CHALLENGES IN INFLUENZA FORECASTING
AND OPPORTUNITIES FOR SOCIAL MEDIA

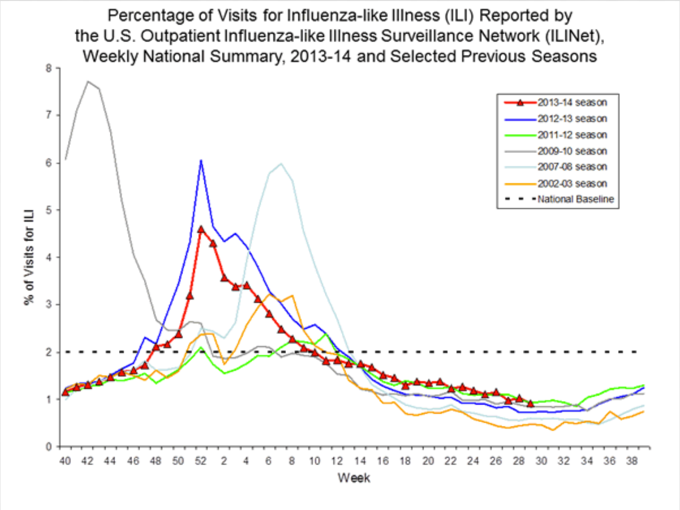
MICHAEL J. PAUL, MARK DREDZE, DAVID BRONIATOWSKI

NOVEL DATA STREAMS FOR INFLUENZA SURVEILLANCE

- New technology allows us to analyze new types of data to infer influenza prevalence
  - Especially data from the Web

- Most (in)famously Google Flu Trends

- Many other promising sources:
  - Social media (especially Twitter)
  - Mobile apps
  - Wikipedia



WHAT ABOUT *FORECASTING?*

- Detection is easy and utility is limited
  - Reliable **forecasting** is important for making preparations and allocating resources

- Google Flu has been shown to improve forecasting
  - Shaman and Karspeck (2012)
  - Nsoesie, Marathe, Brownstein (2013)
  - Dugas et al. (2013)

- Social media hasn't been evaluated yet

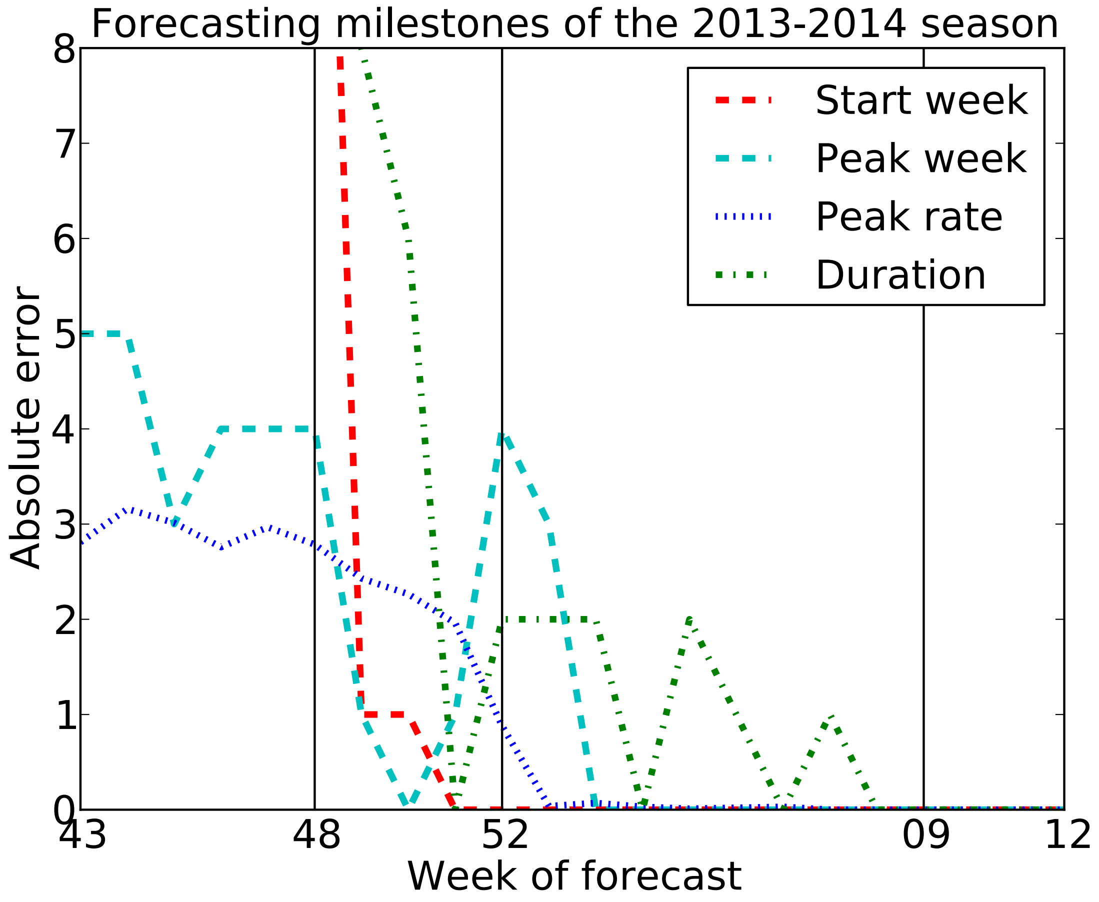
CDC *PREDICT THE FLU* CHALLENGE

CDC Competition Encourages Use of Social Media to Predict Flu

November 25, 2013 — CDC has launched the "Predict the Influenza Season Challenge," a competition designed to foster innovation in flu activity modeling and prediction. The registrant who most successfully predicts the timing, peak and intensity of the 2013-2014 flu season using social media data (e.g., Twitter, internet search data, web surveys) will receive an award of $75,000 and CDC recognition. Full details of the contest requirements – including eligibility rules, how to enter the contest, and scoring – are available via the official contest announcement at https://federalregister.gov/a/2013-28198.

- Contest to forecast the 2013-14 flu season by augmenting existing surveillance with Web data
- Three metrics:
  - Start of season
  - Peak of season
  - Intensity of season (peak rate and duration)

CDC *PREDICT THE FLU* CHALLENGE



INFLUENZA DATA

So what exactly are we trying to predict?

- **ILINet**
  - CDC-run network of thousands of US providers
  - Hospitals report % of outpatients seen for influenza-like illness
  - Weekly reports of estimated ILI prevalence
  - Most commonly used flu metric
- Data is lagged by a week
  - Real time surveillance doesn't exist through traditional means
  - This is why novel data streams can help

FORECASTING MODEL

- Forecasts and current-week nowcasts can be produced using standard time series models with the lagged ILINet data

- Basic autoregressive model:

$$\boxed{y_{w+k} = \alpha_1 \tilde{y}_{w-1} + \alpha_2 \tilde{y}_{w-2} + \alpha_3 \tilde{y}_{w-3}}$$

- This works quite well
  - Especially for nowcasting

FORECASTING MODEL

Can we improve this with social media data?

- Twitter can give estimates for the current week
  - These estimates can be included in the model

$$y_{w+k} = \gamma z_w + \alpha_1 \tilde{y}_{w-1} + \alpha_2 \tilde{y}_{w-2} + \alpha_3 \tilde{y}_{w-3}$$

TWITTER FLU DETECTION

- We used our state-of-the-art Twitter system
  - Lamb et al (2013) and Broniatowski et al (2013)
- Two streams downloading data since Nov 2011
  - 1% sample and stream filtered for health keywords
  - About 4 million per day
- Cascade of tweet classifiers:
  - Relevant to health
  - Relevant to flu
  - Indicates flu infection (vs general awareness)
- Can produce daily or weekly prevalence estimates

$$\frac{\text{\# of tweets classified as flu infection}}{\text{\# of tweets from full sample}}$$

RESULTS

- Mean absolute error when *nowcasting*:

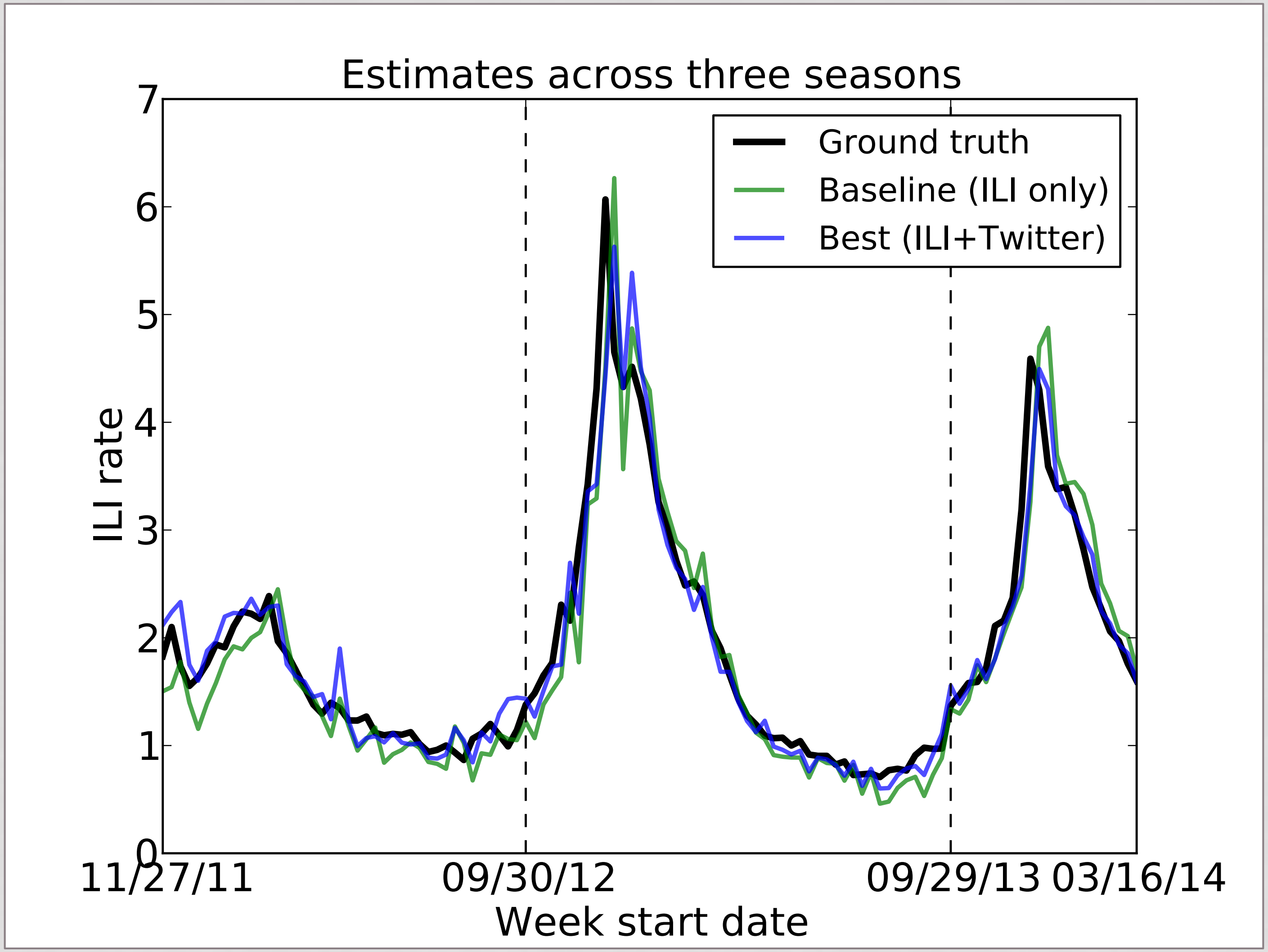
Data	**Average error per season**		
	2011-12	**2012-13**	**2013-14**
ILINet	.19	.30	.35
Twitter	.34	.36	.49
ILINet+Twitter	.15	.21	.21

RESULTS

- Mean absolute error when *forecasting*:

k	ILI Only	ILI+Twitter
0	$.28 \pm .07$	$.19 \pm .03$
1	$.41 \pm .14$	$.29 \pm .06$
2	$.51 \pm .20$	$.37 \pm .09$
3	$.62 \pm .26$	$.47 \pm .12$
4	$.75 \pm .32$	$.56 \pm .15$
5	$.88 \pm .39$	$.65 \pm .19$
6	$.98 \pm .45$	$.75 \pm .23$
7	$1.05 \pm .50$	$.83 \pm .28$
8	$1.12 \pm .54$	$.89 \pm .32$
9	$1.18 \pm .57$	$.93 \pm .34$
10	$1.19 \pm .55$	$.91 \pm .36$

RESULTS



Estimates across three seasons
ILI rate
Week start date
0
1
2
3
4
5
6
7
11/27/11
09/30/12
09/29/13
03/16/14
Ground truth
Baseline (ILI only)
Best (ILI+Twitter)

INFLUENZA DATA REVISIONS

An important caveat about historical data…

- Weekly ILINet values are subject to future revisions
- We were careful to train the models on the data that would have been available at the time of the prediction
  - But we evaluated on the gold standard value from the final report for the season

INFLUENZA DATA REVISIONS

An important caveat about historical data…

- The value initially reported has an average absolute difference from the final value of **.18**

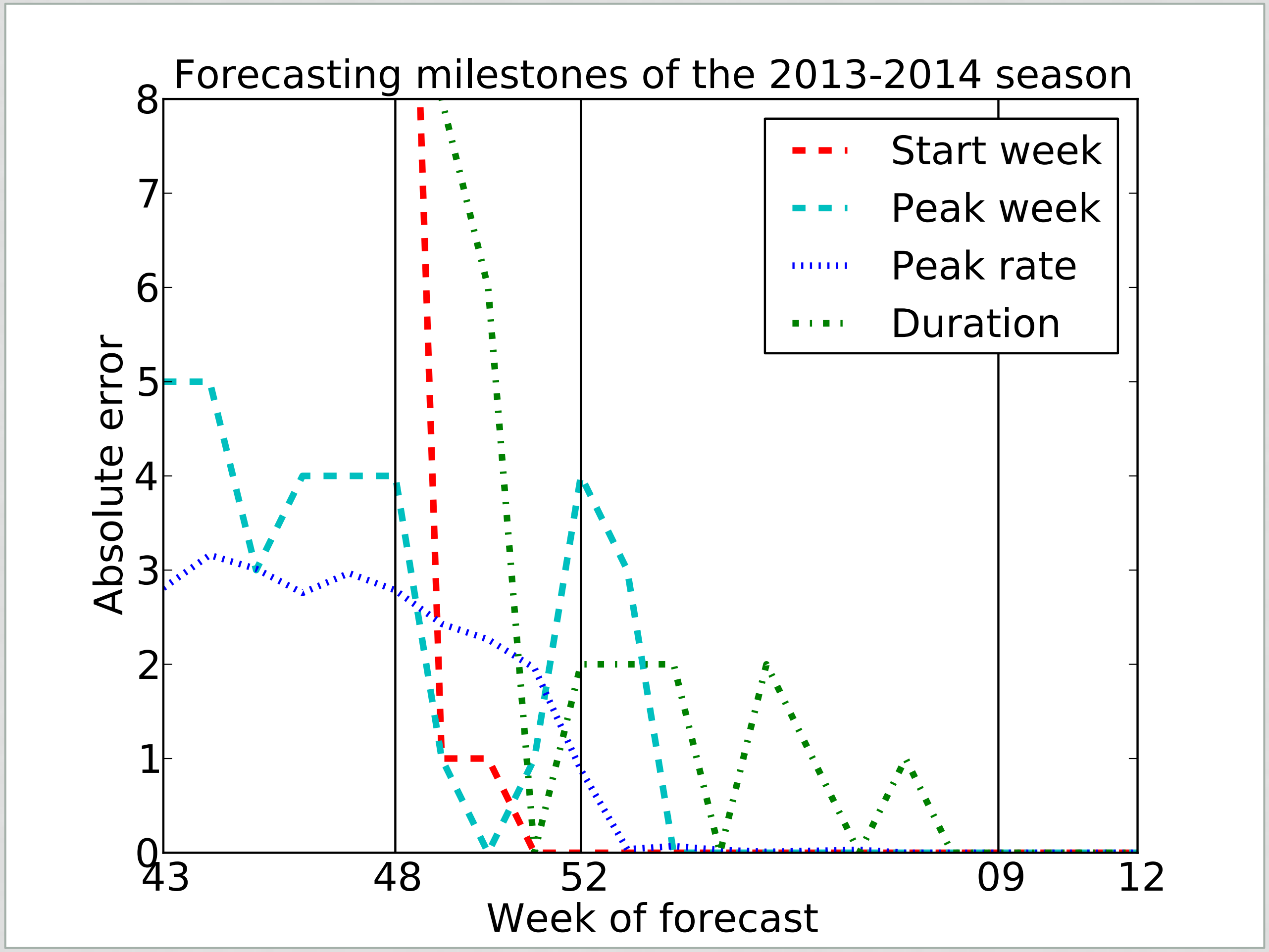
- The value reported after 3 weeks still has an average difference of **.10**

INFLUENZA DATA REVISIONS

Data	Average error per season		
	2011-12	**2012-13**	**2013-14**
ILINet (current)	.19	.30	.35
ILINet (final)	.11	.24	.26

- Error is greatly underestimated when using the final gold values instead of values available at time of forecast

INFLUENZA DATA REVISIONS



COMPARISON TO GOOGLE

- We also compared to Google Flu Trends

Data	Average error per season		
	2011-12	**2012-13**	**2013-14**
ILINet	.19	.30	.35
ILINet+Twitter	.15	.21	.21
ILINet+Google	.20	.44	.28

- Twitter improved nowcasting and forecasting more than Google

CONCLUSION #1

- Twitter improves influenza forecasting
  - For a given level of accuracy, including Twitter can give you 2-4 weeks of additional forecasting ability

- Twitter outperforms Google
  - At least in these three seasons
  - Google recently updated their model so comparison is difficult

CONCLUSION #2

- When using historical data, be careful to use data that actually would have been available at the time of model training
- Others have assumed these were the same
  - Our results showed that this has a substantial effect on performance

CONCLUSION #3

- Always compare to a simple time series baseline
- Our results showed that Twitter by itself is worse than using lagged ILINet data
  - No one had compared to this (using Twitter)
  - But we then showed that you can do even better by combining both!

THANK YOU