



## 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*

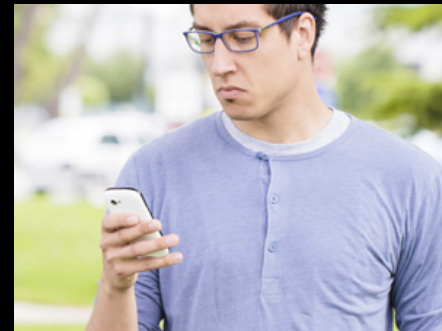
*Instructor, Harvard Medical School*



# Beyond Google searches...



What are doctors searching for?



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?

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

APRIL 28, 2015 · COMMENTARY

 [Print or Save PDF](#)

 [Citation](#)

 [XML](#)

 [Email](#)

 [Tweet](#)

 [Like](#)

 10

## ■ AUTHORS

[Maimuna S. Majumder](#) [Sheryl Kluberg](#) [Mauricio Santillana](#) [Sumiko Mekaru](#) [John S. Brownstein](#)

## ■ ABSTRACT

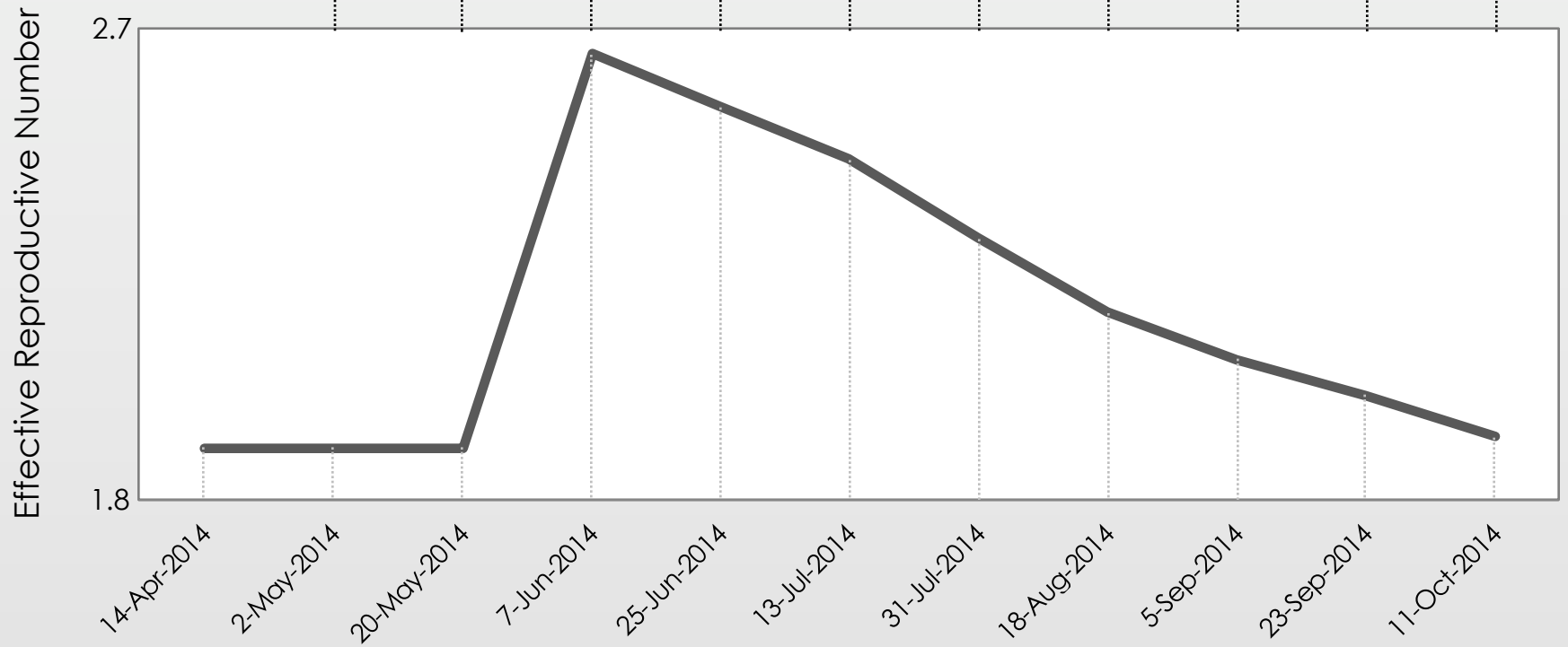
In this commentary, we consider the relationship between early outbreak changes in the observed reproductive number of Ebola in West Africa and various media reported interventions and aggravating events. We find that media reports of interventions that provided education, minimized contact, or strengthened 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.

Strengthening Healthcare

Providing Education

Minimizing Contact

Aggravating Event



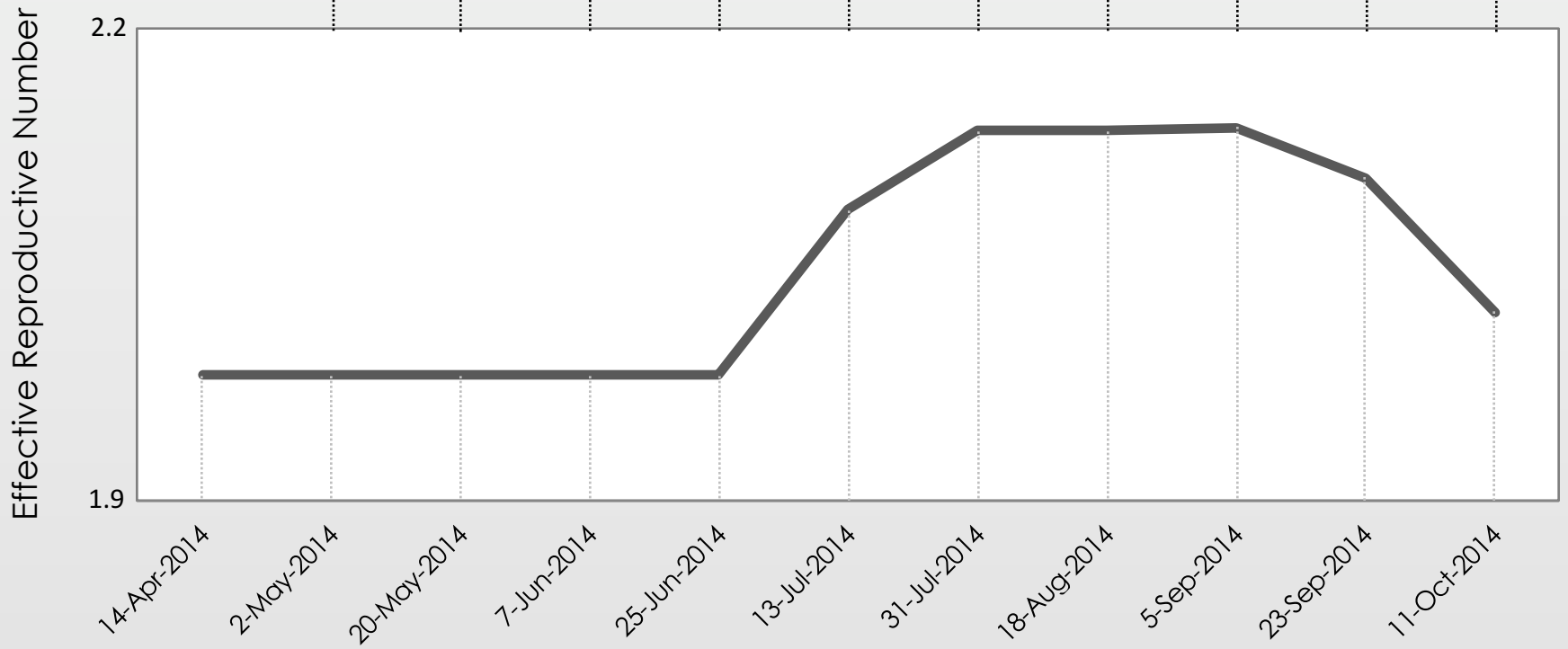
# Sierra Leone

Strengthening Healthcare

Providing Education

Minimizing Contact

Aggravating Event



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)

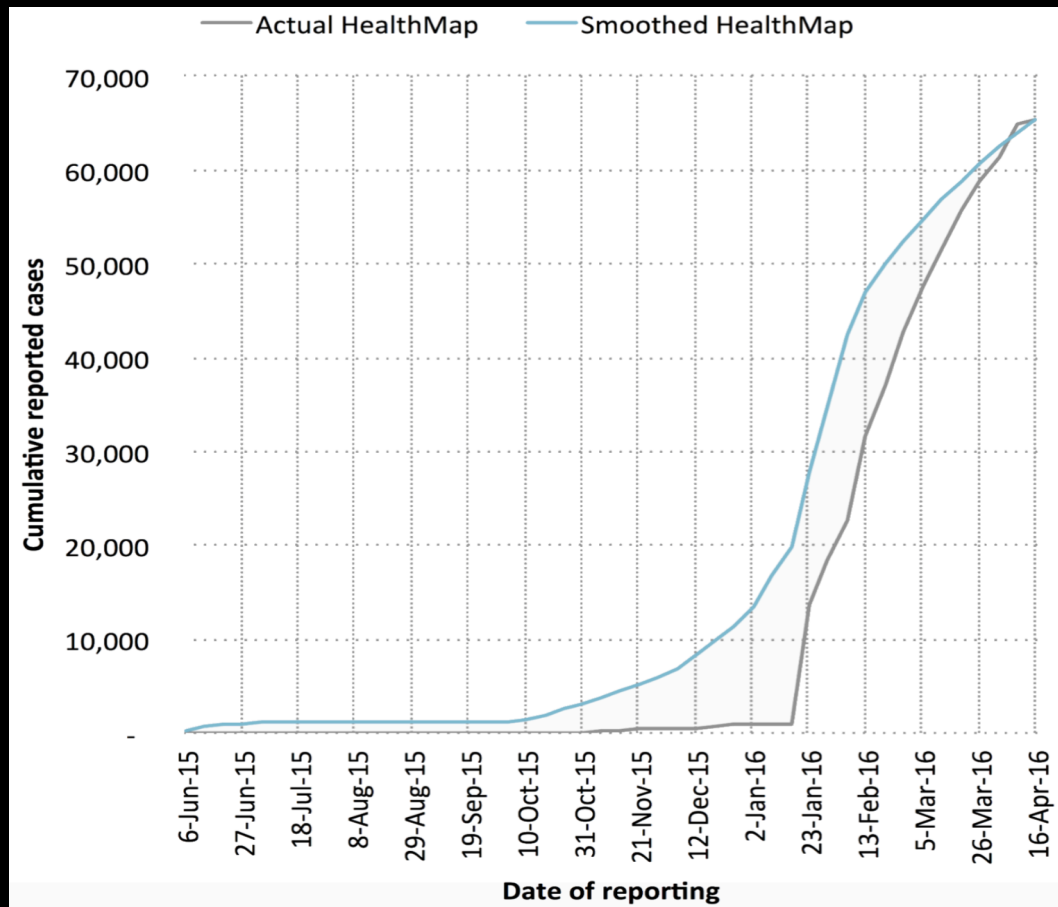
Metrics

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<sup>1,2</sup>, MPH ; Mauricio Santillana<sup>1,3,4</sup>, PhD ; Sumiko R Mearu<sup>1,5</sup>, PhD ; Denise P McGinnis<sup>1</sup>, ScD ; Kamran Khan<sup>6,7</sup>, MD ; John S Brownstein<sup>1,4</sup>, PhD

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

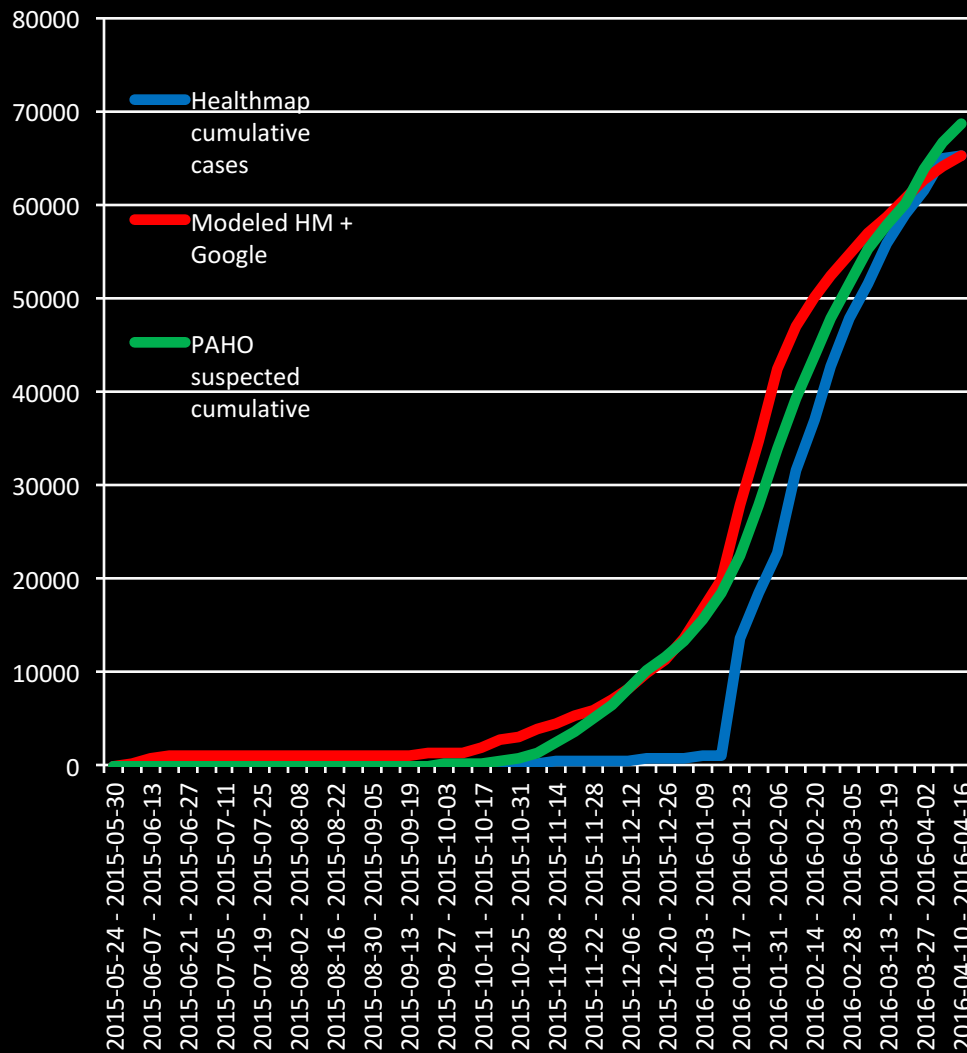


With no access to traditional, government-lead disease surveillance information, we extracted the number of suspected cases as reported by **new reports** as a function of time. We then utilized the time behavior of Google searches of the word “zika” to smooth the news-reported incidence data.



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

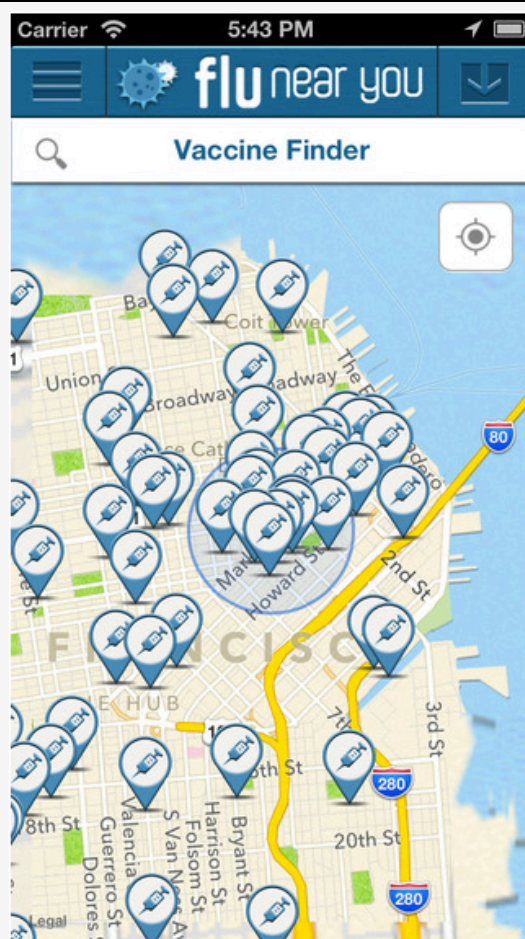
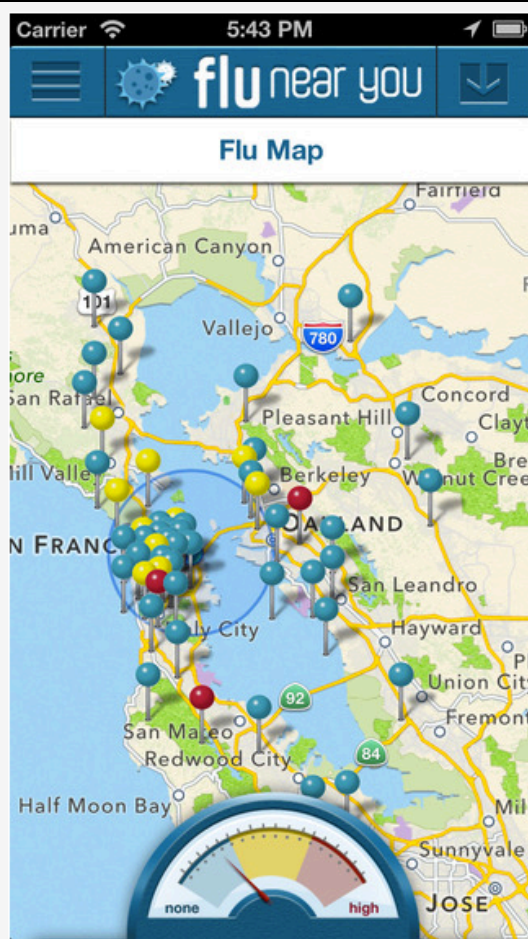
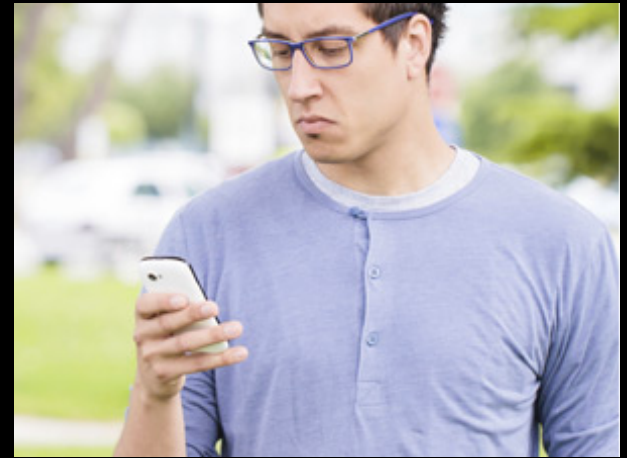
## Zika outbreak



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.

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

flu near you   
do you have it in you?



Launched in 2011

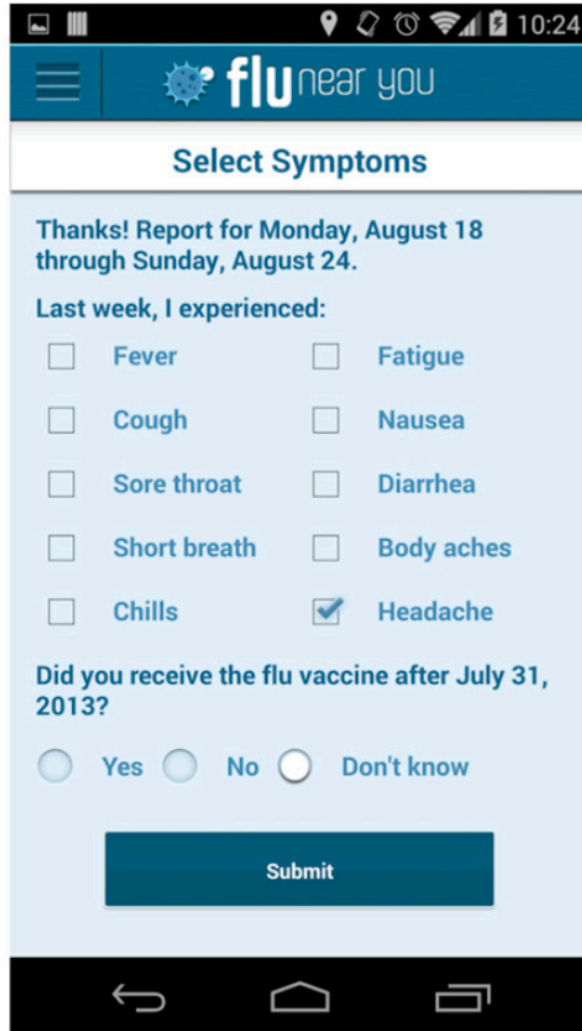
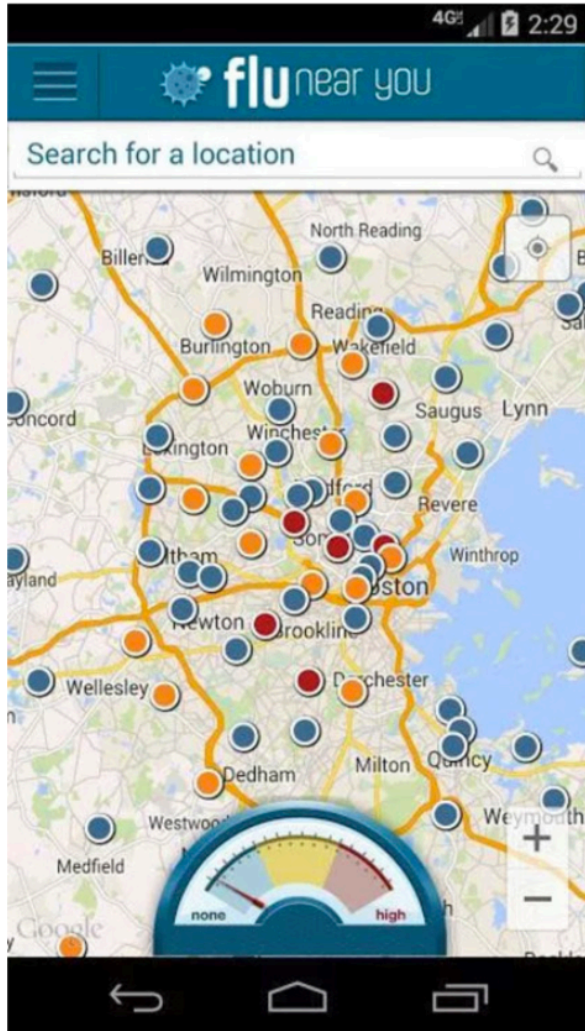





FIGURE 1—Flu Near You mobile interface.

Flu Activity in: **United States**

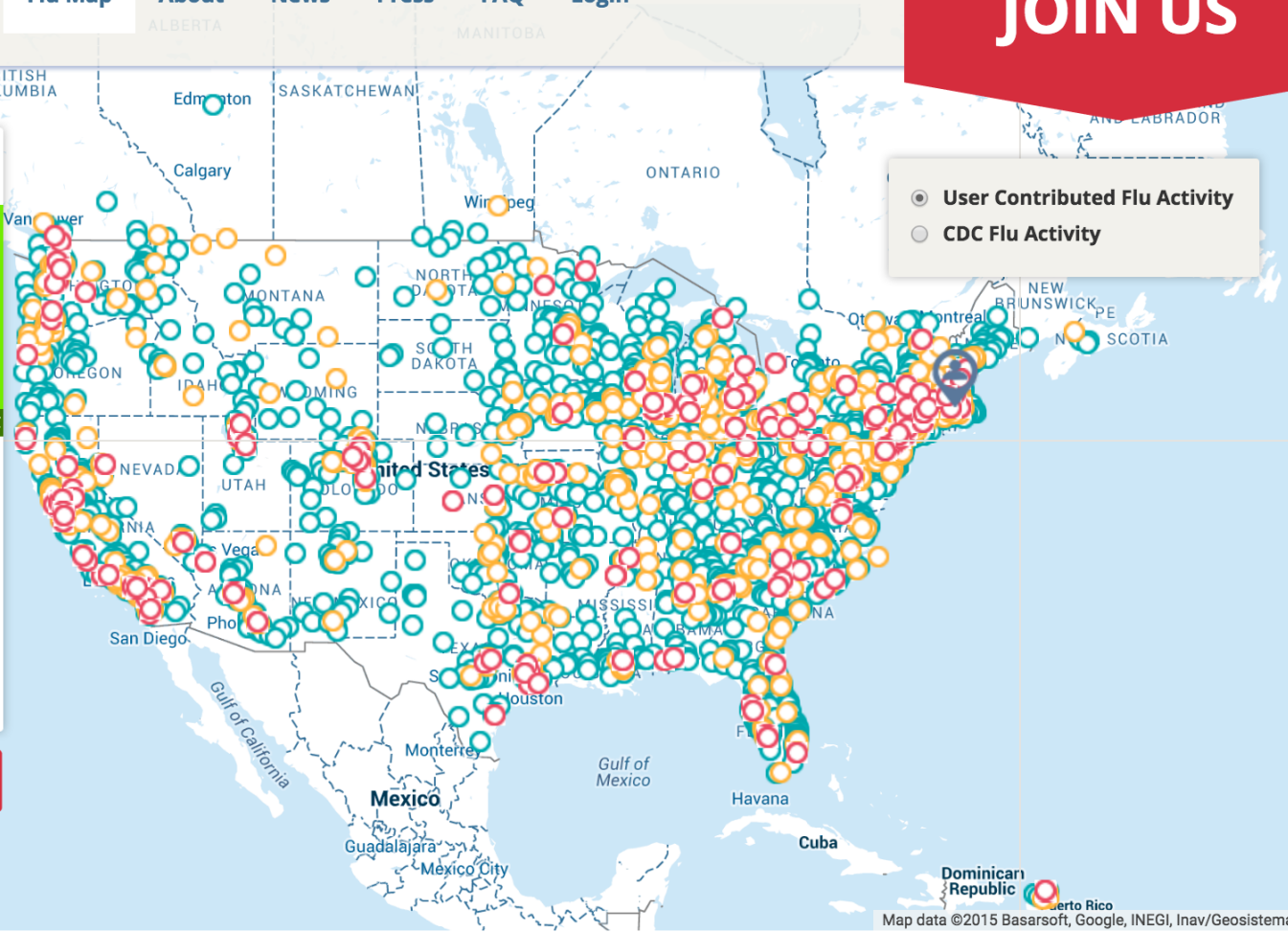


9914 reports this week

 <b>212</b>	Flu-Like Symptoms 2%	↓
 <b>1298</b>	Any symptoms 13%	
 <b>8616</b>	No symptoms 87%	

Enter your City or Zipcode

check for flu





**TABLE 1—Flu Near You User Characteristics and Reporting: United States, 2012–2014**

	2012–2013	2013–2014
Reports per user		
Mean	7.3	10.0
Median	4.0	5.0
IQR	8.0	14.0
Min	1.0	1.0
Max	33.0	33.0
Symptoms, %		
Reported ILI	20.0	16.3
Never reported ILI	80.0	83.7
Gender, no. (%)		
Female	33 698 (53.3)	20 750 (52.0)
Male	24 178 (38.3)	15 956 (40.0)
Missing	5 305 (8.4)	3 176 (8.0)
Total	63 181	39 882
Age group, y no. (%) <sup>a</sup>		
0–12	6 375 (10.1)	3 530 (8.9)
13–17	2 861 (4.5)	2 137 (5.4)
18–29	6 567 (10.4)	3 886 (9.7)
30–39	10 182 (16.1)	5 324 (13.3)
40–49	11 823 (18.7)	6 725 (16.9)
50–59	12 193 (19.3)	8 209 (20.6)
60–69	9 693 (15.3)	7 337 (18.4)
≥ 70	3 487 (5.5)	2 734 (6.9)
Total, no.	63 181	39 882

Note. ILI = influenza-like illness; IQR = interquartile range. Percentages may not add to 100 because of rounding.

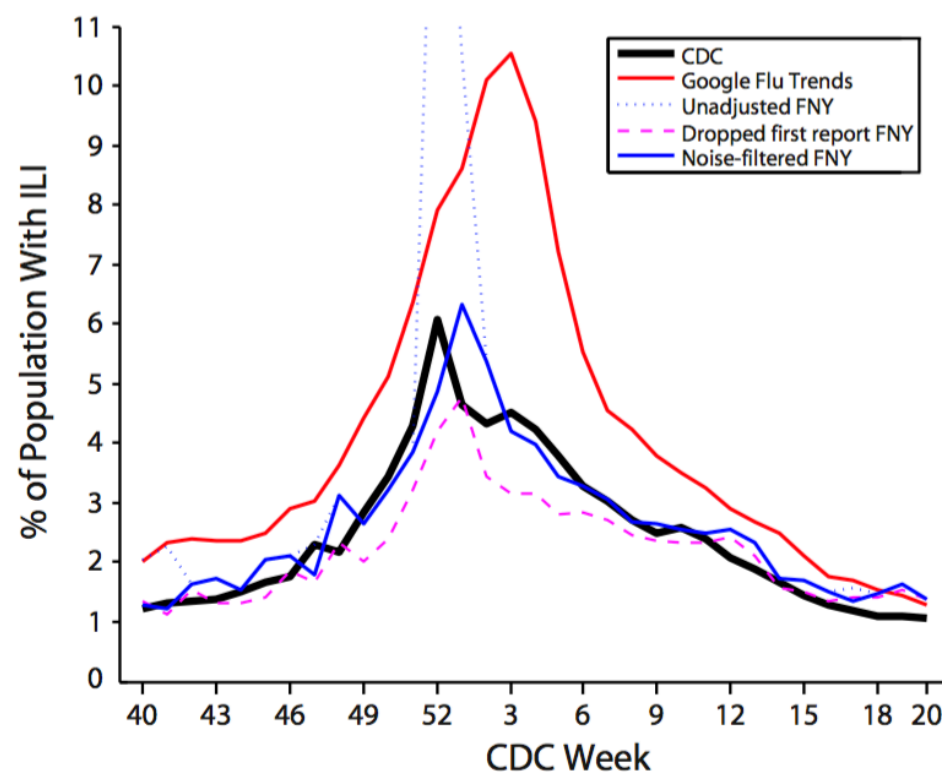
**TABLE 2—Flu Near You Participants by Department of Health and Human Services Surveillance Region: United States, 2012–2014**

HHS Region <sup>a</sup>	FNY 2012–2013, No. (%)	FNY 2013–2014, No. (%)	United States 2013, No. (%)
1	4 450 (7.0)	3 542 (8.9)	14 618 806 (4.6)
2	4 807 (7.6)	2 587 (6.5)	32 165 552 (10.1)
3	7 542 (11.9)	4 402 (11.0)	30 389 522 (9.5)
4	8 572 (13.6)	5 537 (13.9)	62 884 128 (19.7)
5	10 024 (15.9)	6 138 (15.4)	52 082 560 (16.3)
6	5 624 (8.9)	3 292 (8.3)	39 968 891 (12.5)
7	2 689 (4.3)	1 698 (4.3)	13 897 060 (4.3)
8	3 705 (5.9)	2 111 (5.3)	11 335 332 (3.5)
9	10 885 (17.2)	7 459 (18.7)	49 153 335 (15.4)
10	4 822 (7.6)	3 084 (7.7)	13 248 739 (4.1)
Missing	61 (0.1)	32 (0.1)	
Total	63 181 (100.0)	39 882 (100.0)	319 743 925 (100.0)

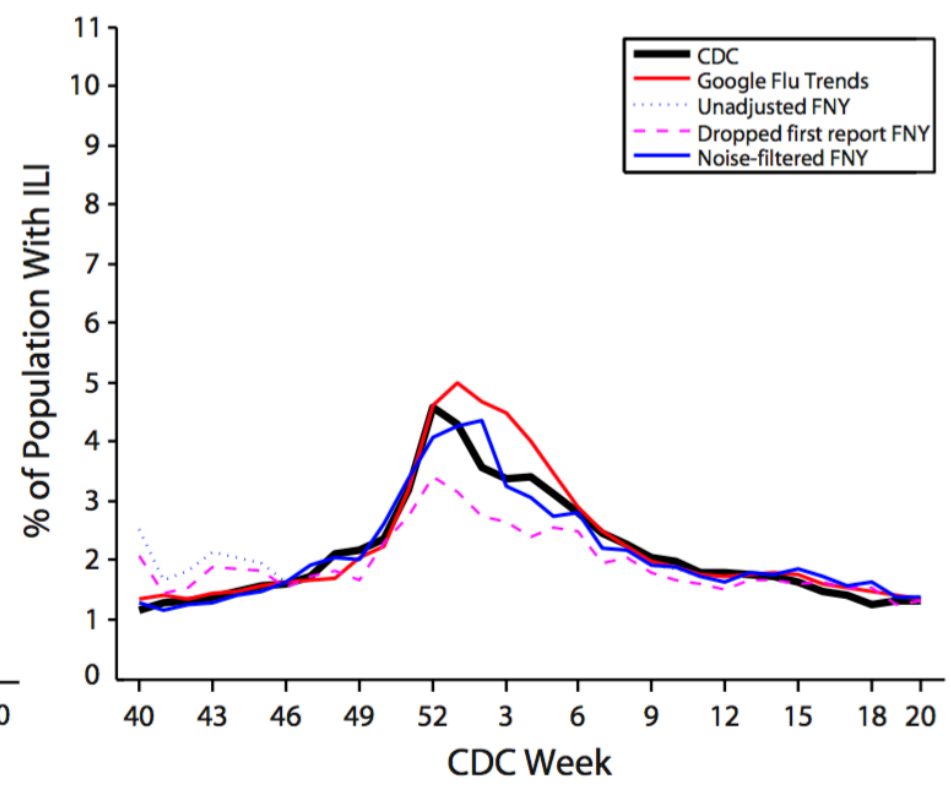
Note. FNY = Flu Near You; HHS = Department of Health and Human Services.

<sup>a</sup>States by HHS Surveillance Region—Region 1: CT, ME, MA, NH, RI, VT; Region 2: NJ, NY, PR; Region 3: DE, DC, MD, PA, VA, WV; Region 4: AL, FL, GA, KY, MS, NC, SC, TN; Region 5: IL, IN, MI, MN, OH, WI; Region 6: AR, LA, NM, OK, TX; Region 7: IA, KS, MO, NE; Region 8: CO, MT, ND, SD, UT, WY; Region 9: AZ, CA, HI, NV; Region 10: AK, ID, OR, WA.

a



b



System Comparison	2012-2013 Pearson	2012-2013 RMSE	2012-2013 Peak ILI Week	2013-2014 Pearson	2013-2014 RMSE	2013-2014 Peak ILI Week
CDC	1.0000	0.0000 % ILI	Dec 24, 2012 (week 52)	1.0000	0.0000 % ILI	Dec 23, 2013 (week 52)
CDC vs FNY raw	0.7847	2.2765 % ILI	Dec 24, 2012 (week 52)	0.9119	0.3912 % ILI	Jan 6, 2014 (week 2)
CDC vs FNY drop-adjusted	0.9353	0.6248 % ILI	Dec 31, 2012 (week 1)	0.9265	0.4936 % ILI	Dec 23, 2013 (week 52)
CDC vs FNY noise-filtered	0.9226	0.5165 % ILI	Dec 31, 2012 (week 1)	0.9658	0.2332 % ILI	Dec 30, 2013 (week 1)

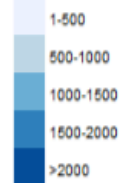
Note. CDC = Centers for Disease Control and Prevention; FNY = Flu Near You; ILI = influenza-like illness; RMSE = root-mean-square error. For context, ILI rates obtained from Google Flu Trends and the raw FNY are shown.

**FIGURE 2—Comparison between report-adjusted FNY, noise-filtered FNY, and CDC influenza-like illness rates for flu seasons (a) 2012–2013 and (b) 2013–2014: United States.**

Unadjusted

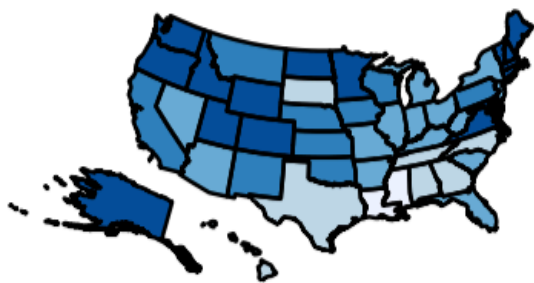


Number of Participants

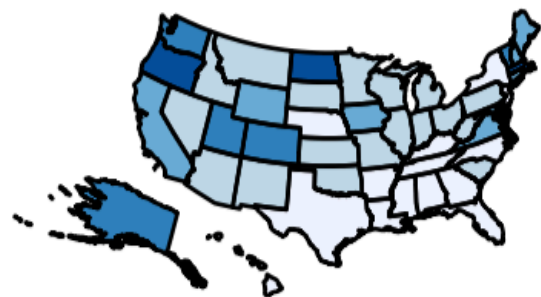


Adjusted

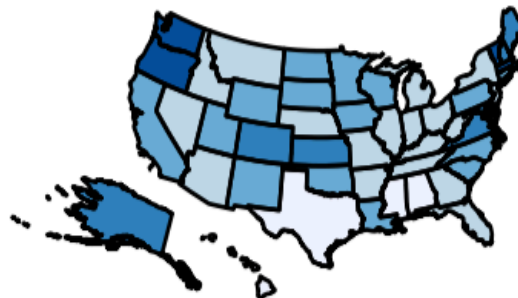
n= 41,523



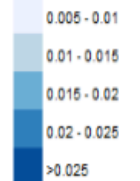
n= 31,869



n= 49,814



Percent of Population

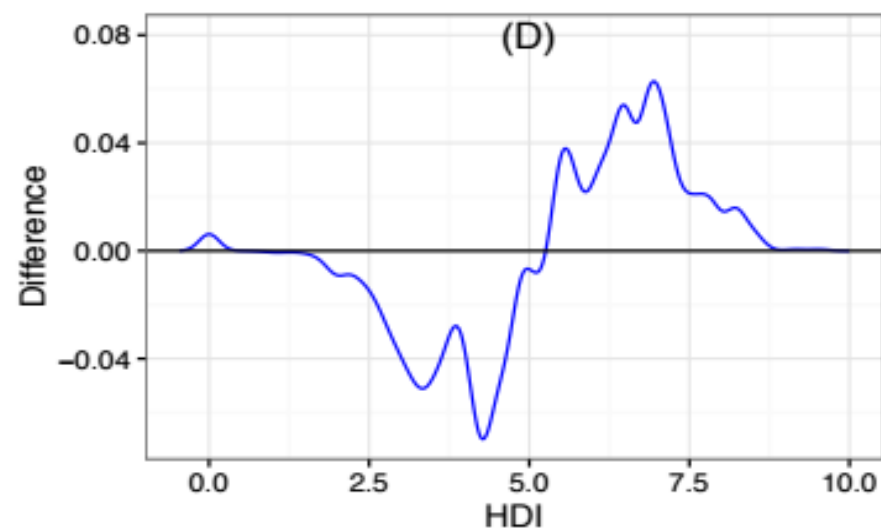
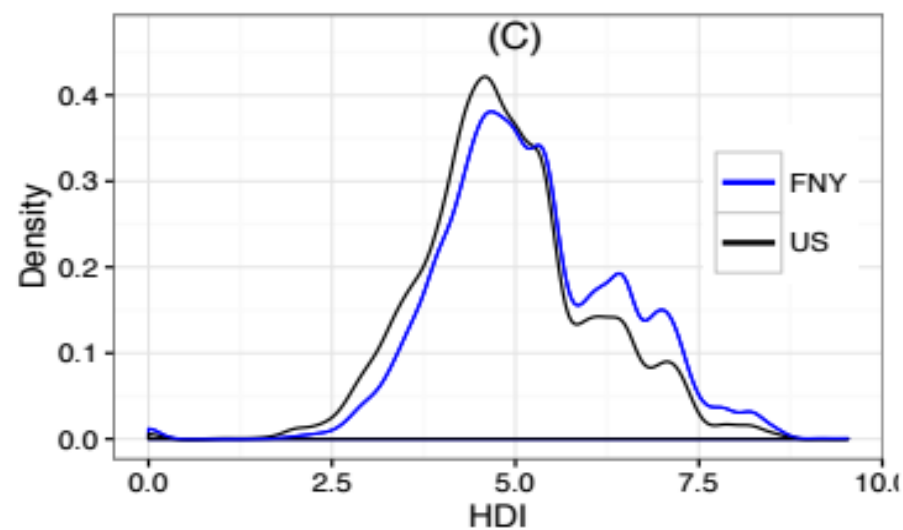
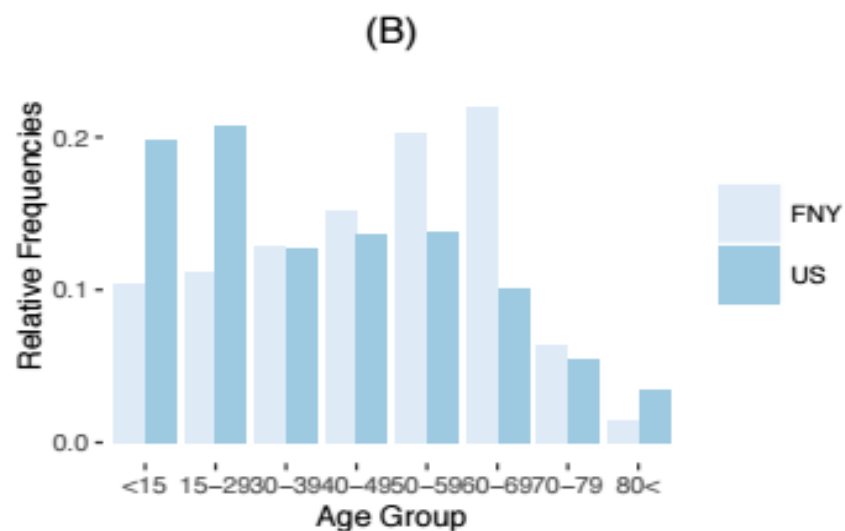
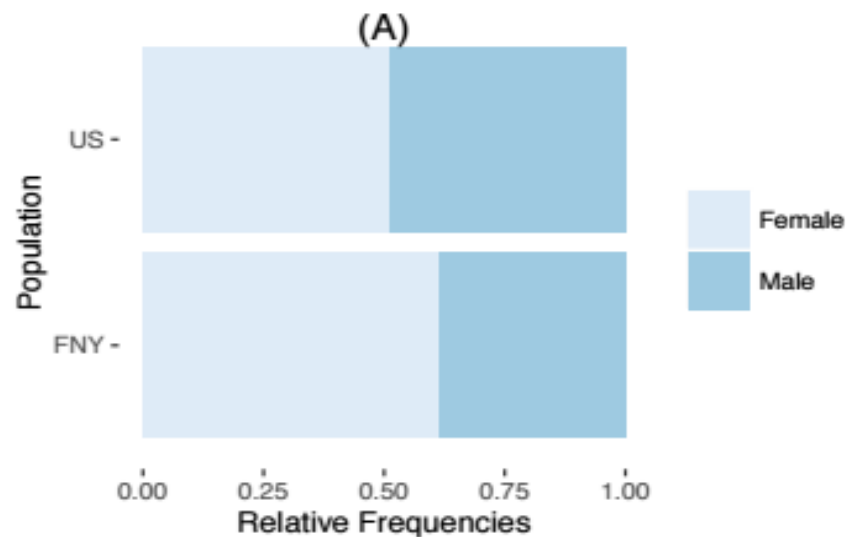


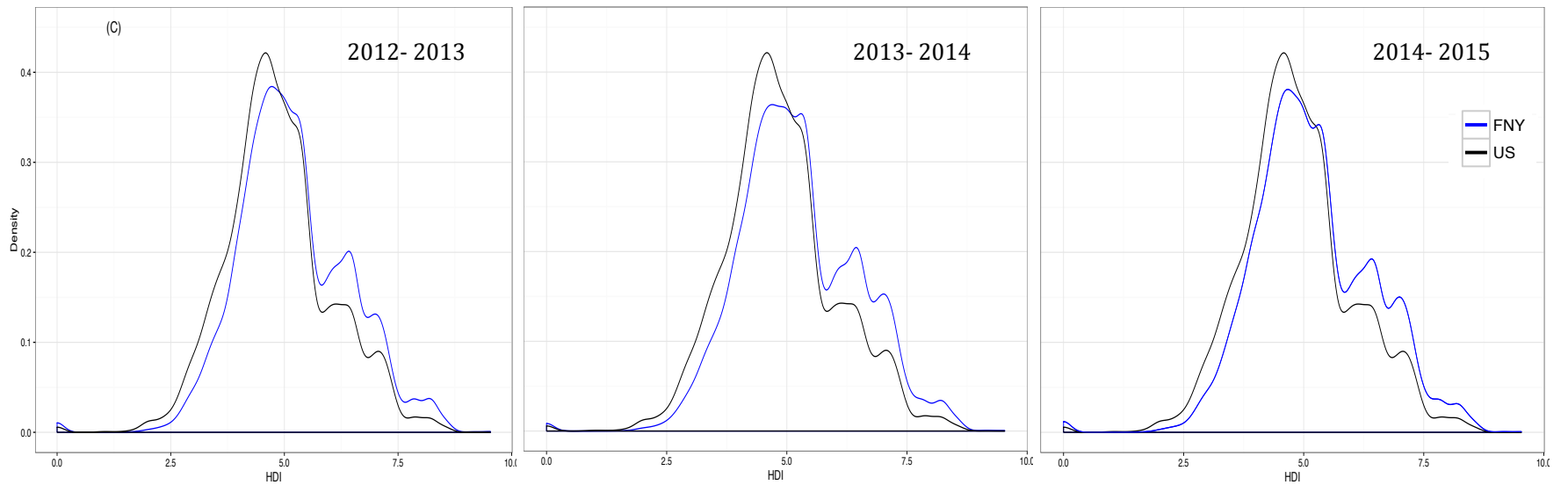
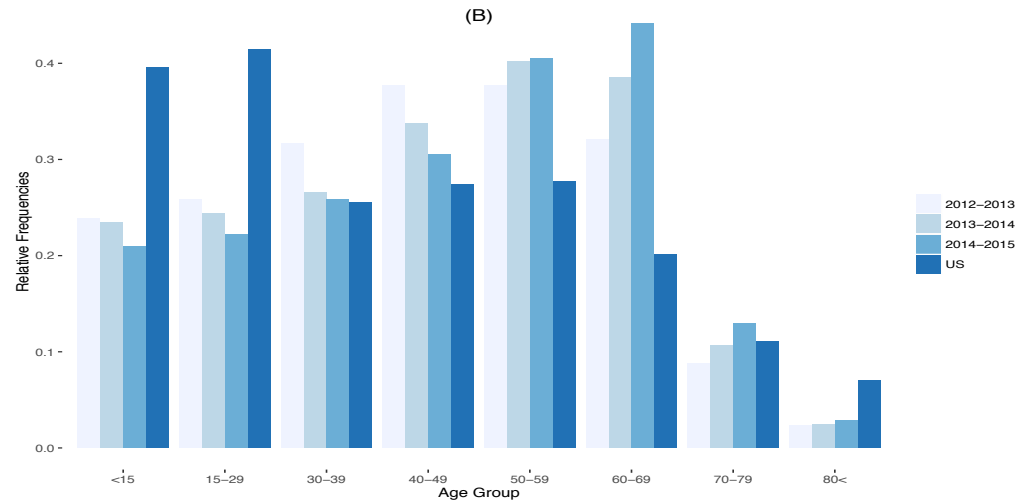
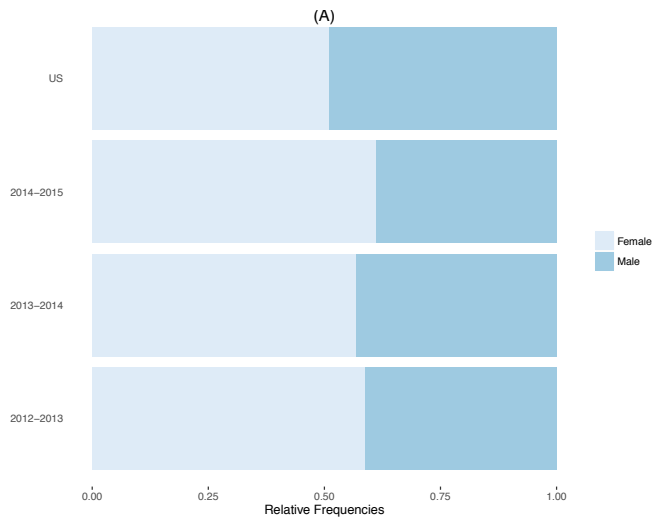
2012- 2013

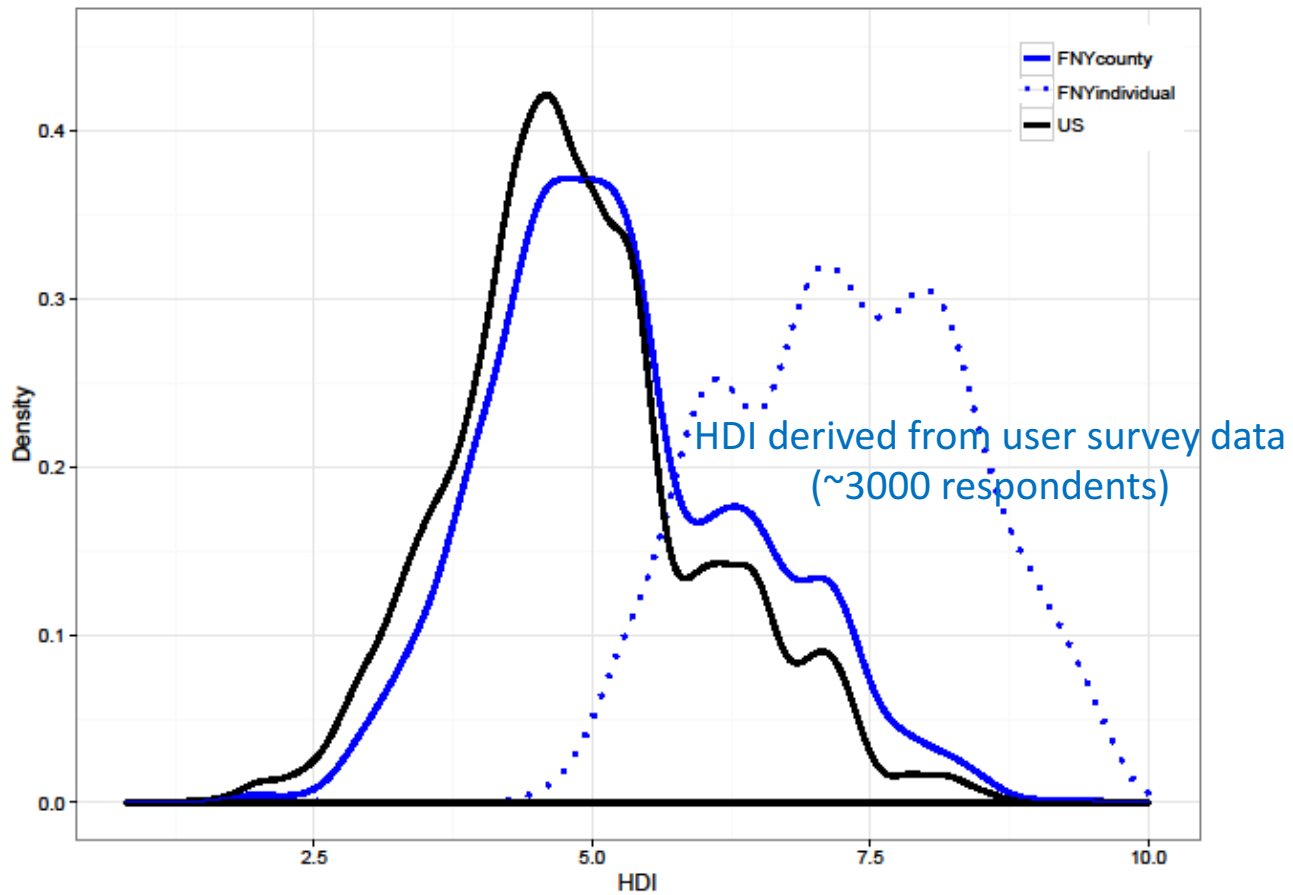
2013- 2014

2014- 2015

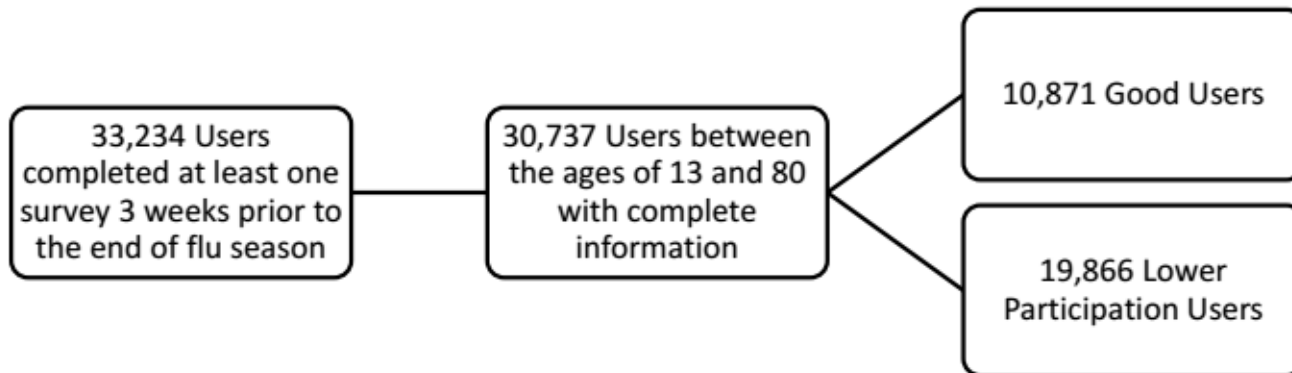








### Consistent users vs sporadic users (2014-2015 flu season)



## Likelihood of being a consistent user

**Table 1:** Summary of Adjusted ORs

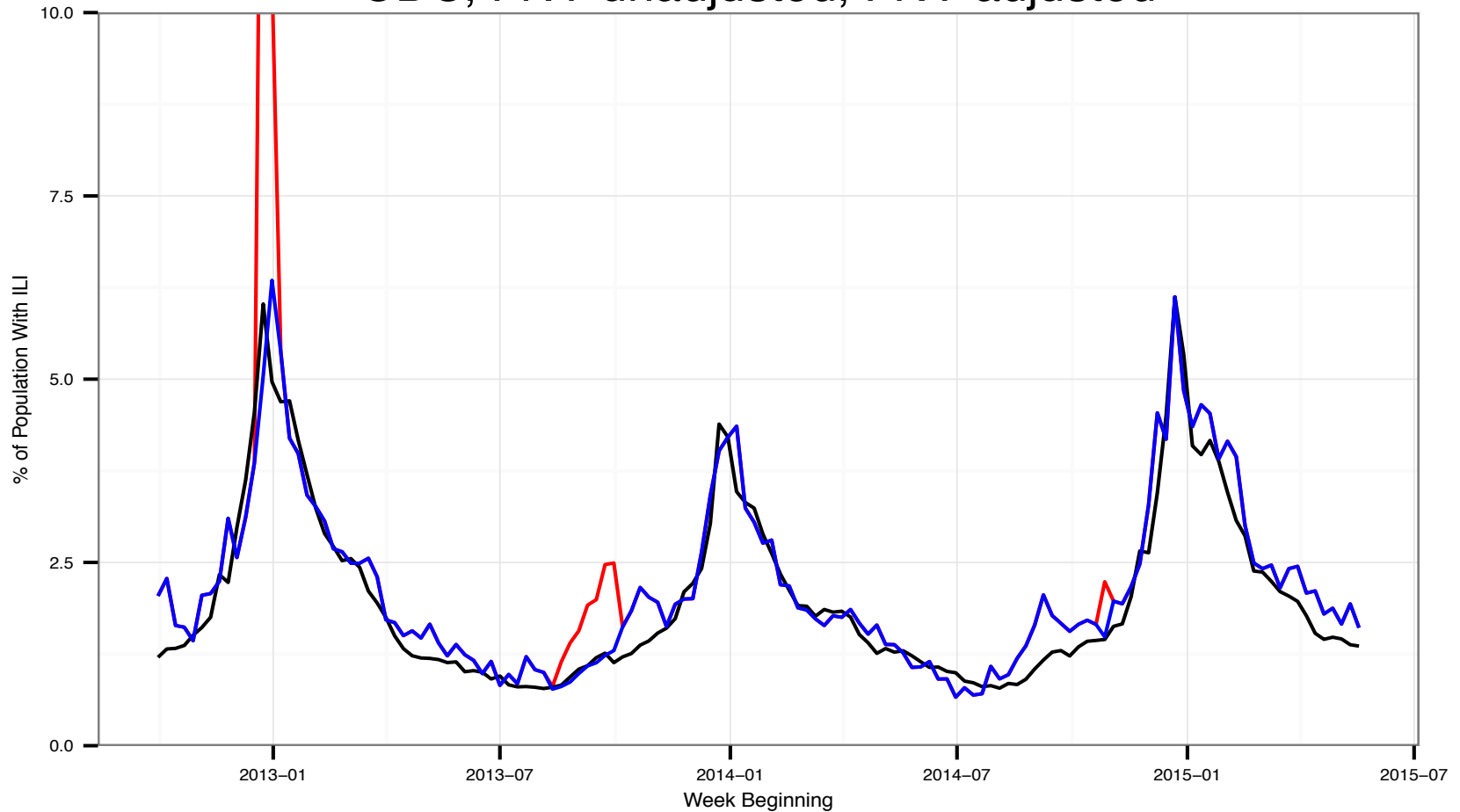
Variable	Reference group	Adjusted OR	95% CI	P-value
<b>Sex</b>				
Female	Male	0.75	(0.71, 0.79)	<0.001
<b>Household Members</b>				
No	Yes	3.29	(3.12, 3.35)	<0.001
<b>HDI</b>	-	1.12	(1.09, 1.14)	<0.001
<b>ILI Status at 1<sup>st</sup> Survey</b>				
Yes	No	0.22	(0.19, 0.25)	<0.001
<b>Age Group</b>				
13- 29	50- 59	0.67	(0.61, 0.74)	<0.001
30- 39	50- 59	0.54	(0.49, 0.58)	<0.001
40- 49	50- 59	0.70	(0.64, 0.75)	<0.001
60- 69	50- 59	1.14	(1.07, 1.23)	<0.001
70- 79	50- 59	1.23	(1.11, 1.36)	<0.001



Overall, females were 25% less likely to be good users compared to men ( $p < 0.001$ ). Users who reported for additional household members had 3.29 times the odds of being good users compared to users who did not report for additional household members ( $p < 0.001$ ), while users who reported symptoms meeting the definition of ILI at the first entry were 78% less likely to be good users compared to users who did not ( $p < 0.001$ ). Each unit increase in HDI is associated with increase in odds of being a good user ( $OR = 1.12$ ,  $p < 0.001$ ). There is an increasing trend in participation with older age ( $p < 0.001$ ).

# National comparison with CDC

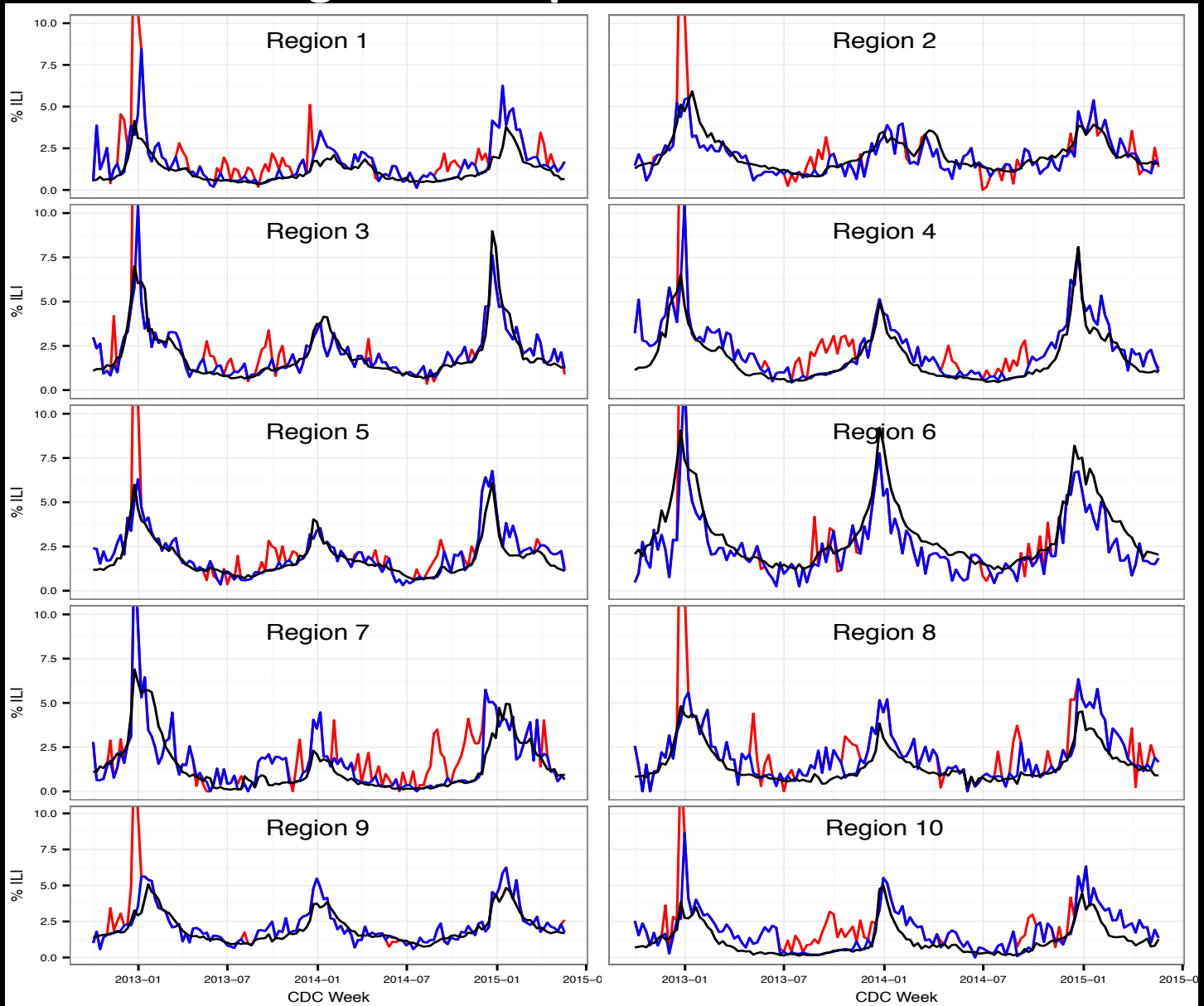
## CDC, FNY unadjusted, FNY adjusted



Plots produced by Kristin Baltrusaitis

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

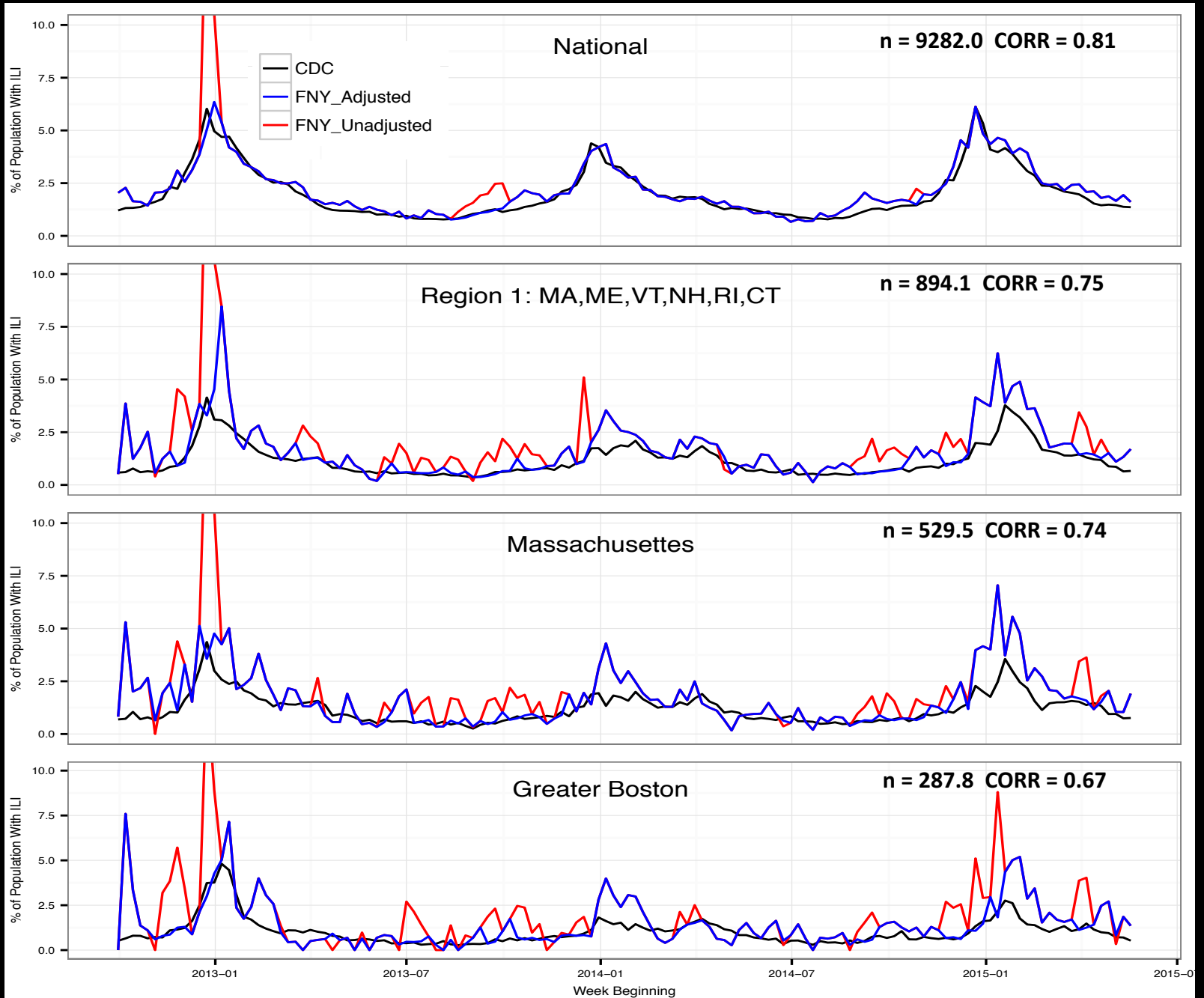
# Regional comparisons with CDC



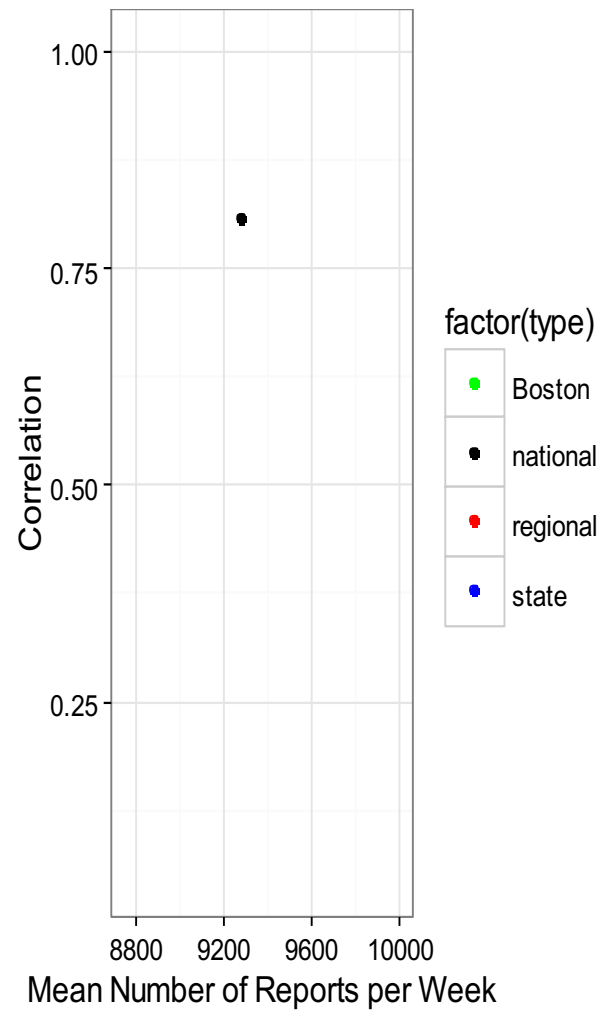
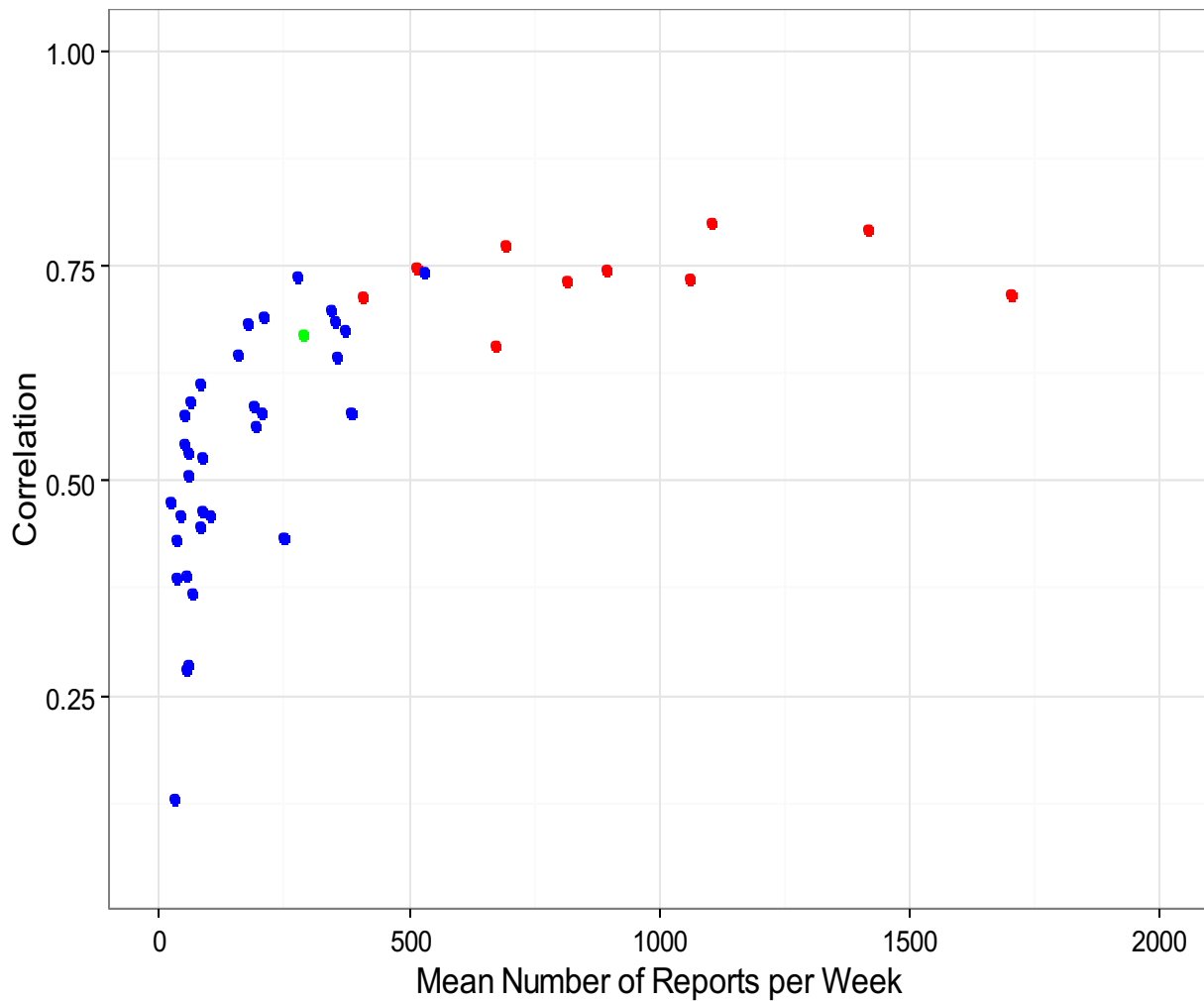
Time Series	Correlation Range	RMSE Range
FNY – raw	0.66 – 0.80	1.33 – 1.77
FNY – CDC adjusted	0.80 – 0.90	0.60 – 1.26

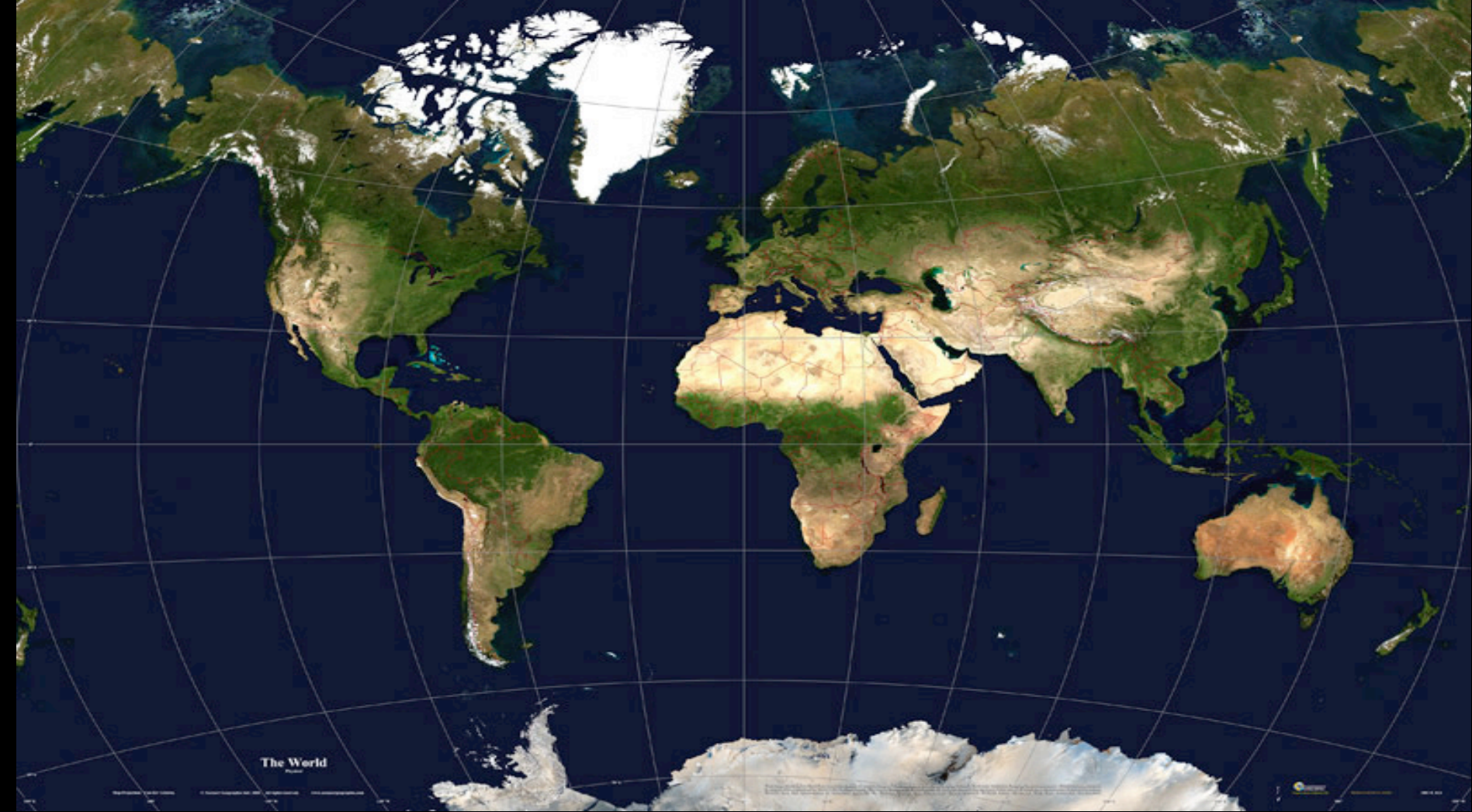


# Correlation of FNY with CDC. Multiple Geographic Scales



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





Thank you!

Contact: [msantill@fas.harvard.edu](mailto:msantill@fas.harvard.edu)