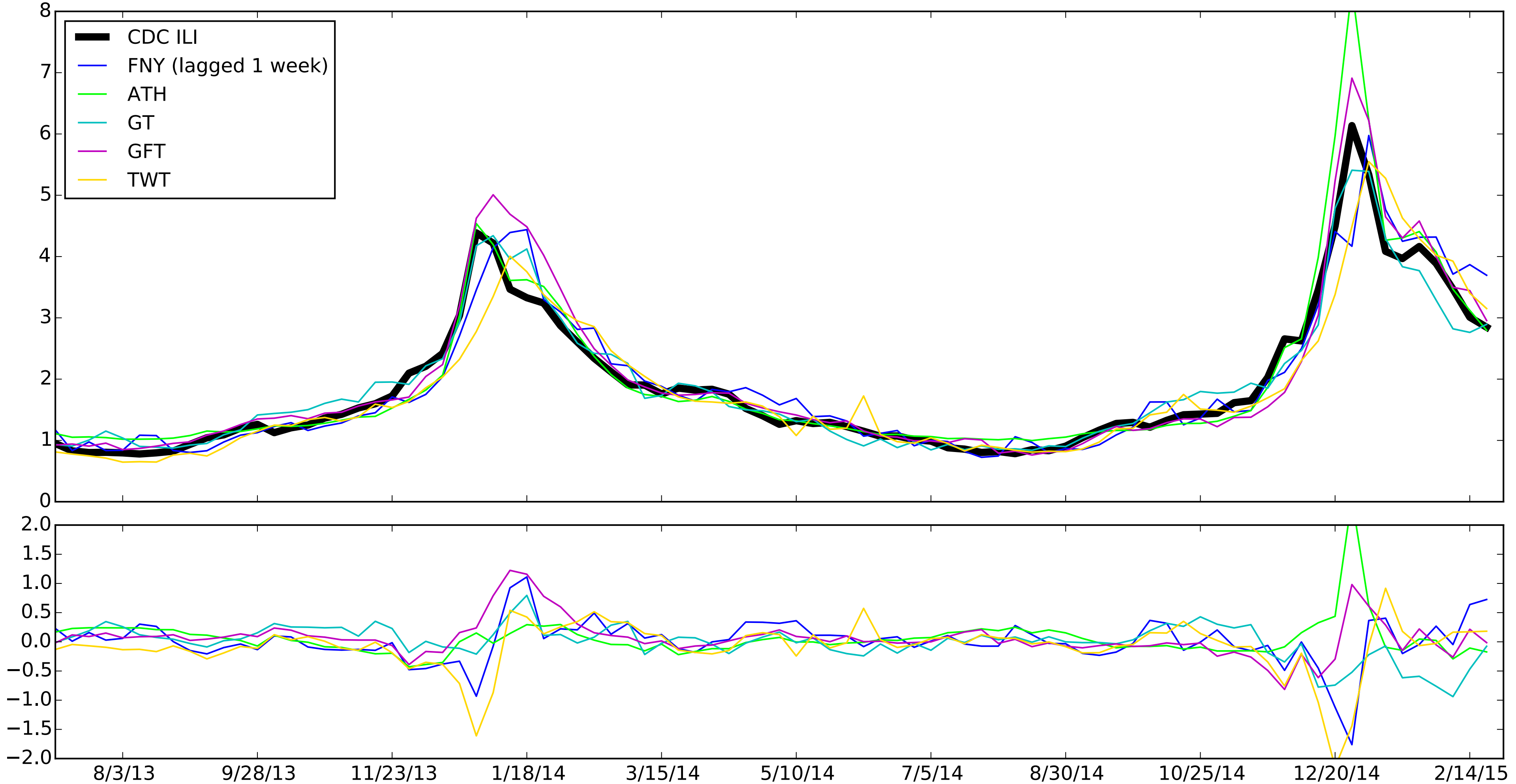
Santillana et al. PLoS Computational Biology, 2015

Performance ensemble

	CORR	RMSE (%ILI)	Rel RMSE (%)	RMAE (%)	Hit Rate
FNY	0.948	0.385	15.9	39.3	65.9
ATH	0.977	0.351	14.1	36.7	**77.7**
GT	0.978	0.245	13.3	42.9	65.9
GFT	0.980	0.333	12.3	35.3	75.3
TWT	0.937	0.414	15.1	50.1	62.4
CDC Baseline	0.930	0.501	18.2	46.7	68.2
CDC Virology	0.923	-	-	-	69.4
SVM (RBF)	**0.989**	**0.176**	**8.27**	**23.6**	69.4

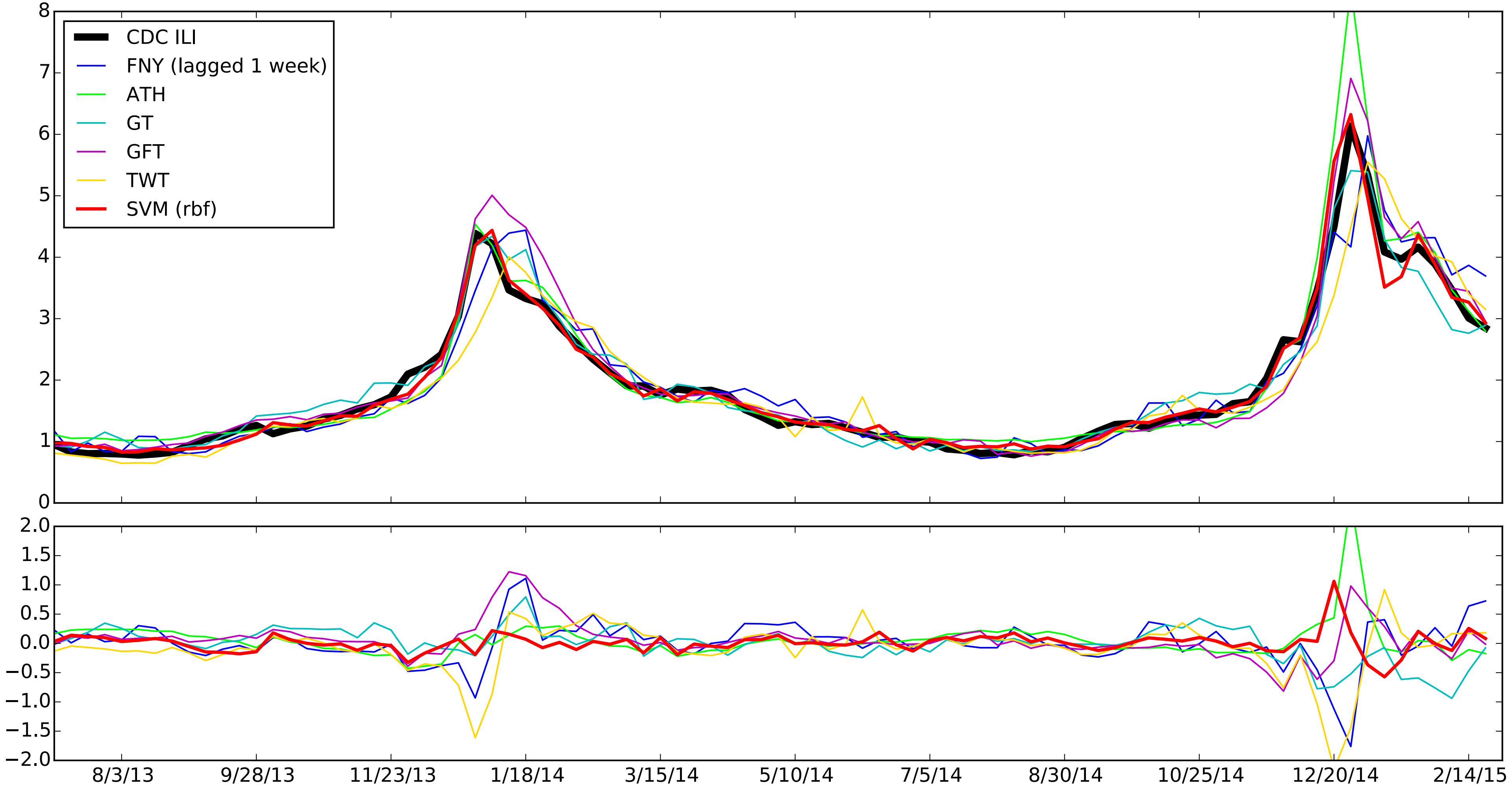
Santillana et al. PLoS Computational Biology, 2015

Performance of individual data sources



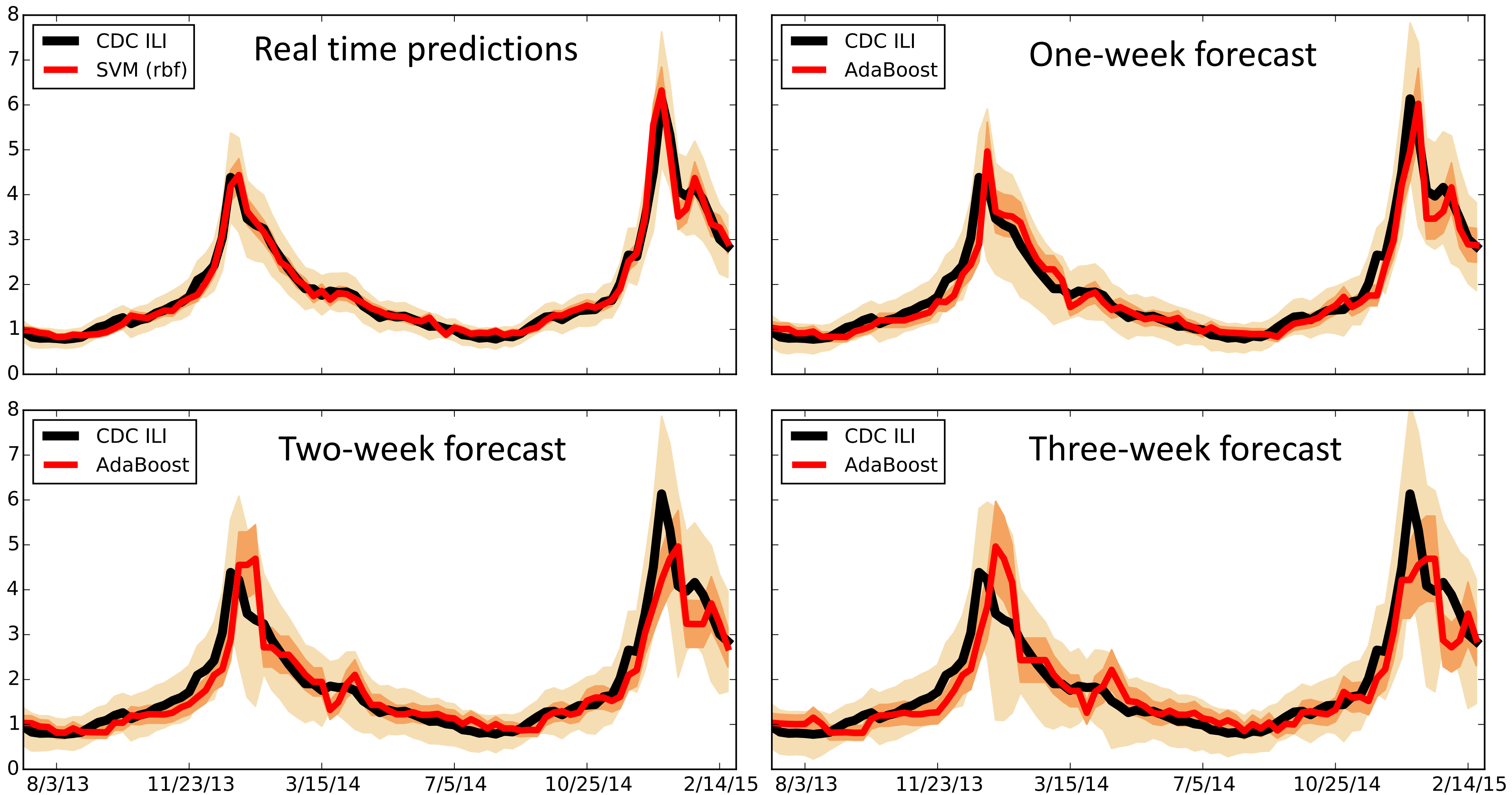
Santillana et al. PLoS Computational Biology, 2015

Performance ensemble



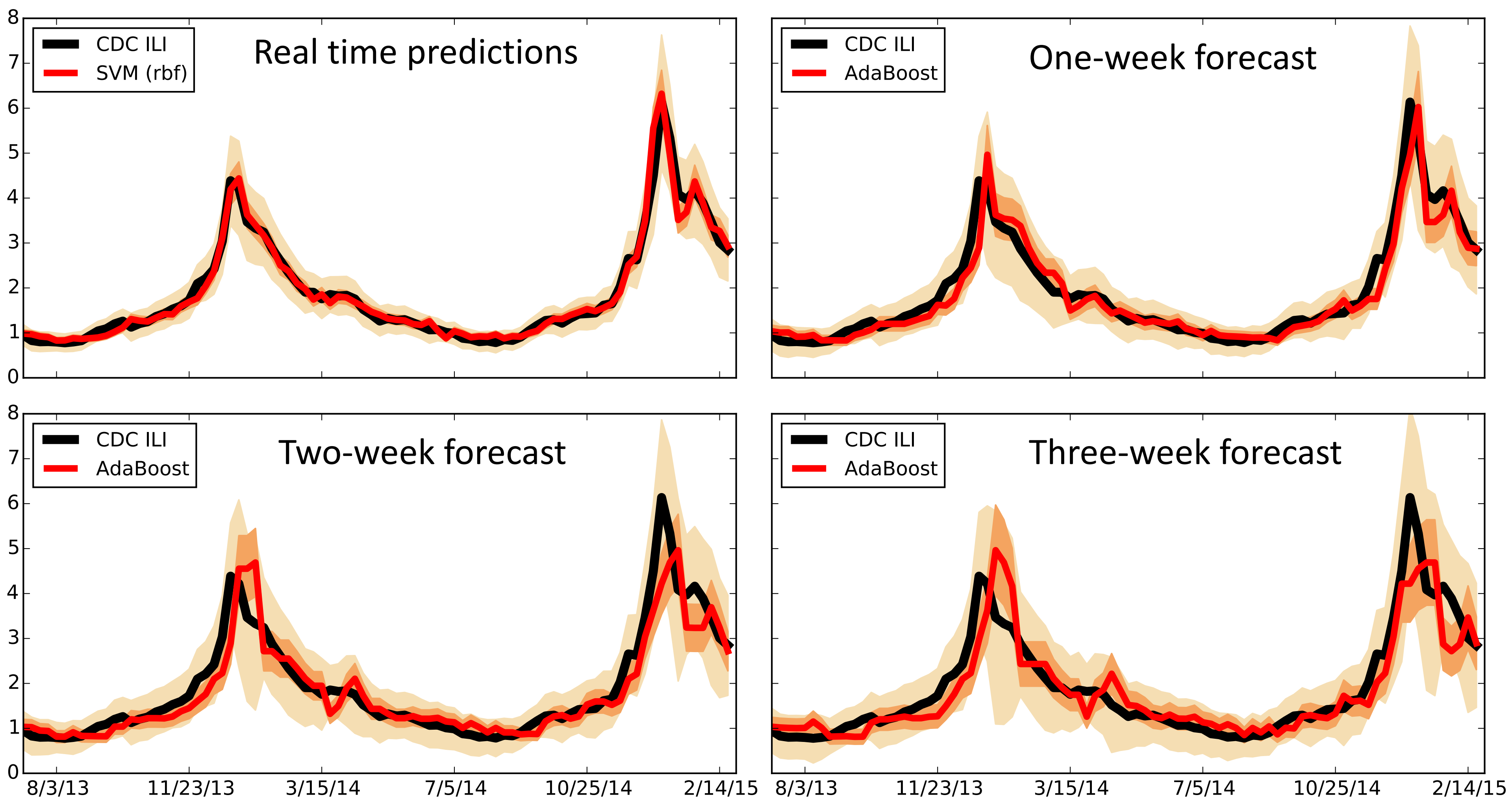
Santillana et al. PLoS Computational Biology, 2015

Real time predictions and Forecasts



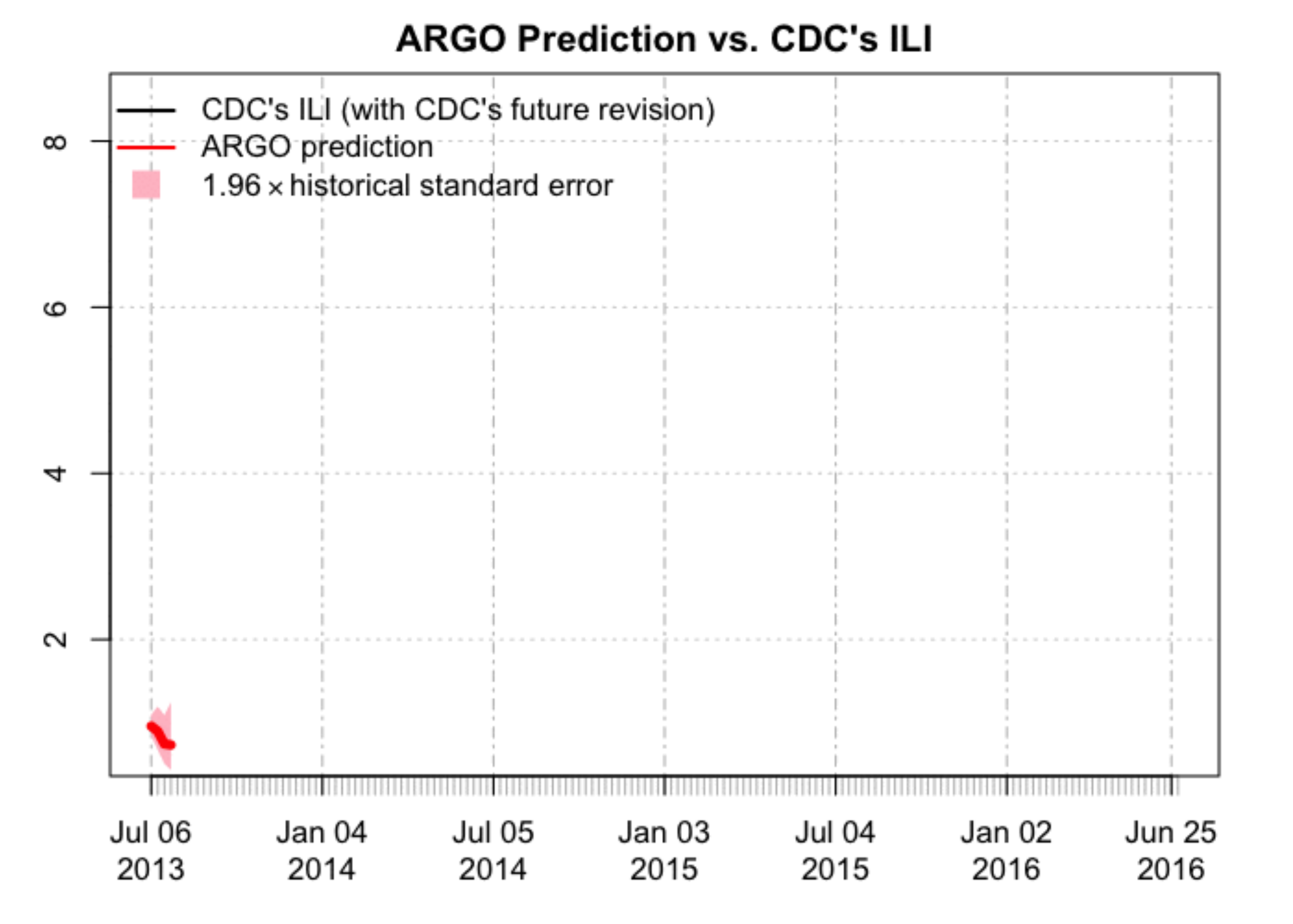
Santillana et al. PLoS Computational Biology, 2015

Real time predictions and Forecasts

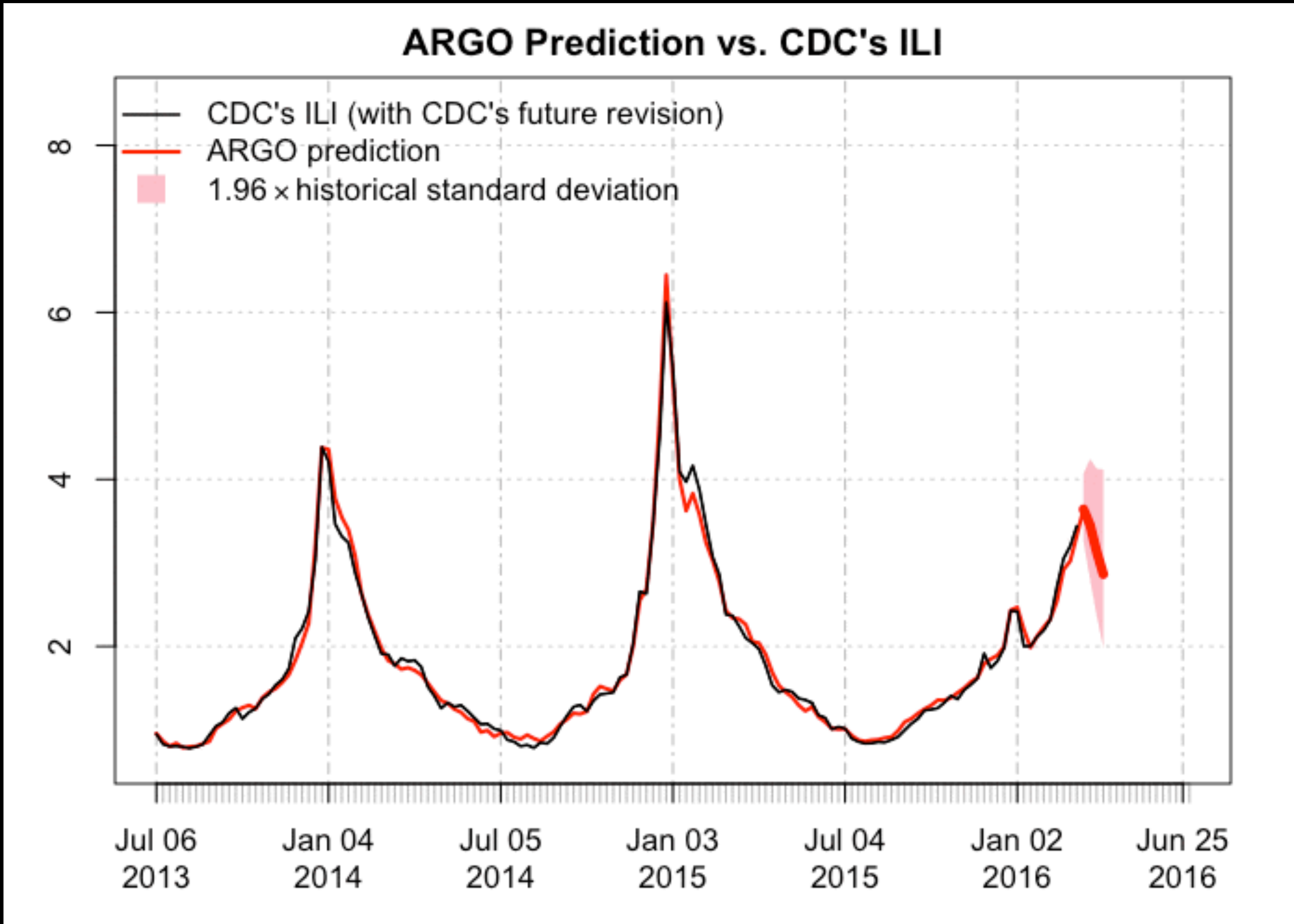


Santillana et al. PLoS Computational Biology, 2015

We are launching the next generation of digital disease surveillance tools



Video produced by Shihao Yang

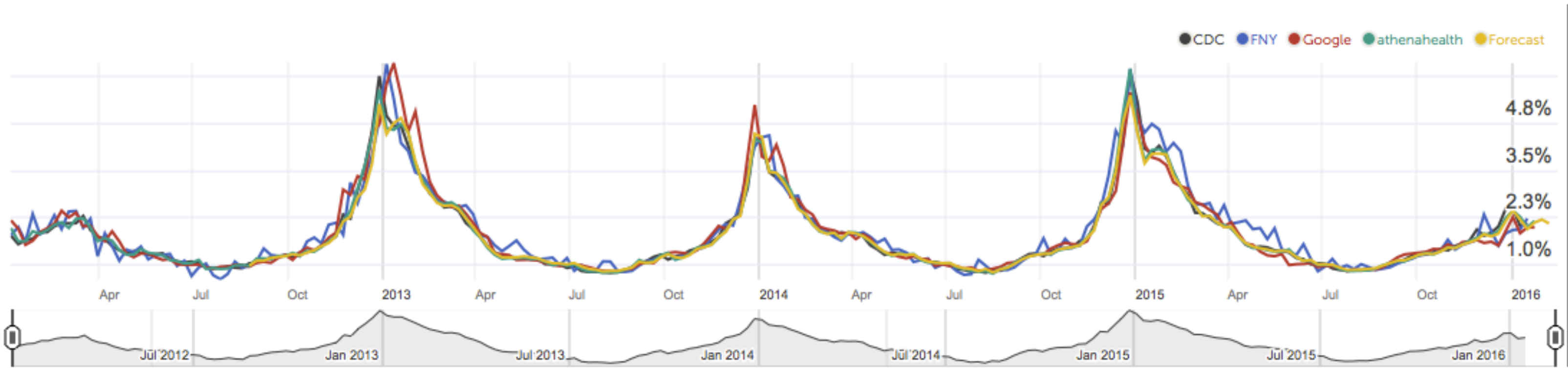


ARGO Prediction vs. CDC's ILI
CDC's ILI (with CDC's future revision)
ARGO prediction
1.96 × historical standard deviation
2
4
6
8
Jul 06 2013
Jan 04 2014
Jul 05 2014
Jan 03 2015
Jul 04 2015
Jan 02 2016
Jun 25 2016

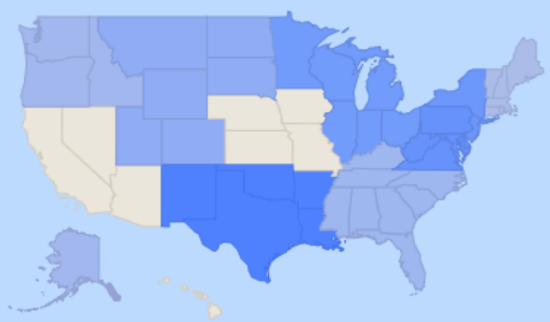
HealthMap Flu Trends

About

United States ▾ | Time range ▾ | Flu ▾ | Forecast 2.2% this week 2.1% next week



☑ CDC ☑ FNY ☑ Google ☑ athenahealth ☑ Forecast

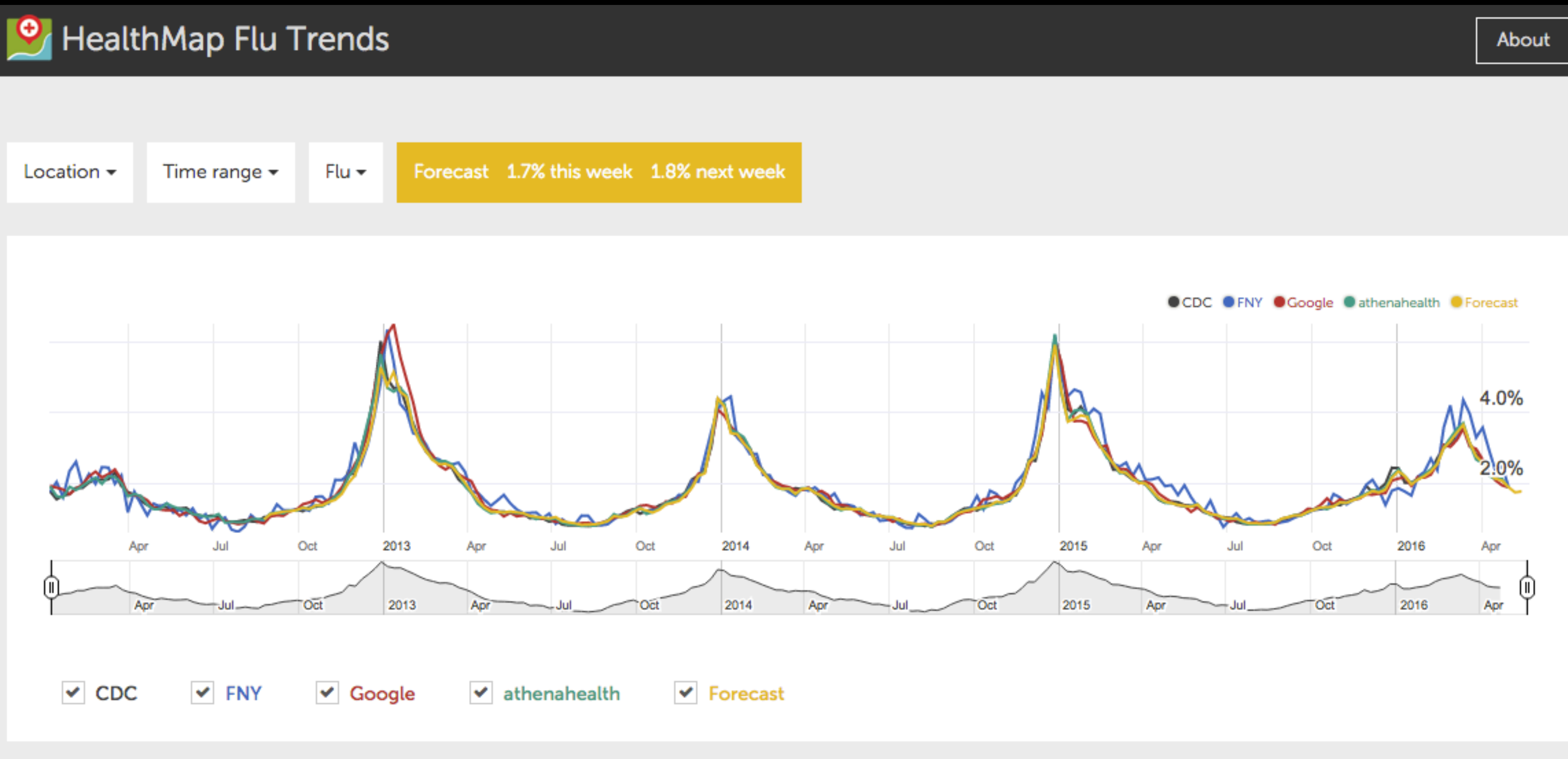


Thanks to Sue Aman, Rachel Chorney, Jeff Andre, Andre Nguyen, John Brownstein and Healthmap team!

HealthMap

Boston Children's Hospital
Until every child is well

HARVARD
MEDICAL SCHOOL



HealthMap Flu Trends
About
Location
Time range
Flu
Forecast 1.7% this week 1.8% next week
CDC
FNY
Google
athenahealth
Forecast
4.0%
2.0%
Apr
Jul
Oct
2013
2014
2015
2016
CDC
FNY
Google
athenahealth
Forecast

HealthMap Flu Trends

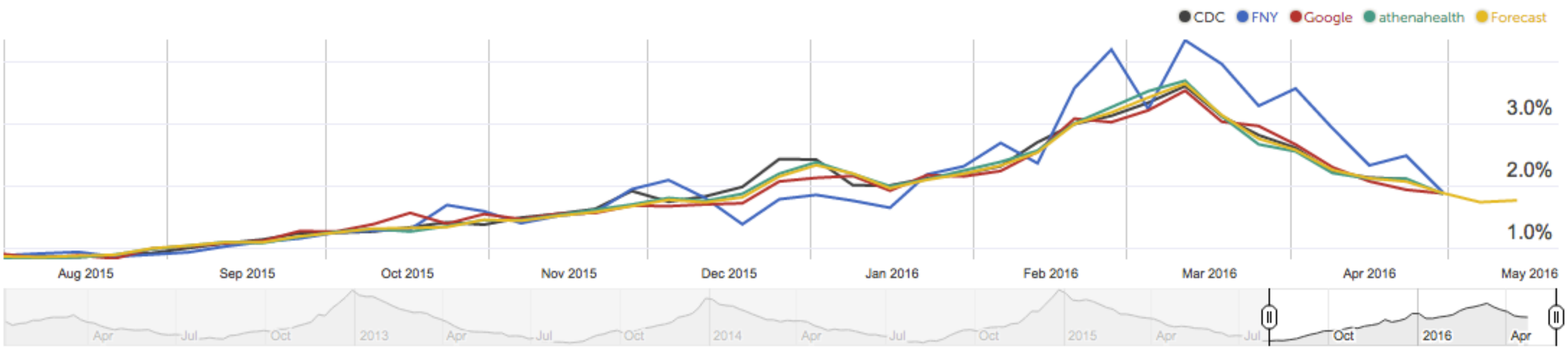
About

Location

Forecast

Flu

Forecast 1.7% this week 1.8% next week

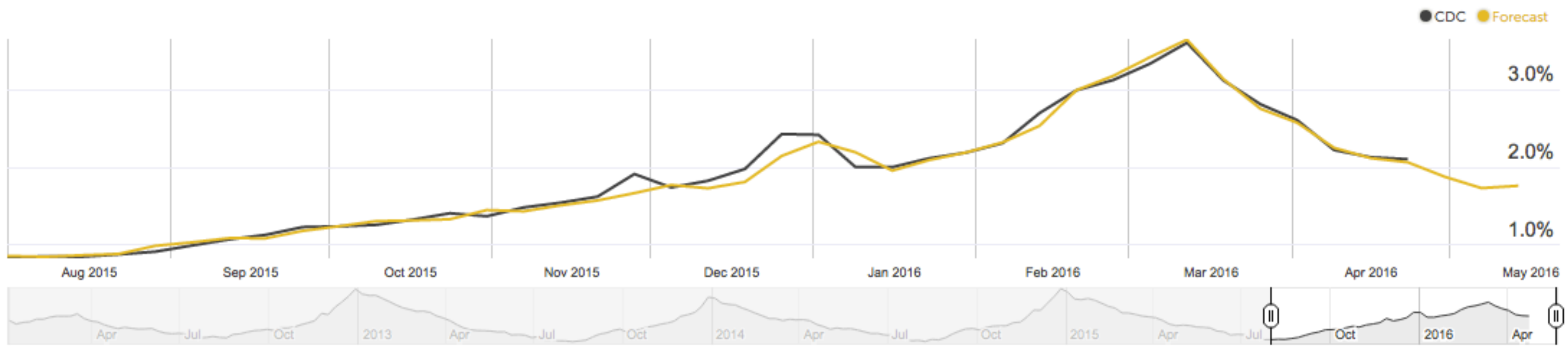


CDC FNY Google athenahealth Forecast

HealthMap Flu Trends

About

Location ▾ | Forecast | Flu ▾ | Forecast 1.7% this week 1.8% next week

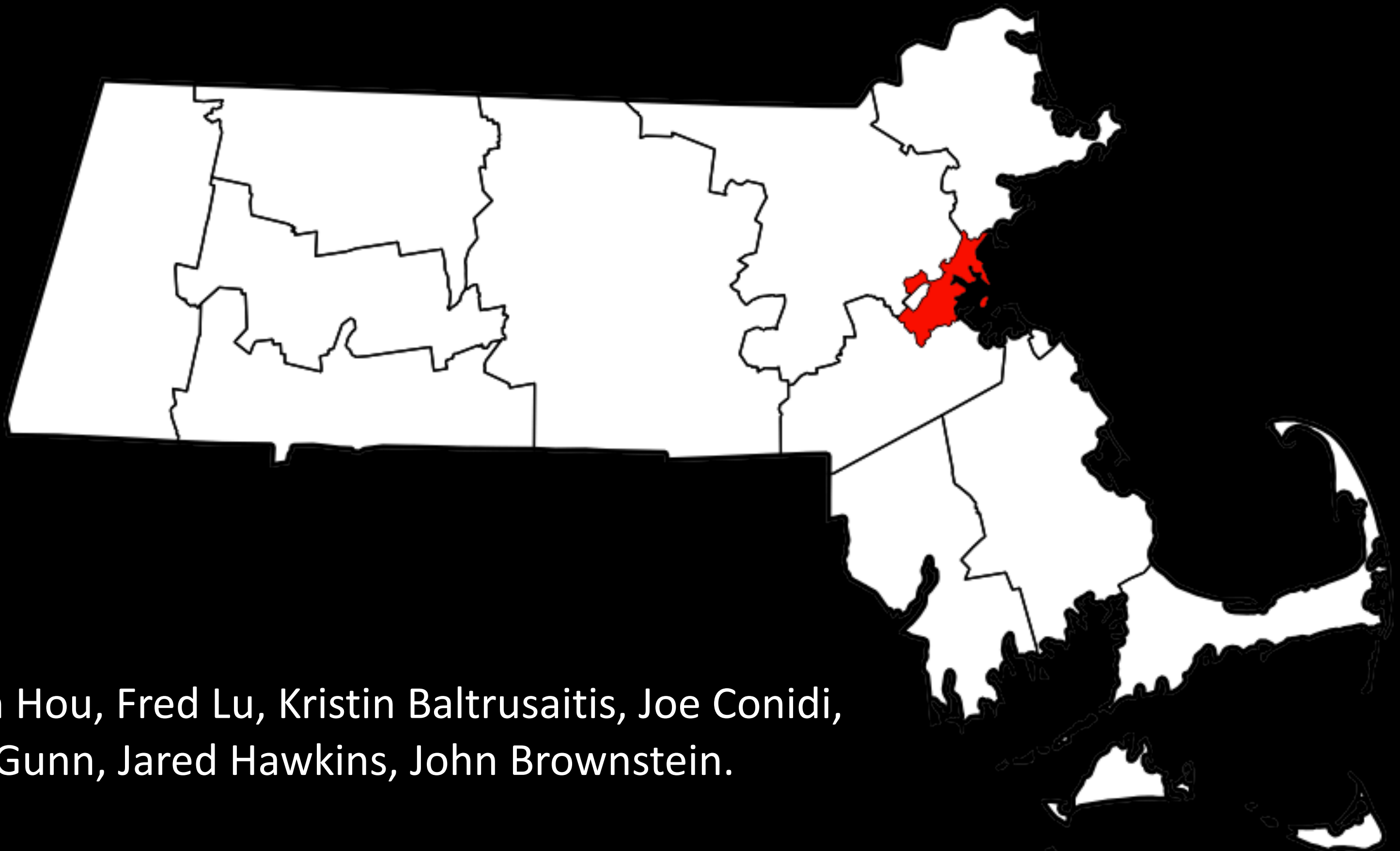


☑ CDC ☐ FNY ☐ Google ☐ athenahealth ☑ Forecast

We will extend out methodology to finer spatial resolutions. Pilot projects:

1. State level: Massachusetts

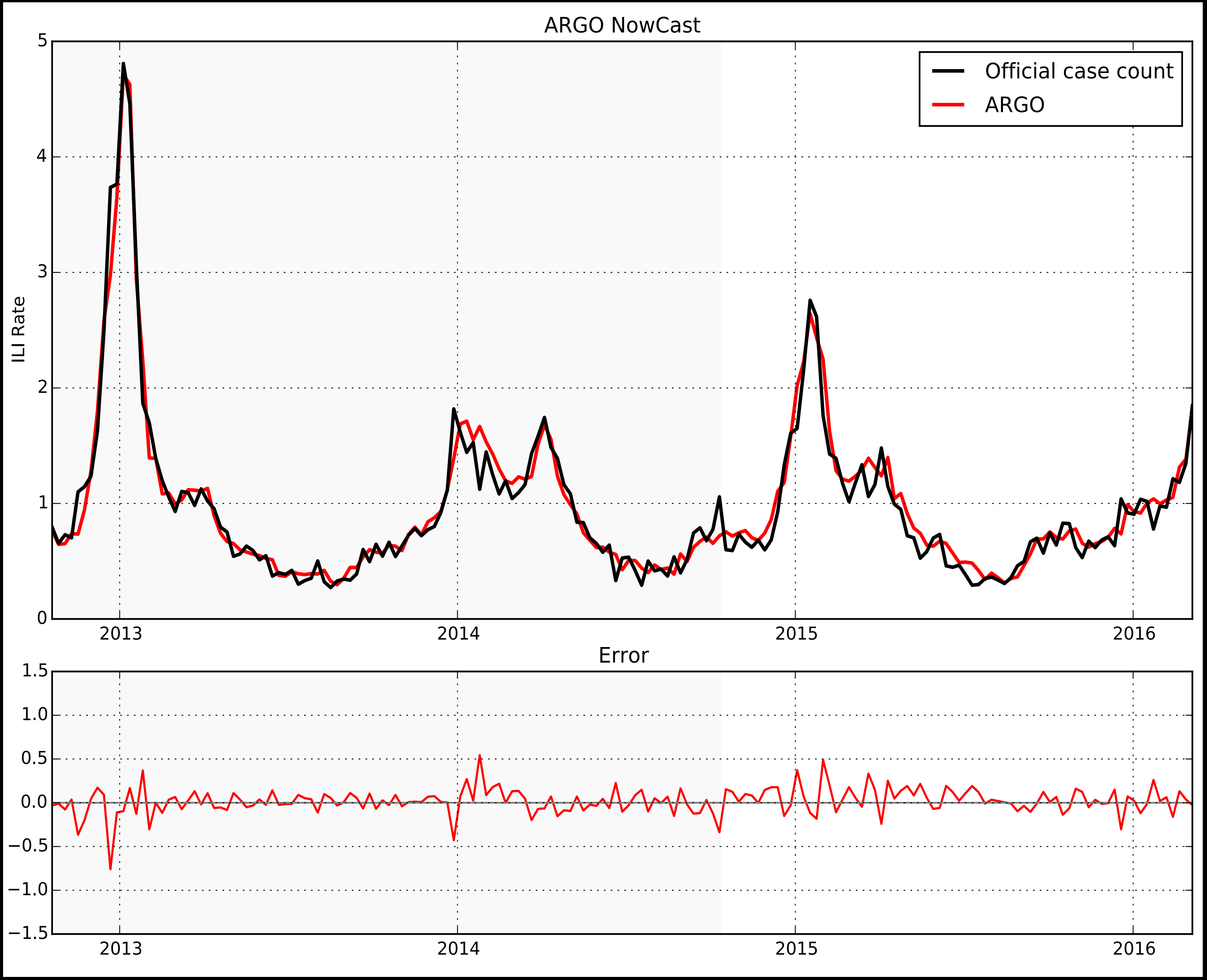
1. City level: Boston



With:

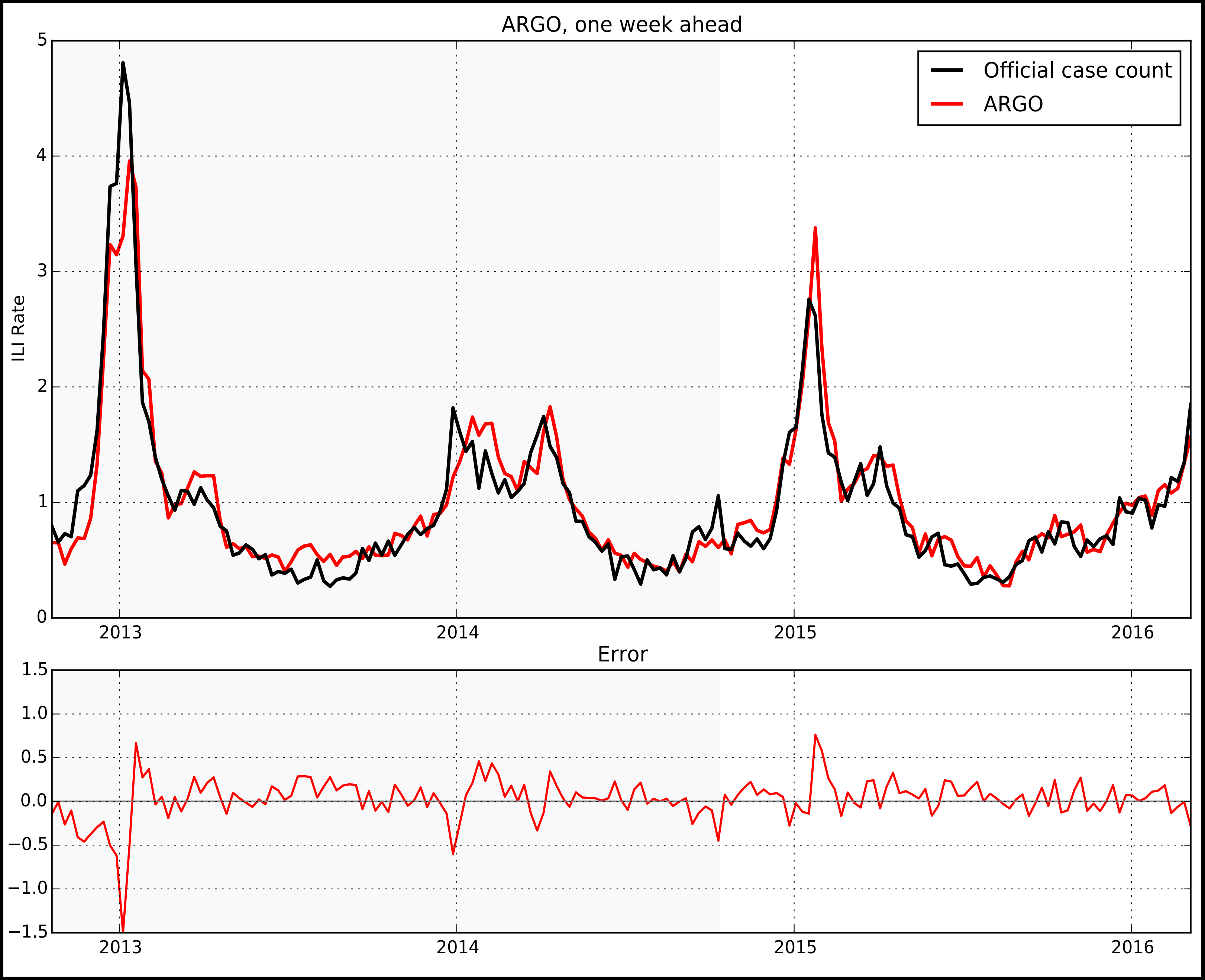
Suqin Hou, Fred Lu, Kristin Baltrusaitis, Joe Conidi, Julia Gunn, Jared Hawkins, John Brownstein.

Using multiple data sources to track flu in **Boston**



Plots produced by Fred Lu and Suqin Hou

Using multiple data sources to **forecast** flu in **Boston**

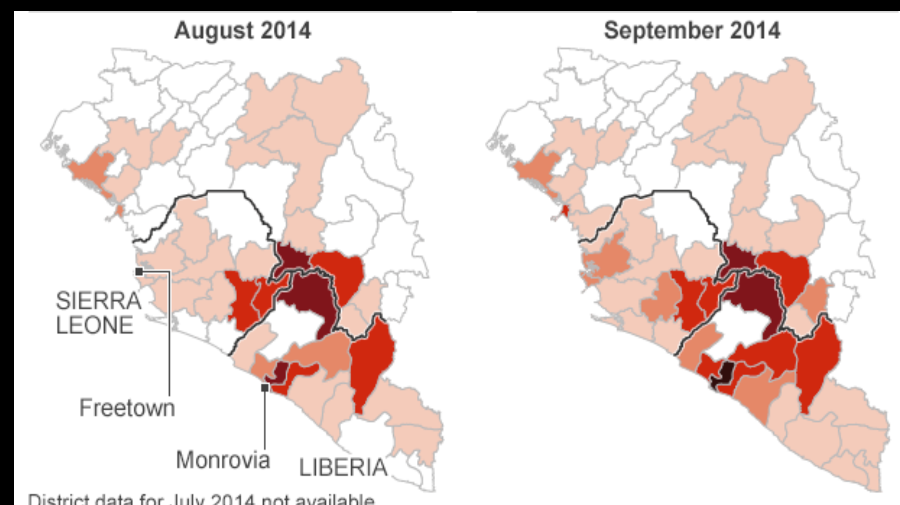


Plots produced by Fred Lu and Suqin Hou

Beyond flu…

Dengue and Zika

Ebola



OPEN ACCESS Freely available online

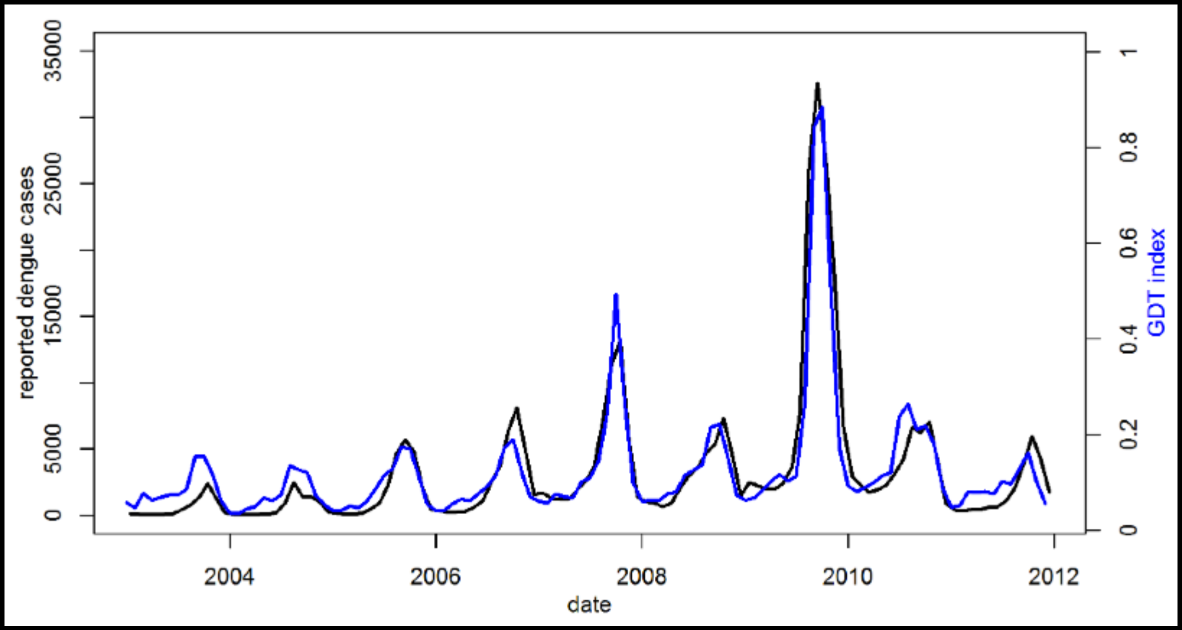
PLOS | NEGLECTED TROPICAL DISEASES

Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends

Rebecca Tave Gluskin[1]*, Michael A. Johansson[2], Mauricio Santillana[3], John S. Brownstein[1]

1 Children's Hospital Informatics Program, Children's Hospital Boston, Boston, Massachusetts, United States of America, **2** Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico, **3** School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

While Google Dengue Trends captures well the national incidence of disease



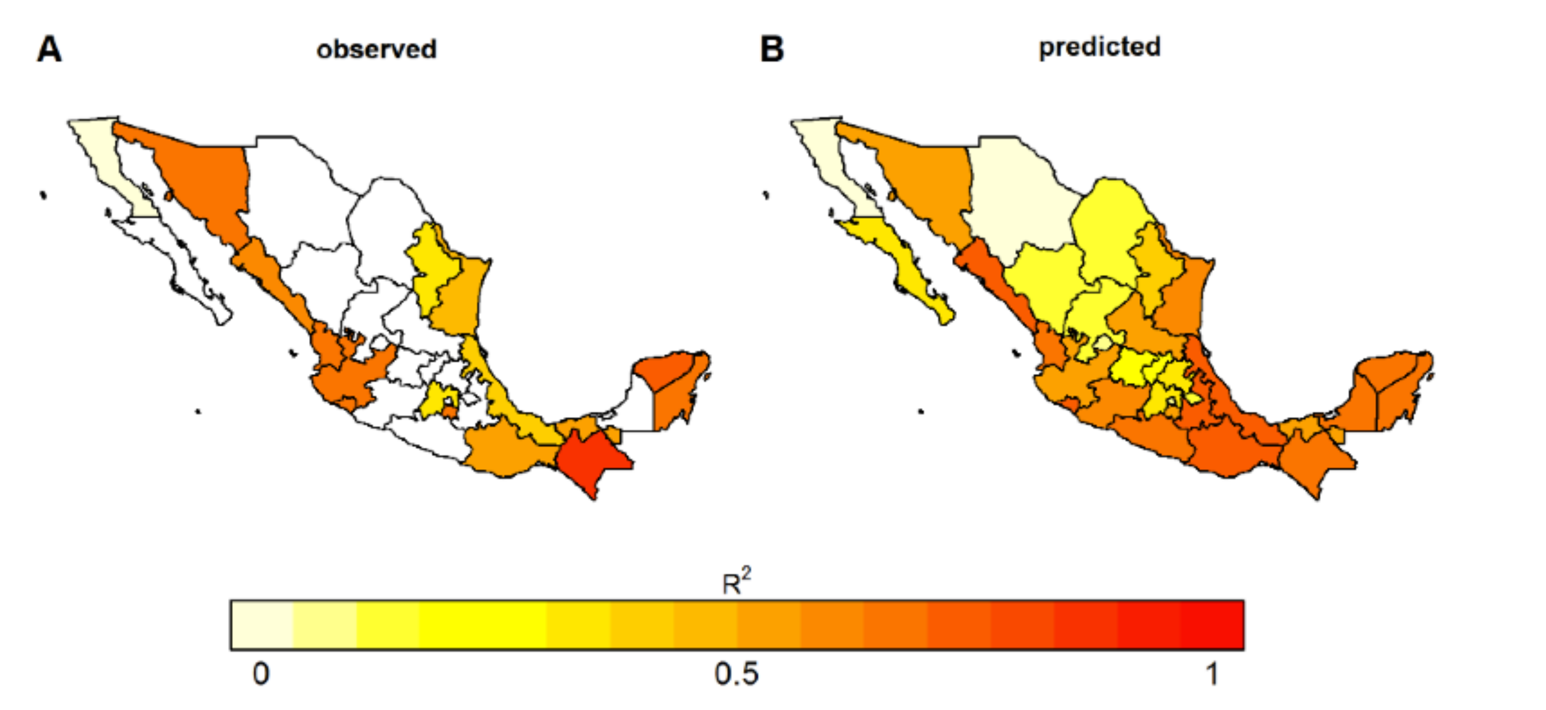
OPEN ACCESS Freely available online
PLOS | NEGLECTED TROPICAL DISEASES

Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends

Rebecca Tave Gluskin[1]*, Michael A. Johansson[2], Mauricio Santillana[3], John S. Brownstein[1]

1 Children's Hospital Informatics Program, Children's Hospital Boston, Boston, Massachusetts, United States of America, **2** Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico, **3** School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

It fails to captures the incidence of dengue at the state level in multiple cases



OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends

Rebecca Tave Gluskin[1]*, Michael A. Johansson[2], Mauricio Santillana[3], John S. Brownstein[1]

1 Children's Hospital Informatics Program, Children's Hospital Boston, Boston, Massachusetts, United States of America, **2** Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico, **3** School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Interestingly, access to internet was not a good indicator of accuracy (challenging assumptions)

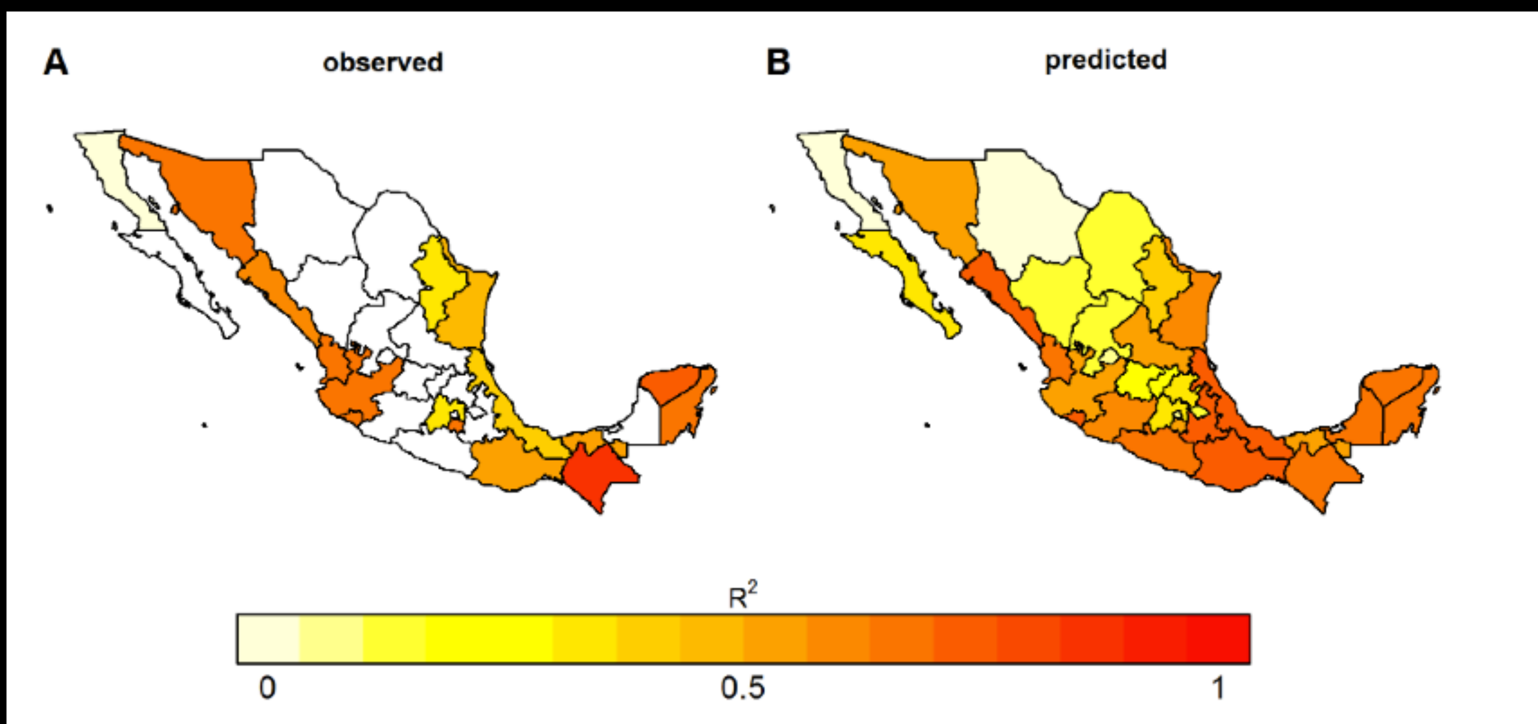
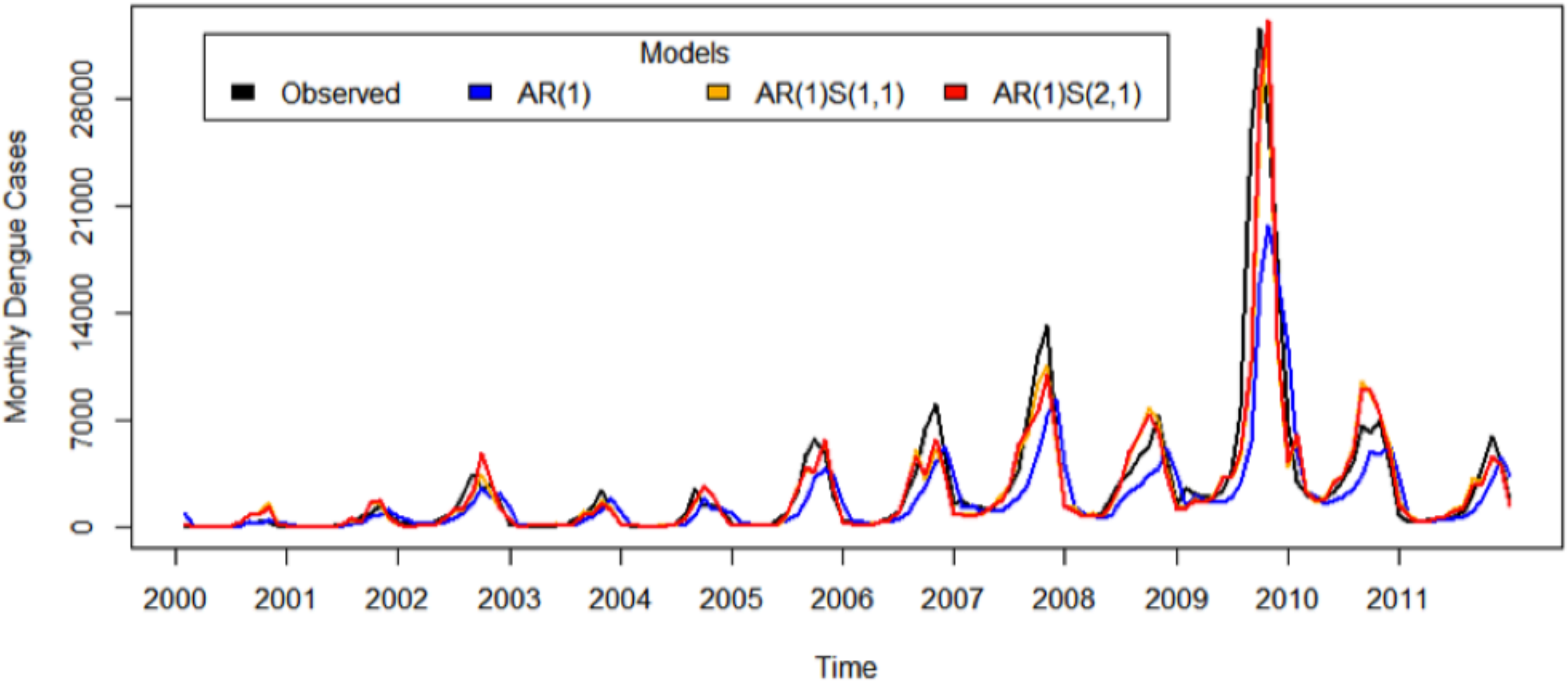


Table 1. Determinants of logit-transformed R^2 between Google Dengue Trends and government reported dengue cases: single covariate models.

	Coefficient	95% Confidence Interval	R^2	AIC[b]
Annual dengue cases[a]	**0.61**	**(0.36, 0.86)**	**0.67**	**43**
Minimum temperature	0.18	(−0.02, 0.37)	0.21	57
Mean temperature	**0.24**	**(0.01, 0.47)**	**0.26**	**56**
Maximum temperature	**0.28**	**(0.02, 0.55)**	**0.27**	**56**
Precipitation[a]	**1.6**	**(0.5, 2.6)**	**0.44**	**52**
Population[a]	−0.4	(−1.3, 0.6)	0.04	60
Population density[a]	−0.05	(−0.9, 0.81)	0	61
Percent youth	0.2	(−0.23, 0.63)	0.07	60
Doctors per 100 k residents	0.01	(−0.01, 0.03)	0.06	60
Potable water	−0.02	(−0.12, 0.07)	0.02	61
Municipal sewage	−0.01	(−0.09, 0.06)	0.01	61
Internet access	−0.06	(−0.15, 0.03)	0.12	59
Household income	−7.2E-05	(−14.5E-05, 0.1E-05)	0.24	57

[a]Log-transformed.
[b]Akaike information criterion.
doi:10.1371/journal.pntd.0002713.t001

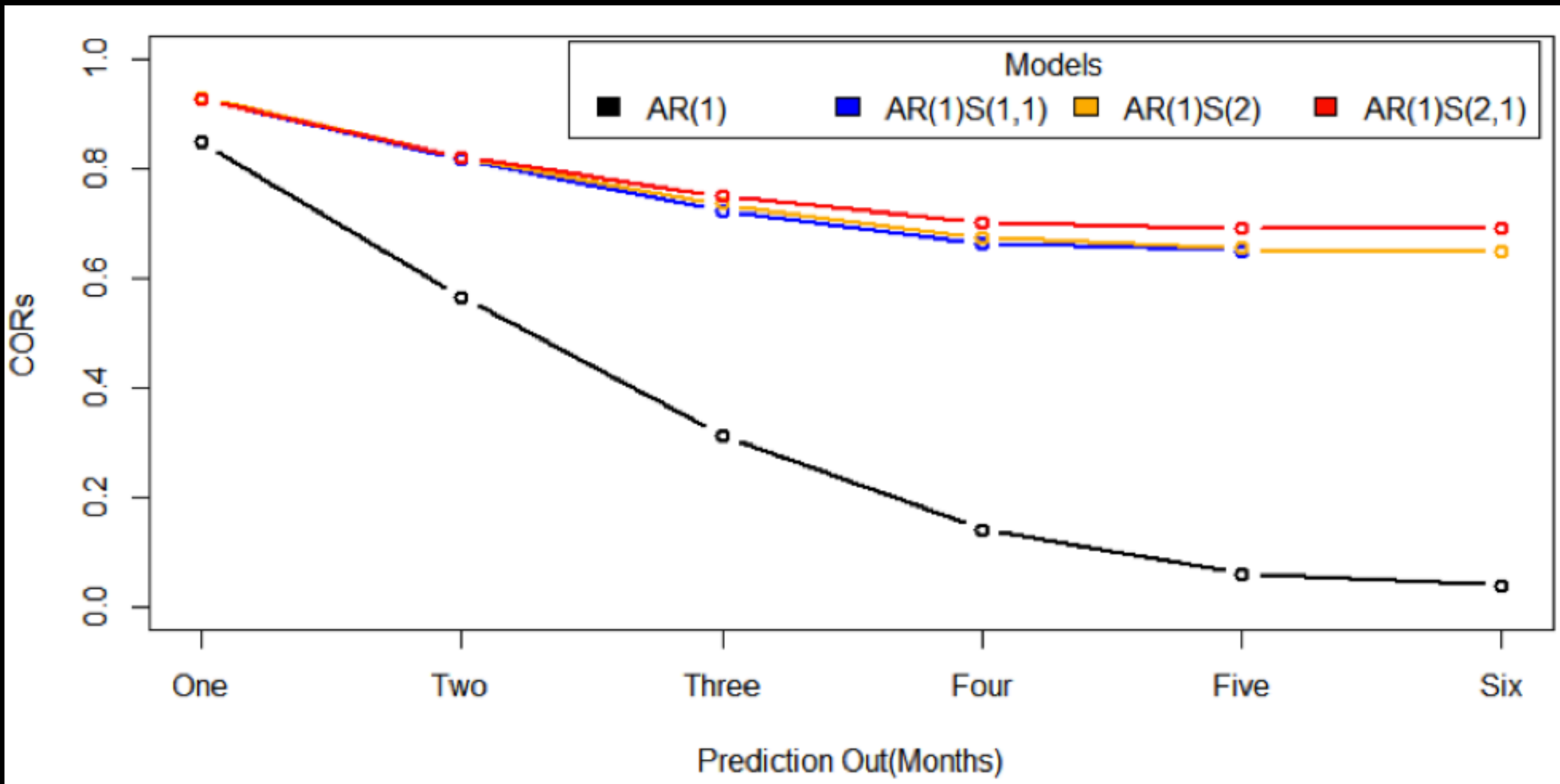
Mexico Dengue incidence (Country-level)



Jonansson et al. *Submitted*

Mexico Dengue incidence (Country-level)

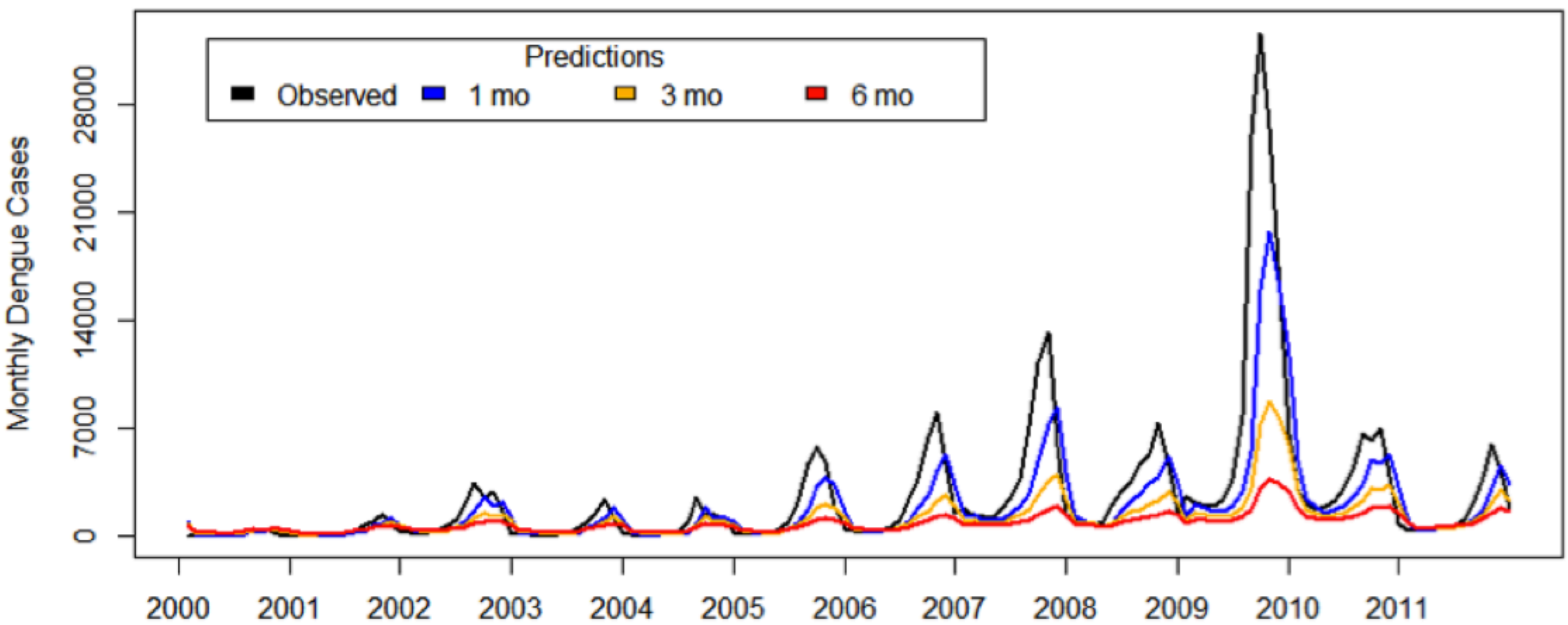
Accuracy of predictions decreases as time horizon grows



Jonansson et al. *Submitted*

Mexico Dengue incidence (Country-level)

Accuracy of predictions decreases as time horizon grows

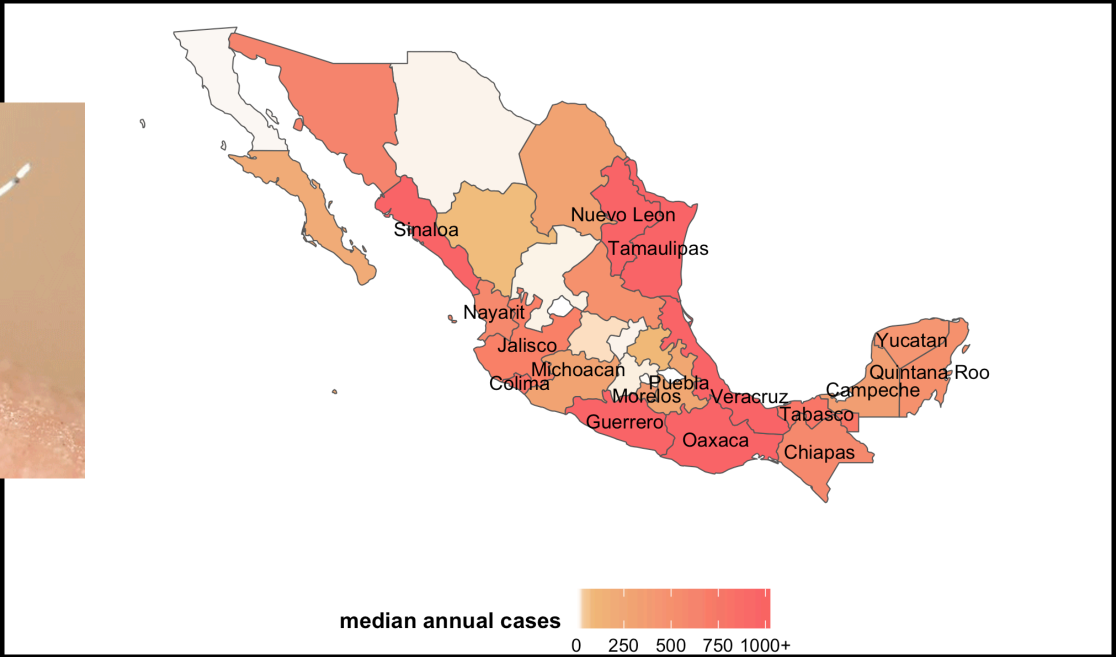


Predictions using an AR1 (constrained) model

Jonansson et al. *Submitted*

Forecasting Dengue Incidence in Mexico

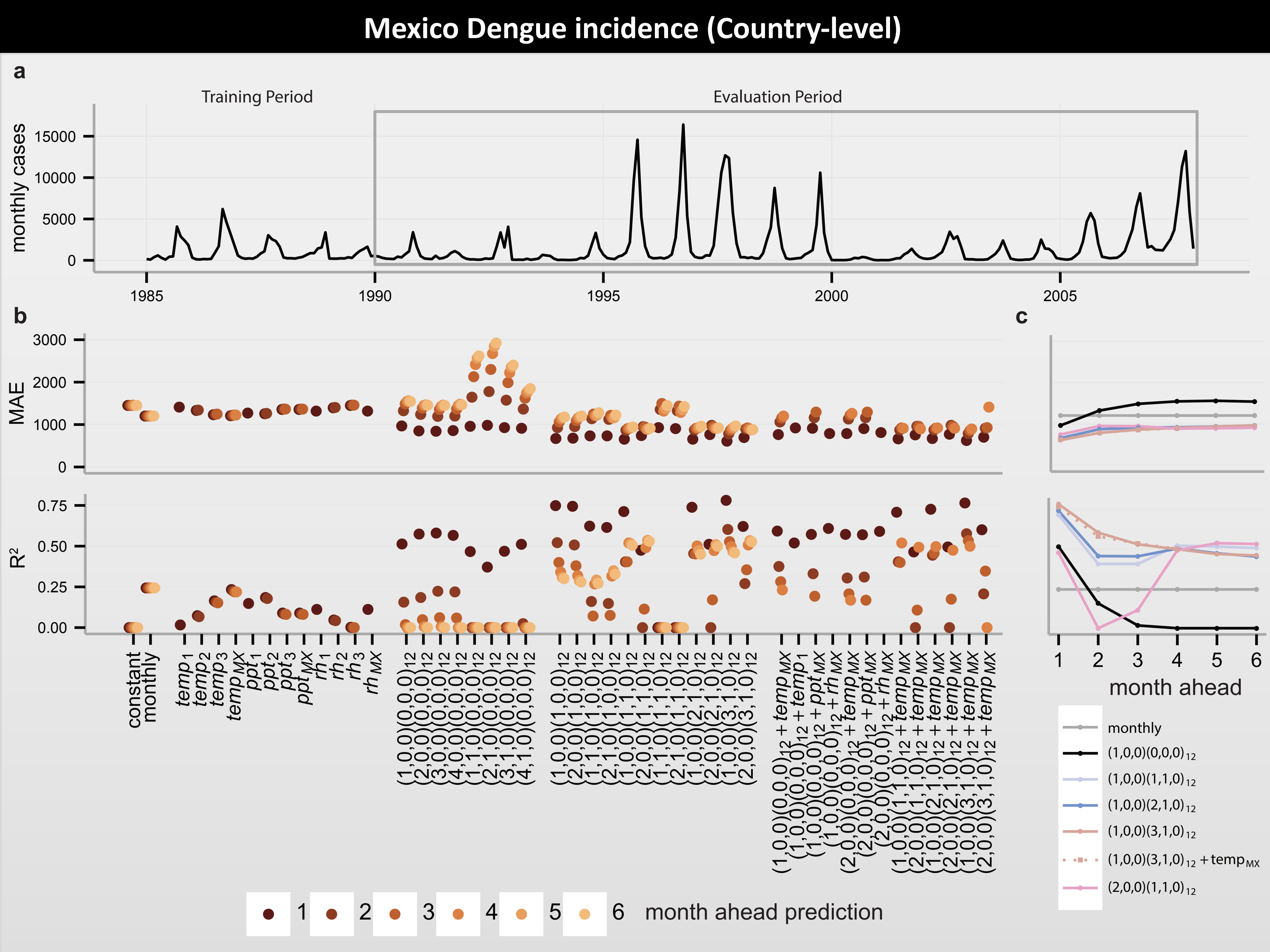
Establishing a prediction baseline



Team:

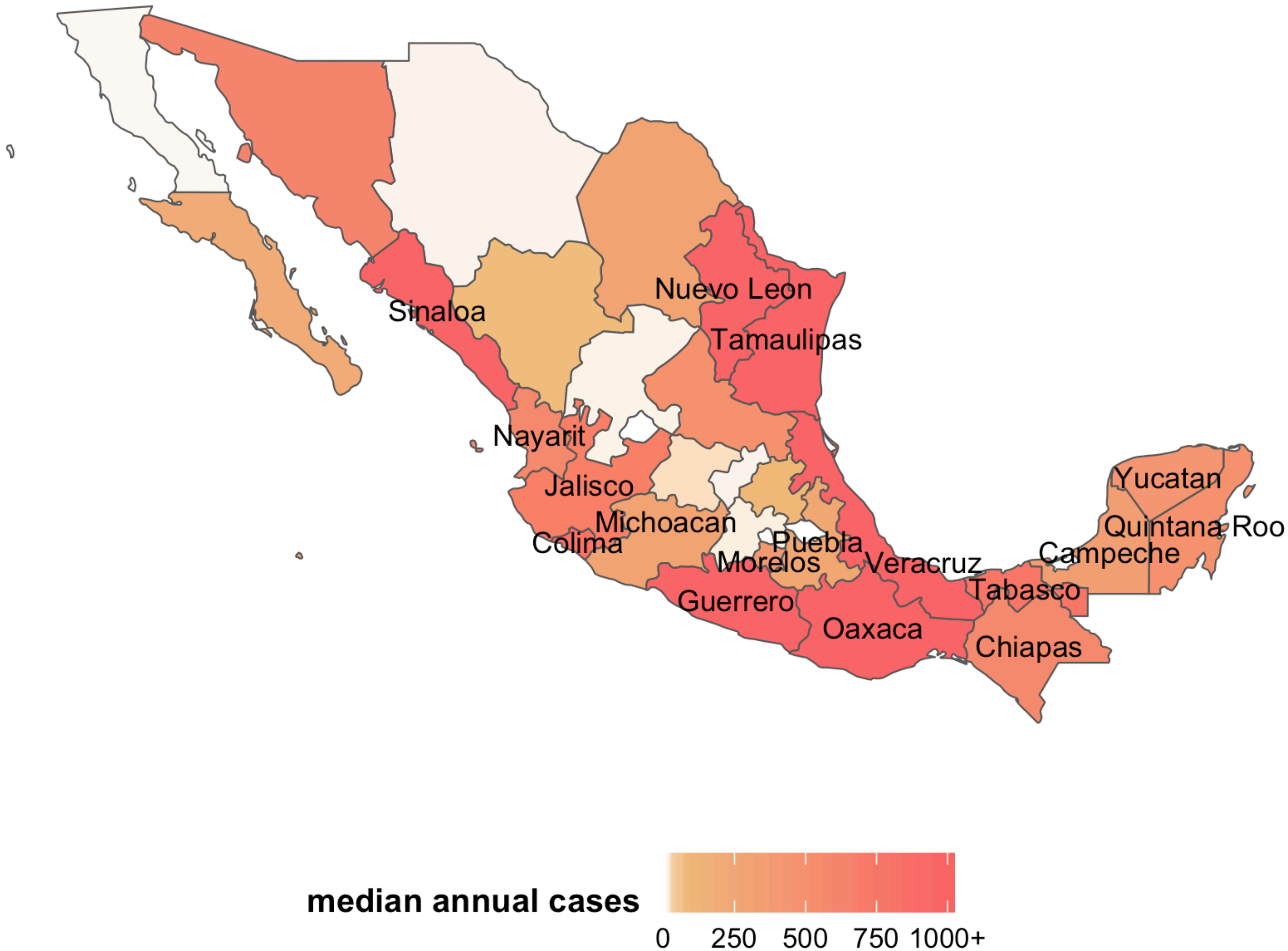
Mauricio Santillana (BCH, Harvard),
Michael Johansson (CDC Puerto Rico),
Aditi Hota (Columbia Univ),
John Brownstein (BCH, Harvard),
Nick Reich (Umass Amherts)

Establishing a baseline



Mexico Dengue incidence (Country-level)
a
Training Period
Evaluation Period
monthly cases
15000
10000
5000
0
1985
1990
1995
2000
2005
b
MAE
3000
2000
1000
0
R²
0.75
0.50
0.25
0.00
constant
monthly
c
month ahead
1 2 3 4 5 6
monthly
(1,0,0)(0,0,0)12
(1,0,0)(1,1,0)12
(1,0,0)(2,1,0)12
(1,0,0)(3,1,0)12
(1,0,0)(3,1,0)12 + tempMX
(2,0,0)(1,1,0)12
1 2 3 4 5 6 month ahead prediction

How about higher resolution geographically? Actionable information

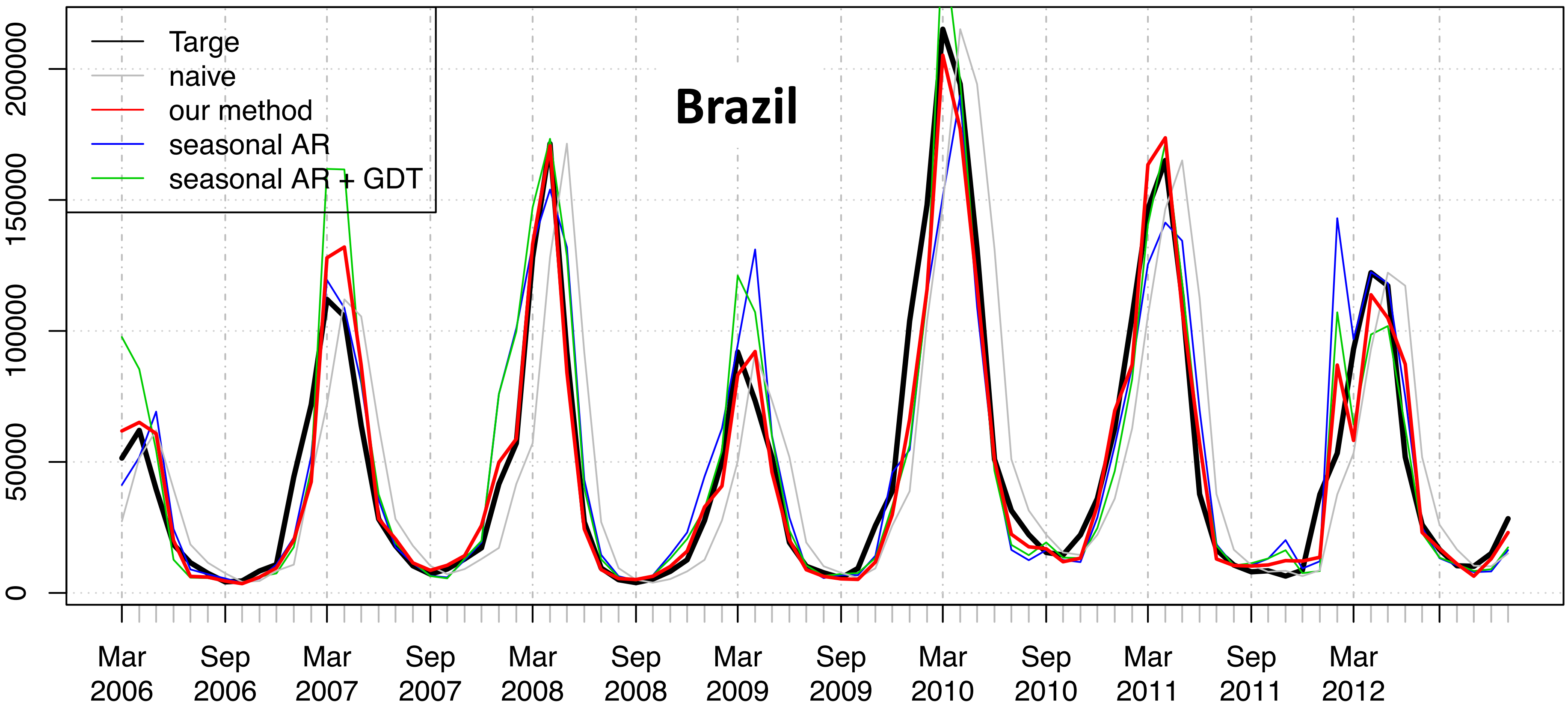


Jonansson et al. *In preparation*

Forecasting Dengue Incidence using multiple data sources

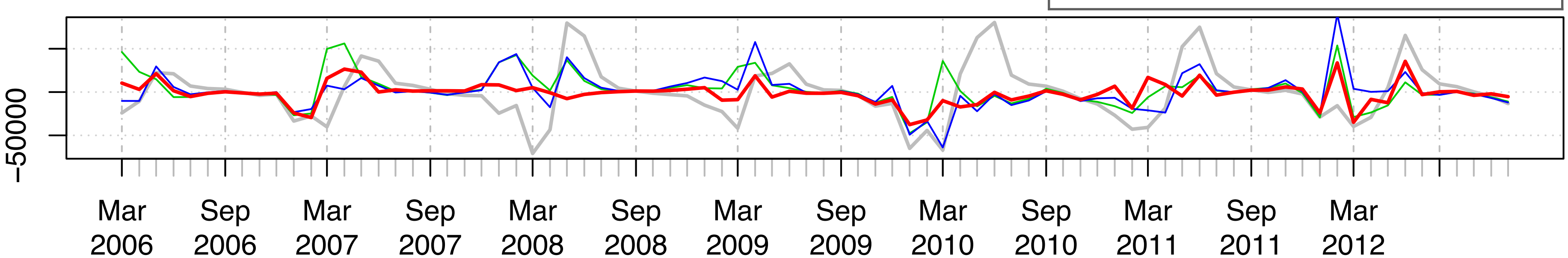
Break Dengue

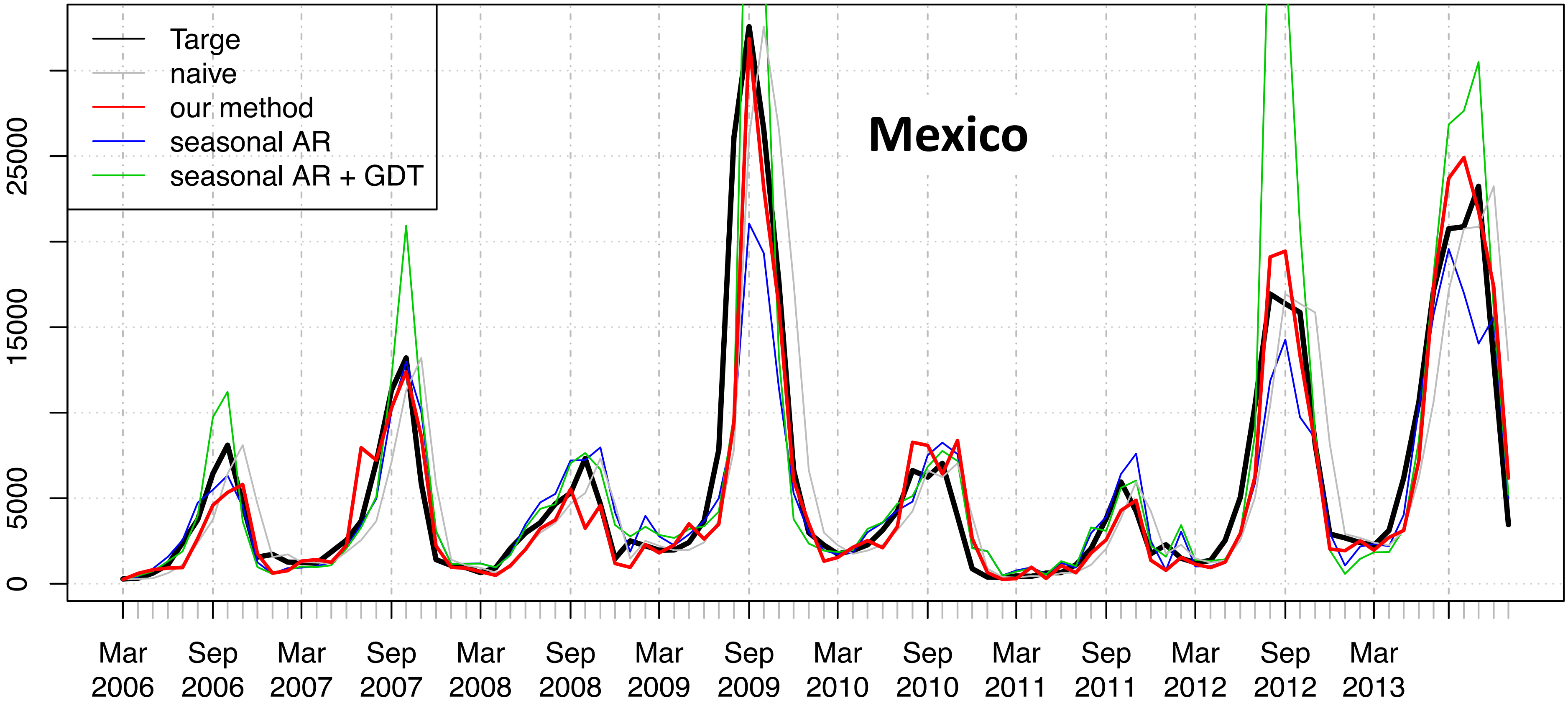
HealthMap



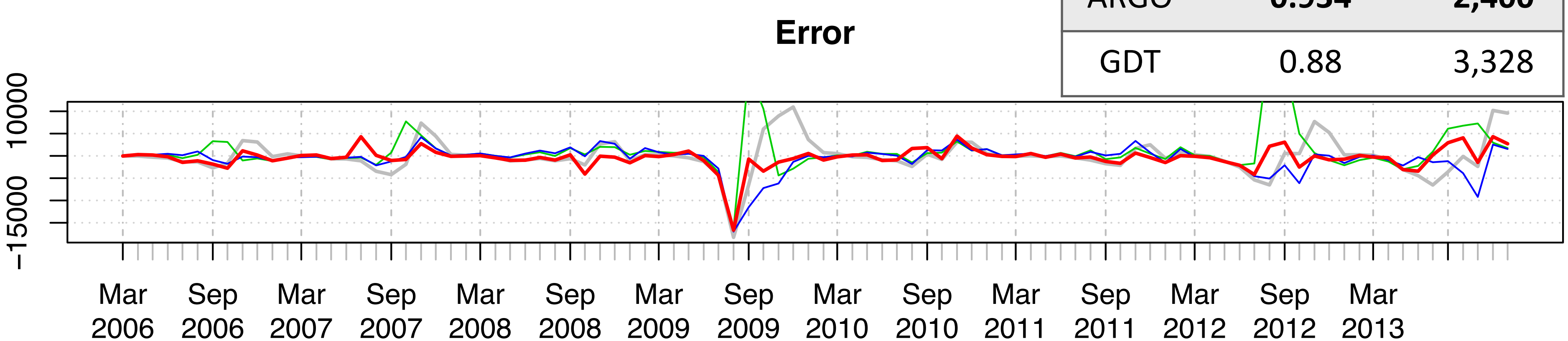
	Correlation	RMSE
ARGO	**0.966**	**12,884**
GDT	0.916	20,349

Error

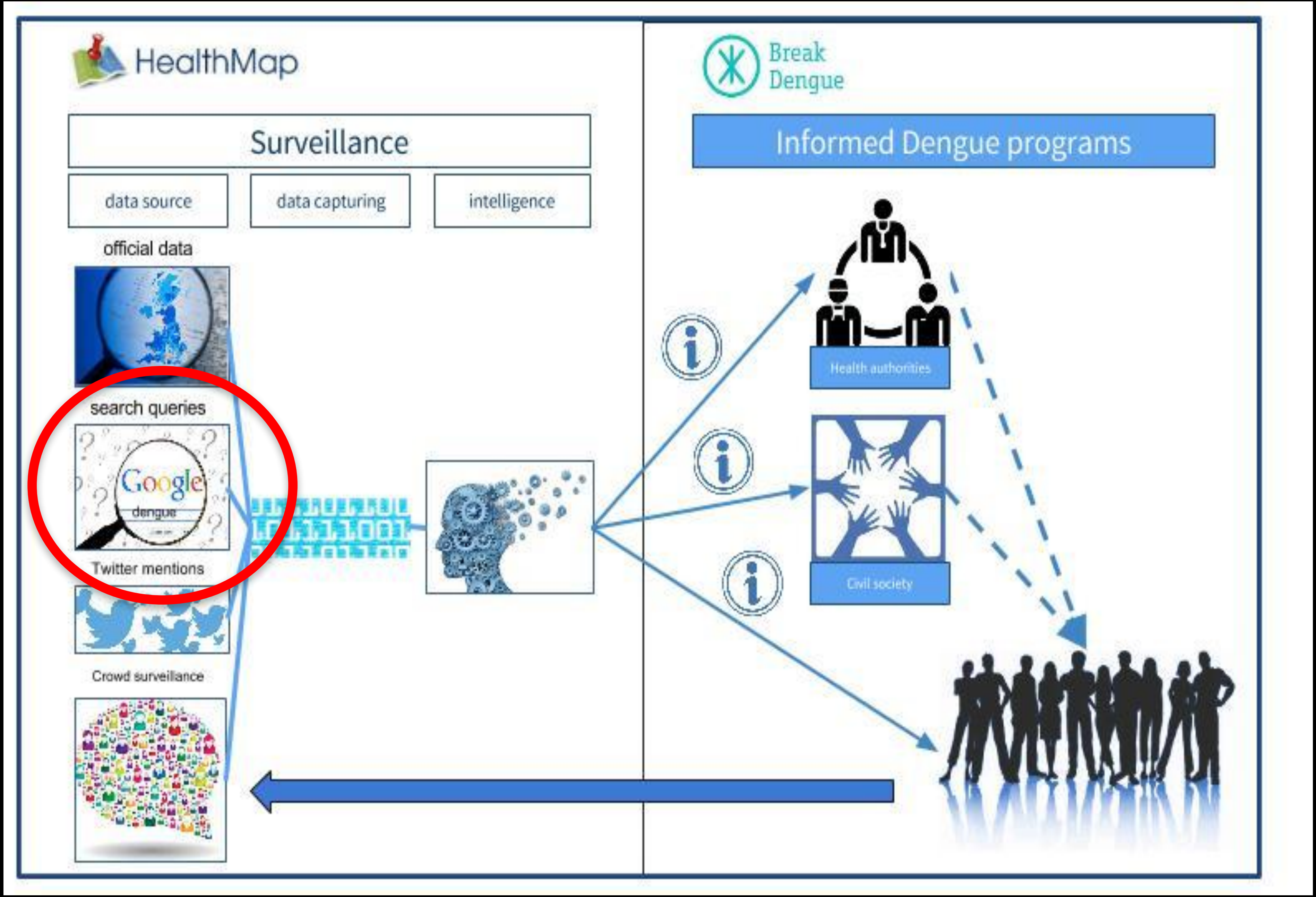




	Correlation	RMSE
ARGO	**0.934**	**2,400**
GDT	0.88	3,328



Partnerships are an essential element of this



CALL FOR PARTNERSHIPS

DEVELOPMENT OF REAL-TIME TRACKING TOOLS

TO MONITOR DENGUE WORLDWIDE – HEALTHMAP & BREAK DENGUE

NICHOLAS BROOKE & MAURICIO SANTILLANA

The global incidence of dengue is estimated at 390 million cases per year worldwide [1]. Endemic in many Asian and Latin American countries, dengue has become a leading cause of hospitalization and death among children in these regions [2] and contributes to substantial economic loss for governments and households [3]. Real-time dengue surveillance, therefore, is critical for identifying areas where transmission is ongoing or

methods to produce real-time and forecast estimates of flu incidence using Google search data [4,5,6], Twitter [7], Wikipedia [8], crowd-sourced participatory disease surveillance tools (such as Flu Near You) [9], and clinician's databases (such as UpToDate) [10]. These methods produce real-time estimates of flu activity that closely track the flu incidence as reported by traditional public health surveillance systems, such as Influen-

ple locations around the world by leveraging disparate data sources including: traditional clinical reporting systems, crowd sourced disease surveillance tools, internet-based services such as Google, Twitter and Wikipedia. In addition, our aim would be to empower local authorities with actionable information that may help them establish targeted alert systems to prevent further disease spread.

What if we could use **news reports** as a way to modulate predictions produced with models?

An example from the Ebola outbreak in 2015

Healthmap.org

Every hour

24/7

93

Public & private sources

15

Languages

100,000

Web

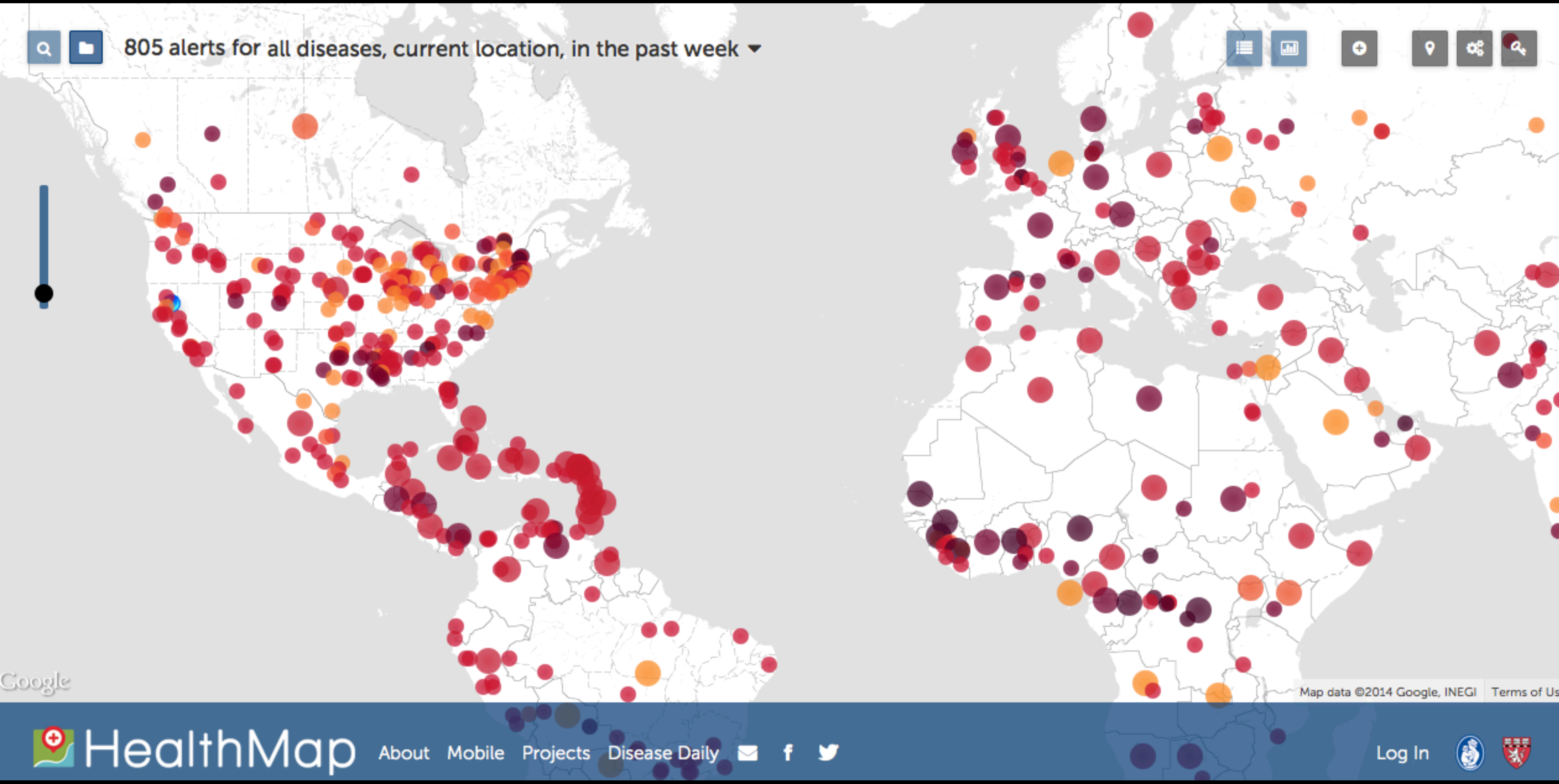
2,000

alerts per day, precisely placed in

10,000+

locations

HealthMap brings together disparate data sources, including online news aggregators, eyewitness reports, expert-curated discussions and validated official reports, to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health.



Through an automated process, updating 24/7/365, the system monitors, organizes, integrates, filters, visualizes and disseminates online information about emerging diseases in nine languages, facilitating early detection of global public health threats

HealthMap

About Mobile Projects Disease Daily

Google

Unilever

skoll global threats fund

USAID FROM THE AMERICAN PEOPLE

IARPA BE THE FUTURE

NLM

CDC

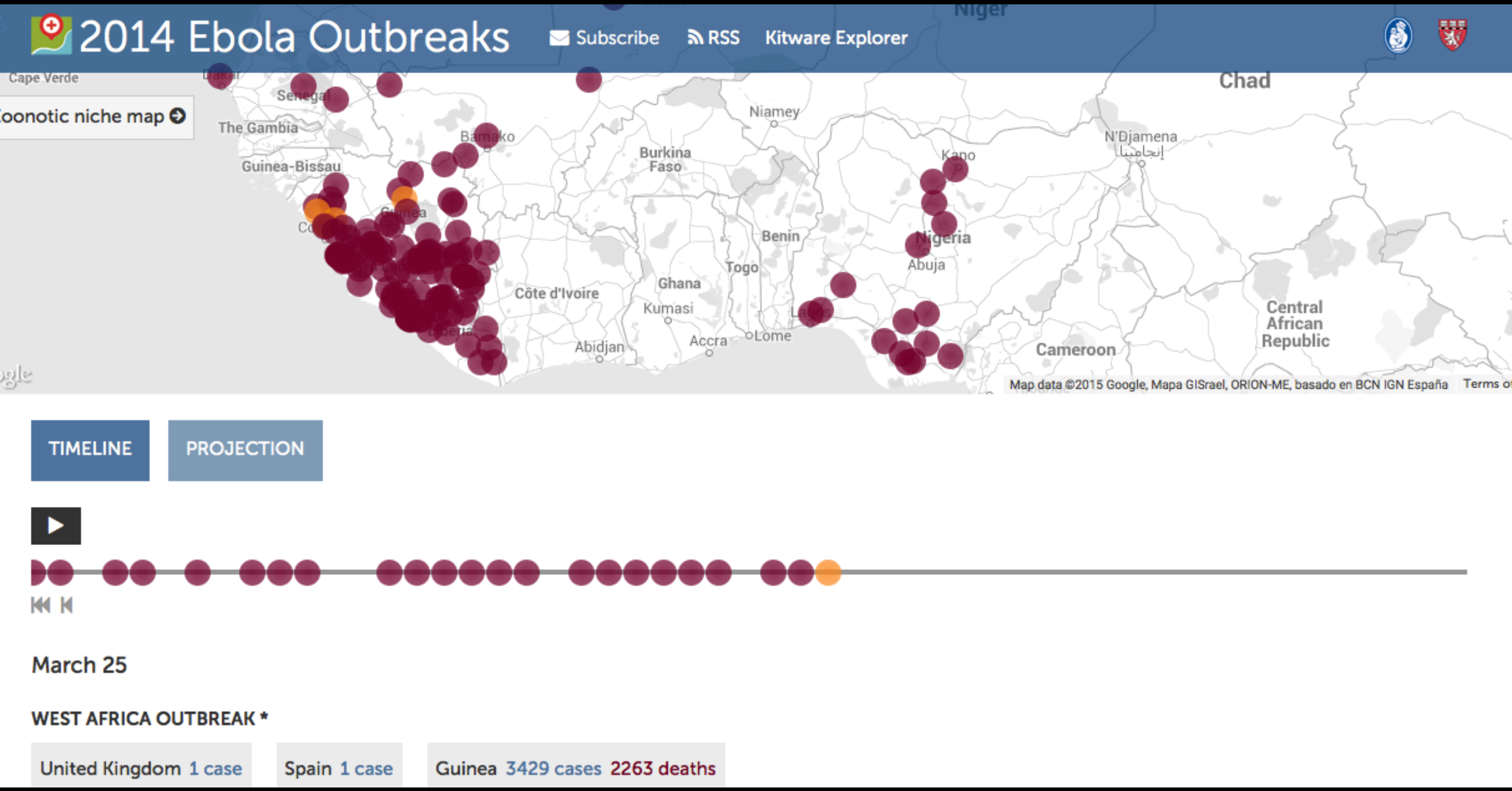
CIHR IRSC

BILL & MELINDA GATES foundation

amazon.com

MERCK

Recent success story: Ebola outbreak identification and tracking



http://www.healthmap.org/ebola/#timeline

PLOS | CURRENTS OUTBREAKS

2014 Ebola Outbreak: Media Events Track Changes in Observed Reproductive Number

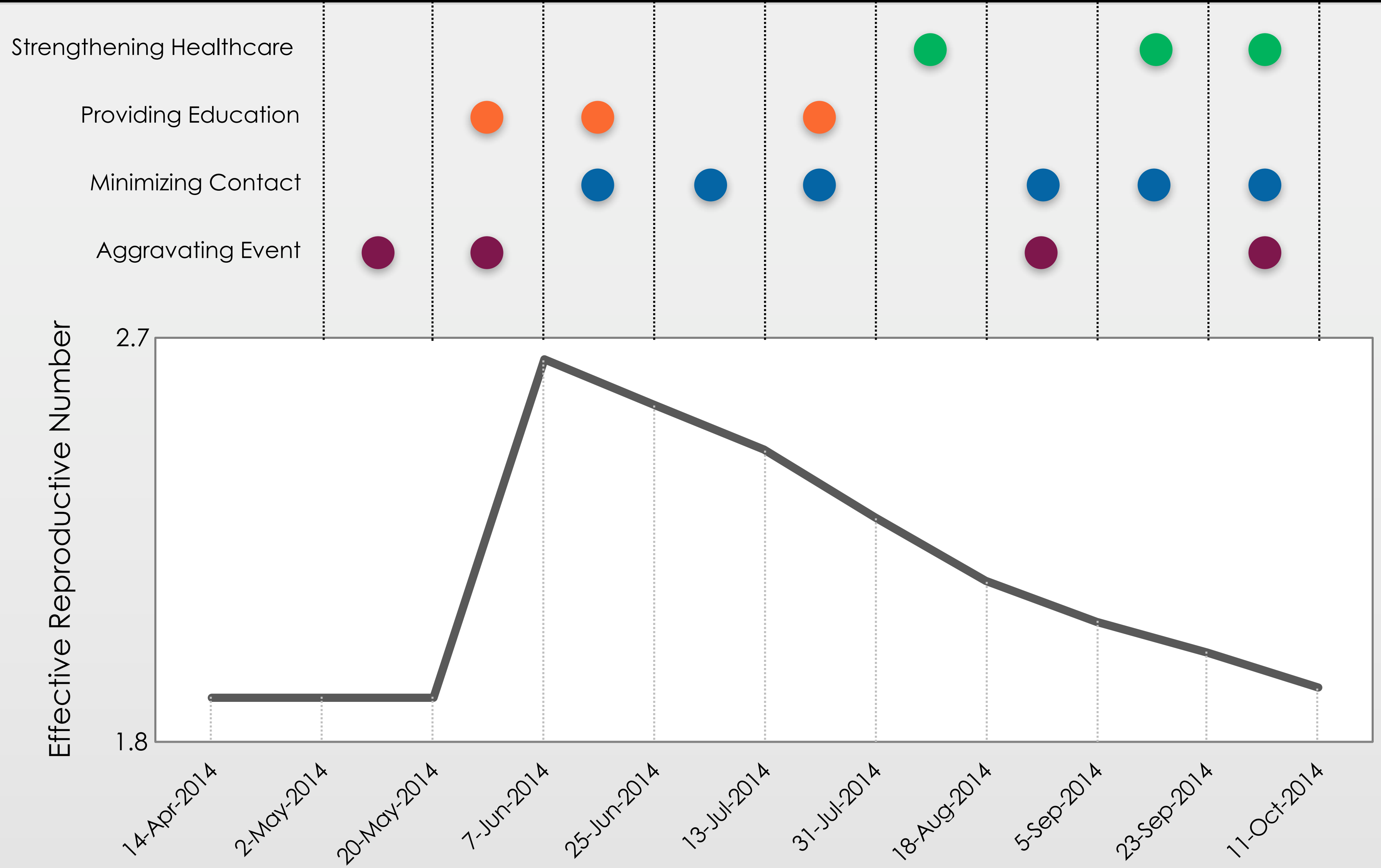
APRIL 28, 2015 · COMMENTARY

AUTHORS

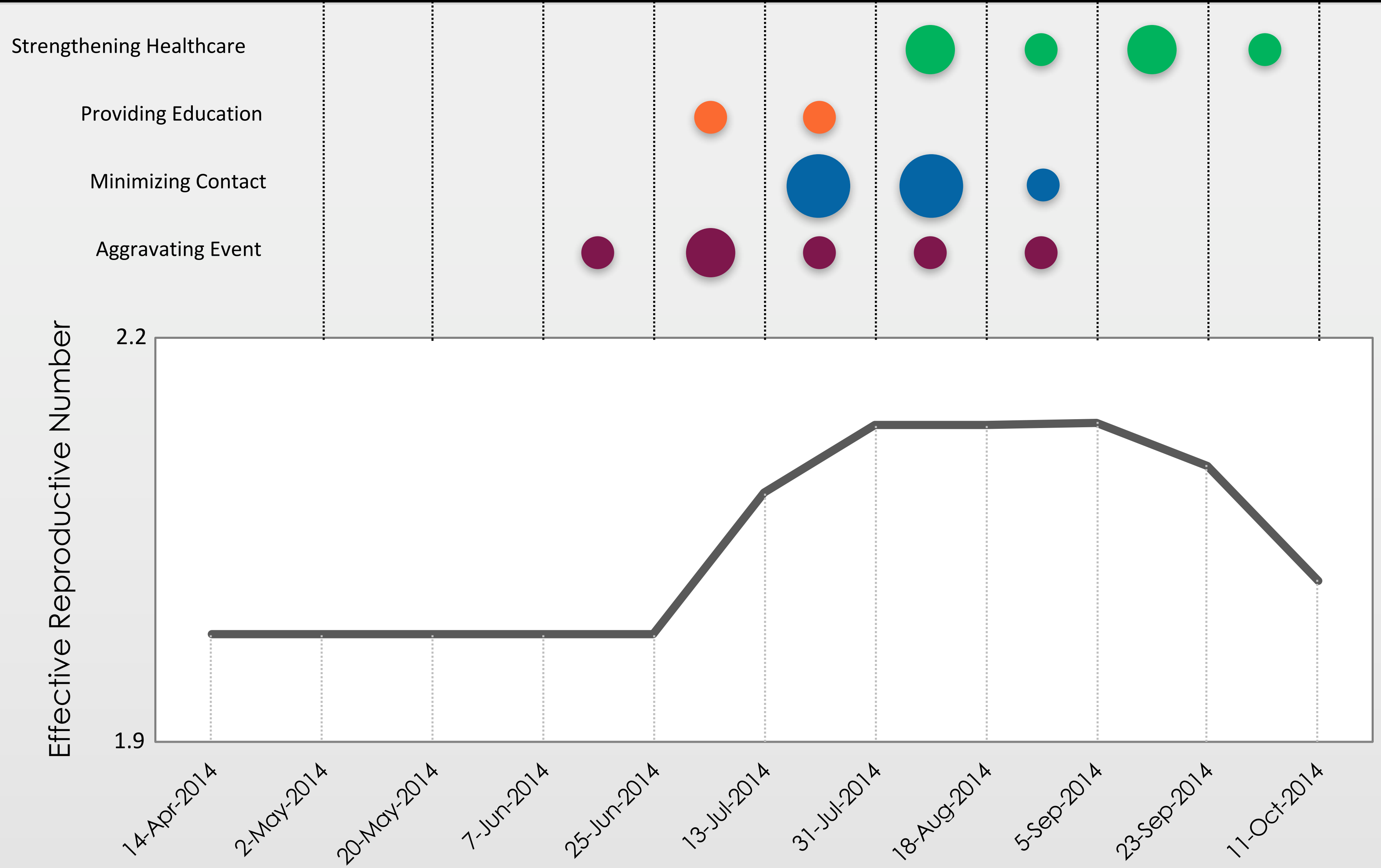
Maimuna S. Majumder Sheryl Kluberg Mauricio Santillana Sumiko Mekaru John S. Brownstein

ABSTRACT

In this commentary, we consider the relationship between early outbreak changes in the observed reproductive number of Ebola in West Africa and various media reported interventions and aggravating events. We find that media reports of interventions that provided education, minimized contact, or strengthened healthcare were typically followed by sustained transmission reductions in both Sierra Leone and Liberia. Meanwhile, media reports of aggravating events generally preceded temporary transmission increases in both countries. Given these preliminary findings, we conclude that media reported events could potentially be incorporated into future epidemic modeling efforts to improve mid-outbreak case projections.



Sierra Leone


Strengthening Healthcare
Providing Education
Minimizing Contact
Aggravating Event
Effective Reproductive Number
2.2
1.9
14-Apr-2014
2-May-2014
20-May-2014
7-Jun-2014
25-Jun-2014
13-Jul-2014
31-Jul-2014
18-Aug-2014
5-Sep-2014
23-Sep-2014
11-Oct-2014

Liberia

A more recent contribution on the 2015 Latin American **Zika** outbreak

A more recent contribution on the 2015 Latin American **Zika** outbreak

JMIR Publications

JMIR Public Health and Surveillance

Published on 01.06.16 in Vol 2, No 1 (2016): Jan-Jun

This paper is in the following e-collection/theme issue:

Infoveillance, Infodemiology and Digital Disease Surveillance Infodemiology and Infoveillance

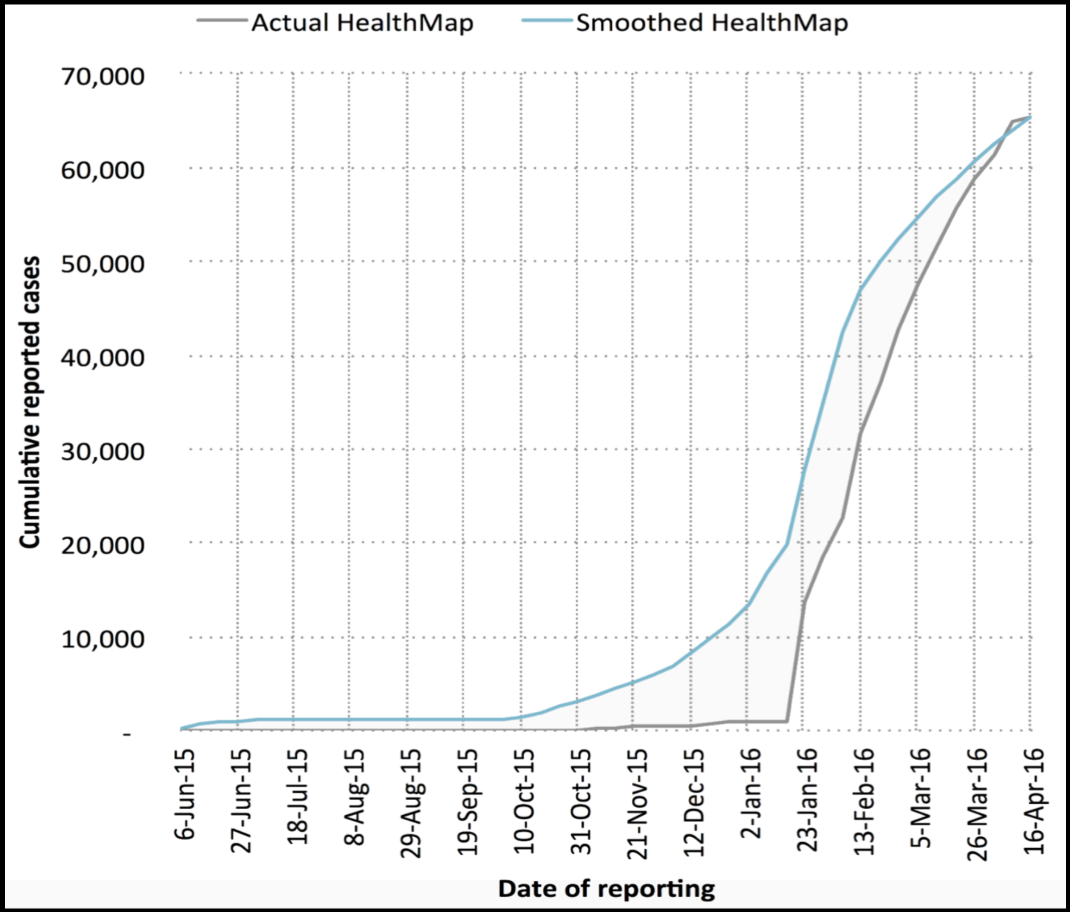
Article | Cited By (0) | Tweetations (29) | Metrics

Original Paper

Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak

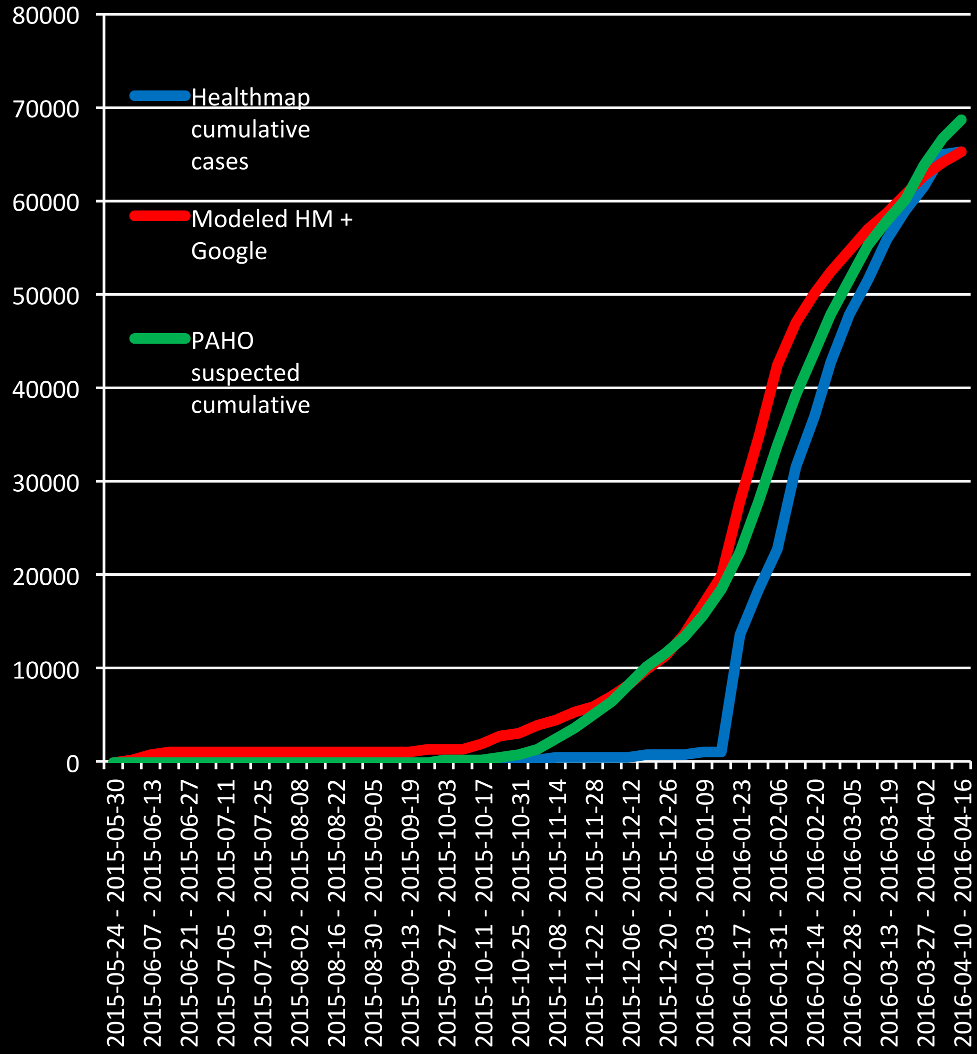
Maimuna S Majumder[1,2], MPH ; Mauricio Santillana[1,3,4], PhD ; Sumiko R Mekaru[1,5], PhD ; Denise P McGinnis[1], ScD ; Kamran Khan[6,7], MD ; John S Brownstein[1,4], PhD

A more recent contribution on the 2015 Latin American **Zika** outbreak

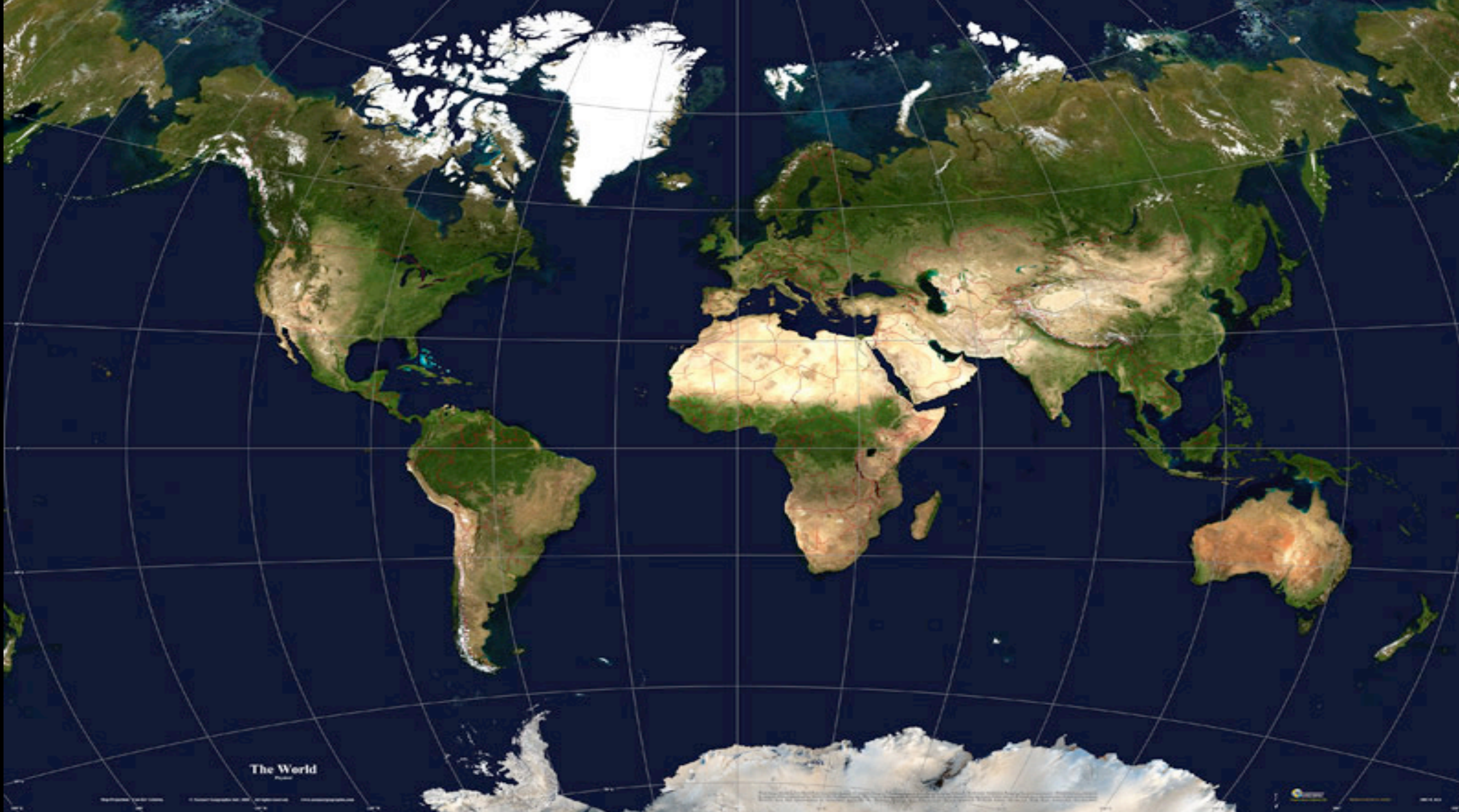


With no access to traditional, government-lead disease surveillance information, we extracted the number of suspected cases as reported by **new reports** as a function of time. We then utilized the time behavior of Google searches of the word “zika” to smooth the news-reported incidence data.

A more recent contribution on the 2015 Latin American **Zika** outbreak



When we gained access to government-lead disease surveillance information, we found great similarity with the curve we produced ahead of the publication of this information.



Thank you!

Contact: msantill@fas.harvard.edu