Identifying Change Points and Forecasting Influenza Trends Using Diverse Influenza-like Illness Surveillance Data in Houston, TX (2012 - 2016)

Susannah Paul, BS, Eunice Santos, MPH, Roberd M. Bostick, MD, MPH

Rollins School of Public Health, Emory University, 1516 Clifton Road, Atlanta, GA 30322

AIMS

- Use influenza-like illness (ILI) data to estimate magnitude, time of peak influenza activity, and start and end of influenza season
- Use historical ILI data to forecast potential activity

BACKGROUND

- Burden of influenza in the United States (U.S.) varies by
  - circulating strains
  - vaccine efficacy & effectiveness
  - changes in virus structure
  - season onset
- 2015 - 2016: Influenza activity peaked mid-March
  - One of later peaks on record
  - 25 million influenza illnesses
  - 11 million influenza-associated medical visits
  - 310,000 influenza-related hospitalizations
  - 12,000 pneumonia & influenza deaths (mortality count likely higher)
- 2013: Spot vaccine shortages in certain regions
  - 29 children died
  - 5,249 hospitalizations as of 2nd week of January
  - Public vaccine demand increased
- 2009: declaration of Public Health Emergency of International Concern due to influenza pandemic
  - Initially discovered via lab confirmations in April
  - H1N1 strain that caused pandemic now considered a seasonally circulating strain
- ILI used as indicator of influenza activity
  - Data gathered from sentinel providers or health care facilities via reporting or automation
  - **ILI = fever (at least100 ℉/37.8 ℃) + cough or sore throat without another known cause**
- How does influenza surveillance help public health?
  - Allows public health professionals time for **preparedness** & **prevention**
  - If activity identified early enough, can reduce influenza-associated mortality & morbidity
  - Advancements in technology provides **novel resources** to track health-seeking behaviors associated with influenza

METHODS: Change Point Analysis

- Use to detect multiple, subtle changes to evaluate whether significant change occurred
- Graph **cumulative sum (CUSUM) charts** to monitor average influenza-like activity at a time

$S_i = S_{i-1} + (X_i - \bar{X})$,

where X_i = weekly percentage of ILI,
S_{i-1} = previous week's cumulative sum, and
$\bar{X}$ = average.

- Generate **1,000 bootstrap samples** to estimate confidence interval around point of change
  - Samples represent random data that mimic behavior of cumulative sum if no change
- Calculate **mean square error (MSE)** as estimator of change

$MSE = \sum_{i=1}^{m} (X_i - \overline{X_1})^2 + \sum_{i=m+1}^{n} (X_i - \overline{X}_2)^2$, where

$$\overline{X_1} = \frac{\sum_{i=1}^{m} X_i}{m} \text{ and}$$

$$\overline{X}_2 = \frac{\sum_{i=m+1}^{n} X_i}{n-m}.$$

METHODS: ARIMA Forecasting

Autoregressive integrated moving average (ARIMA) models

- describe **autocorrelations** in data
- **forecast** time series

$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ where,

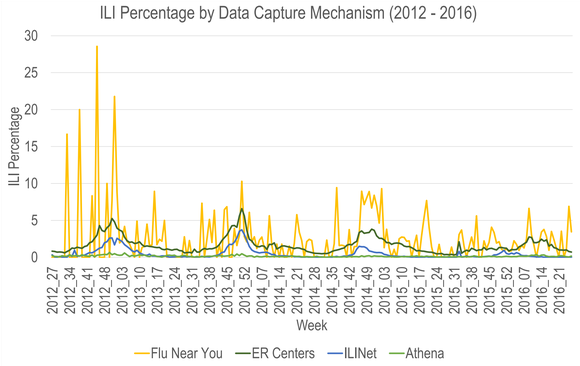
θ = moving average parameters of order q,
ϕ = autoregressive parameters of order p,
$\hat{y}_t$ = prediction estimates at time t,
y_{t-p} = lagged values of y, and
e = error term

Peak Weeks per Year and ILI % by Data Source

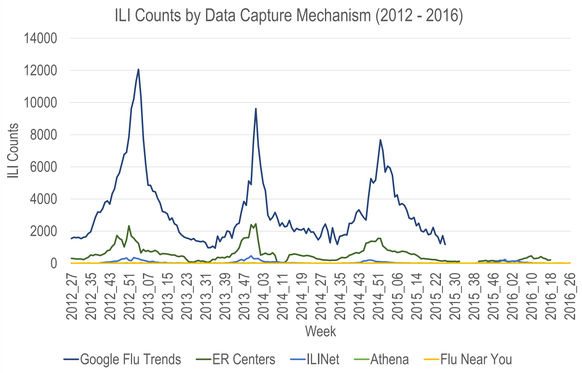
Weeks	Data Source	Peak Week[4]	Peak ILI %[4]	Mean[4]	Mean (Oct-May)
2012 - 2013	Flu Near You	2012 45th	28.57	3.29	3.84[1]
	Athena	2012 50th	0.66	0.24	0.26[1]
	ER Centers	2012 51st	5.25	1.94	2.41[1]
	ILINet	51st 2012	2.62	0.72	1.01[3]
2013 - 2014	Flu Near You	2013 51st	10.29	2.22	2.48[1]
	Athena	2013 51st	0.48	0.15	0.17[1]
	ER Centers	2013 51st	6.59	1.87	2.35[1]
	ILINet	2013 51st	3.73	0.67	0.96[1]
2014 - 2015	Flu Near You	2014 37th	9.43	2.77	3.56[2]
	Athena	2014 50th	0.21	0.09	0.11[2]
	ER Centers	2014 51st	3.80	1.55	2.00[2]
	ILINet	2014 46th	1.48	0.28	0.41[2]
2015 - 2016	Flu Near You	2016 25th	6.9	1.90	2.12[3]
	Athena	2016 12	0.37	0.13	0.14[3]
	ER Centers	2016 9th	2.76	1.24	1.49[3]
	ILINet	51st 2015	0.96	0.18	0.23[5]

[1]Mean calculated using weeks 40 - 22
[2]Mean calculated using weeks 40 - 21
[3]Mean calculated using weeks 39 - 21
[4]The time interval is week 27 through week 26 of following year
[5]Mean calculated using weeks 38 - 20

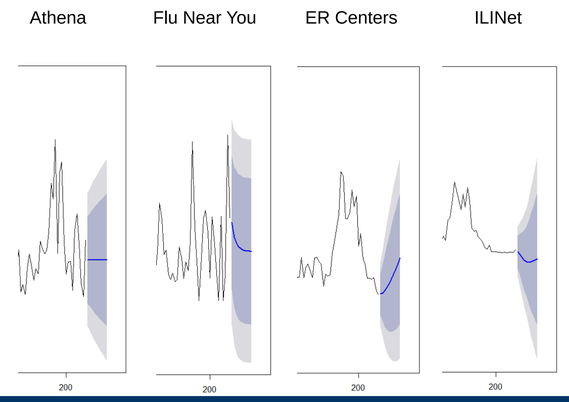
Peak Weeks of ILI Percentages and Counts



- Similar peaks among data sources for counts and percentages



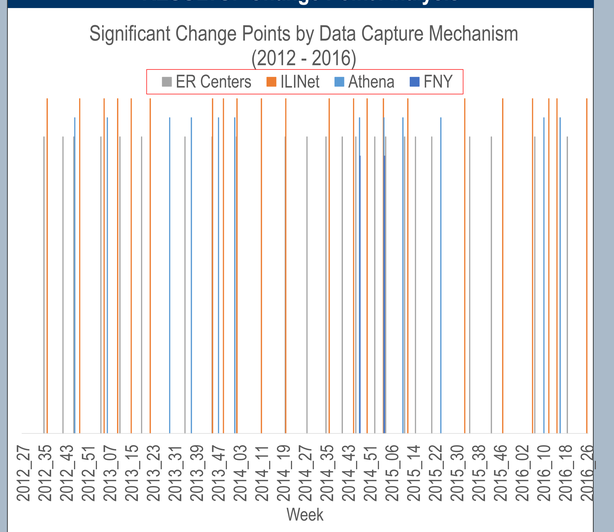
RESULTS: Forecasting



CONCLUSION: Forecasting

- All sources differ in forecasted trend
- Wide confidence interval (gray bands indicate 95% CIs)

RESULTS: Change Point Analysis



CONCLUSIONS: Change Point Analysis

- Clustered bands indicate likely time of significant change in average ILI activity
- ILINet more comprehensive in capturing season start & end
- FNY contains only 2 significant change points

Future Directions

- Utilize all available data sources to fit multivariate model that forecasts one point estimate more accurately
- Include other variables that influence influenza activity (age, vaccination)
- Evaluate validity of forecasts by comparing to observed data
- Compare sources to laboratory test results

REFERENCES

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. Otexts.com. Retrieved from https://www.otexts.org/fpp/8

Kass-Hout, T. a, Xu, Z., McMurray, P., Park, S., Buckeridge, D. L., Brownstein, J. S., … Groseclose, S. L. (2012). Application of change point analysis to daily influenza-like illness emergency department visits. *Journal of the American Medical Informatics Association : JAMIA*, *19*(6), 1075–1081. https://doi.org/10.1136/amiajnl-2011-000793

Taylor, W. A. (2000). Change-Point Analysis: A Powerful New Tool For Detecting Changes. Retrieved from http://www.variation.com/cpa/tech/changepoint.html

All analysis was completed using R 3.2, Microsoft Excel, and Change-Point Analyzer 2.3.

ACKNOWLEDGEMENTS

- Houston Health Department Office of Surveillance and Public Health Preparedness
- Harvard, Boston Children's Hospital, and Skoll Global Threats Fund
- Texas Department of State Health Services