
Research and Applications

Internet search query data improve forecasts of daily emergency department volume

Sam Tideman,¹ Mauricio Santillana,^{2,3} Jonathan Bickel,^{2,3} and Ben Reis^{2,3,4}

¹Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA, ²Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts, USA, ³Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA, and ⁴Predictive Medicine Group, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

Corresponding Author: Sam Tideman, MS, 4901 Searle Parkway Suite 200, Skokie, IL 60077, USA; samtideman@gmail.com

Received 20 March 2019; Revised 25 July 2019; Editorial Decision 5 August 2019; Accepted 6 August 2019

ABSTRACT

Objective: Emergency departments (EDs) are increasingly overcrowded. Forecasting patient visit volume is challenging. Reliable and accurate forecasting strategies may help improve resource allocation and mitigate the effects of overcrowding. Patterns related to weather, day of the week, season, and holidays have been previously used to forecast ED visits. Internet search activity has proven useful for predicting disease trends and offers a new opportunity to improve ED visit forecasting. This study tests whether Google search data and relevant statistical methods can improve the accuracy of ED volume forecasting compared with traditional data sources.

Materials and Methods: Seven years of historical daily ED arrivals were collected from Boston Children's Hospital. We used data from the public school calendar, National Oceanic and Atmospheric Administration, and Google Trends. Multiple linear models using LASSO (least absolute shrinkage and selection operator) for variable selection were created. The models were trained on 5 years of data and out-of-sample accuracy was judged using multiple error metrics on the final 2 years.

Results: All data sources added complementary predictive power. Our baseline day-of-the-week model recorded average percent errors of 10.99%. Autoregressive terms, calendar and weather data reduced errors to 7.71%. Search volume data reduced errors to 7.58% theoretically preventing 4 improperly staffed days.

Discussion: The predictive power provided by the search volume data may stem from the ability to capture population-level interaction with events, such as winter storms and infectious diseases, that traditional data sources alone miss.

Conclusions: This study demonstrates that search volume data can meaningfully improve forecasting of ED visit volume and could help improve quality and reduce cost.

Key words: emergency department visit prediction, emergency medicine, google search, predictive modeling, healthcare big data analytics

INTRODUCTION

Emergency departments (EDs) in the United States have been growing increasingly crowded. From 2001 to 2008, ED visit volume grew roughly twice as fast as the total U.S. population, while the number of total hospital beds shrank by about 198 000.^{1,2} Between 1993

and 2003, the number of EDs in the United States shrank by 425, while total visit volume increased by 26%.³ Similar trends are being reported across the world in studies from the United Kingdom,⁴ Australia,⁵ Singapore,⁶ Brazil,⁷ and South Korea.⁸ Furthermore, the responsibilities of EDs have expanded beyond traditional acute care

to include indigent care, disaster response, observational care, mental health services, and other services traditionally associated with primary care.^{9,10} For patients, this increased crowding has been shown to be associated with longer patient wait times, worsened quality of care,^{11,12} worsened health outcomes,¹³ risks to patient safety,¹⁴ delayed access to life-saving treatment,^{15,16} and lower patient satisfaction.¹⁷ For staff, crowding is associated with lower job satisfaction and increased career burnout.¹⁸ Improved forecasting of visit volume could help EDs better manage their staffing and other resources to accommodate fluctuations and reduce crowding. Additionally, improved forecasting accuracy could allow for decreased staffing during times of low demand and reduced departmental costs.

A number of previous studies have examined the relationship of different variables to ED visits. The majority of previous studies attempting to predict ED visit volume are descriptive in nature, assessing correlations among visit volume, calendar variables, and sometimes historical weather data.^{19–23} Some of these studies found weather to be correlated with ED visits, whereas others did not. The direction and magnitude of these correlations seems to vary by location. Many of the studies note patterns in ED visits relating to holidays, month of the year, and other special events. Almost all studies agree that day of the week is one of the strongest trends in ED visits, while exactly which particular days are busy seems to vary. Special local events such as flu outbreaks,²⁴ heat waves,²⁵ blackouts,²⁶ sporting events,²⁷ and publicity over Presidential heart surgery²⁸ have also been shown to have effects on local ED visits.

There is a growing body of literature that uses statistical forecasting methods to predict daily ED visit counts, testing model accuracy on an out-of-sample dataset, with generalizable measures of accuracy.^{2–8,29,30} We believe that this is the correct approach to allow comparisons across departments, data sources, and modeling techniques and best simulate the real-world performance of a forecasting model. Previous studies represent a diverse collection of EDs from 7 different countries and average daily visit counts ranging from 33 to 441. These studies have used a variety of methods such as linear regression, random forest regression, neural nets, and time series methods such as SARIMA (seasonal autoregressive integrated moving average). While these studies make use of weather and calendar variables, none of them take advantage of the growing number of potential data sources available through the Internet. Ekström et al³¹ utilized novel data sources by using internet traffic to a Swedish health department website to predict ED visits in the Stockholm area. Mean absolute percent error (MAPE), the average percent difference between the forecast and the actual visit count, is the most commonly used metric because it allows comparisons across departments of different sizes. Looking at the results of these previous studies (Figure 1), we can see that despite the diversity in department characteristics and statistical methods, many studies achieve similar overall MAPE. There also appears to be a downward trend with department size, suggesting that forecasts for larger departments can achieve a lower MAPE, as would be expected due to the law of large numbers.

Internet search volume has been seen as a possible additional data source for public health monitoring. In 2008, Polgreen et al³² used Yahoo searches for influenza surveillance. In 2009, Google released Google Flu Trends (GFT), a model based on anonymized, aggregated search activity of selected terms that predicts the official influenza-like illness levels in the United States reported by the Centers for Disease Control and Protection.^{33,34} However, GFT's early predictive performance was disappointing and led to doubts about the validity of its original approach.³⁵ A growing body of research

has attempted to expand upon GFT's results and has demonstrated noticeable improvements.^{36–40} Notably, Yang et al⁴¹ demonstrated how LASSO (least absolute shrinkage and selection operator) variable selection of autoregressive linear regression combined with dynamic recalibration can yield substantially better influenza-like illness predictions than can previous approaches.

The methods learned from predicting influenza and other diseases using search volume data^{42–47} could also be used to improve forecasts of ED visit volume beyond the accuracy of previous data sources alone. The techniques used in our study can easily be applied to other departments to forecast visit volume and potentially improve staffing practices.

MATERIALS AND METHODS

ED visit volume

This study was conducted at the Boston Children's Hospital (BCH) ED, an urban pediatric academic level 1 trauma center in Boston, Massachusetts. The BCH ED treats patients generally 21 years of age and younger, and handles an average of about 60 000 visits every year. The department's daily arrival counts were collected from BCH's electronic data warehouse for the dates of January 1, 2009, to December 31, 2015. The local institutional review board approved this study. Visits were classified into days based on the time of arrival as recorded in the electronic medical record. The model used the actual weather and search volume for the day being predicted for the purpose of analyzing the additional usefulness of the search volume data. These data would not be available when making actual predictions about the future. This was a retrospective study with the data split into a training set and an out-of-sample testing set. All models were fit using the training data and tested for accuracy on the testing set. The data used to train each of the proposed models included only historical data before January 1, 2014. The out-of-sample model test period included data from January 2, 2014, to December 31, 2015. We can see in the visit data in Figure 2 that there are 2 extreme events. In the middle of 2009, there is a spike in visits associated with the swine flu epidemic. In early 2015, there is a drop associated with extreme snowfall. This explains the differences in the minimum and maximum values between the training set and test set, as shown in Table 1. The swine flu event shapes the expectations of the trained model and the will models struggle to predict the response to an unprecedented snow storm.

Autoregressive variables

Lag variables representing the visit count from a specified interval of days before the present day were created. Figure 3 shows an autocorrelation function plot of daily ED arrivals from the training set. A strong weekly pattern in the data is evident from the peaks in the autocorrelation function plot for lags divisible by 7. There is also a strong annual trend, with the correlation bottoming out around 182 days in the past and strengthening around 365 days. For the models predicting next-day visits, we created variables for lags from 1 to 14, 21, 350, 357, and 364, corresponding to periods of high autocorrelation in the training set.

Calendar variables

Indicator variables were created for each day of the week, as well as major holidays: New Year's Day, Presidents' Day, July 4th, Halloween, Thanksgiving, Christmas, and New Year's Eve. Referencing the

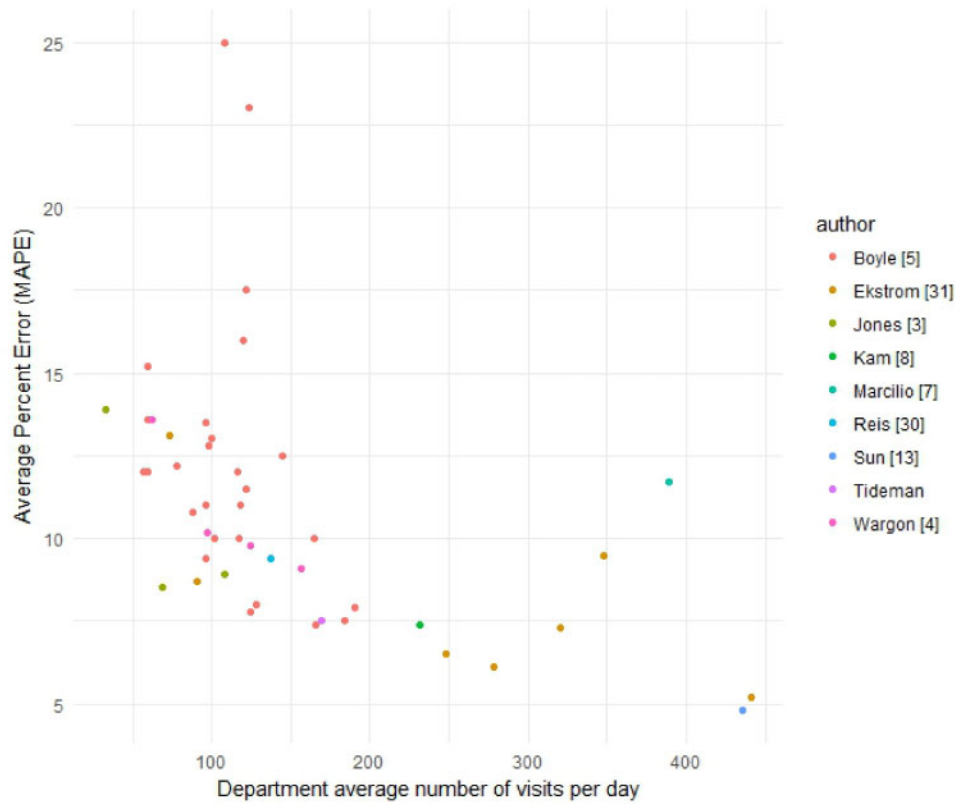


Figure 1. Accuracy of previous studies. MAPE: mean absolute percent error.

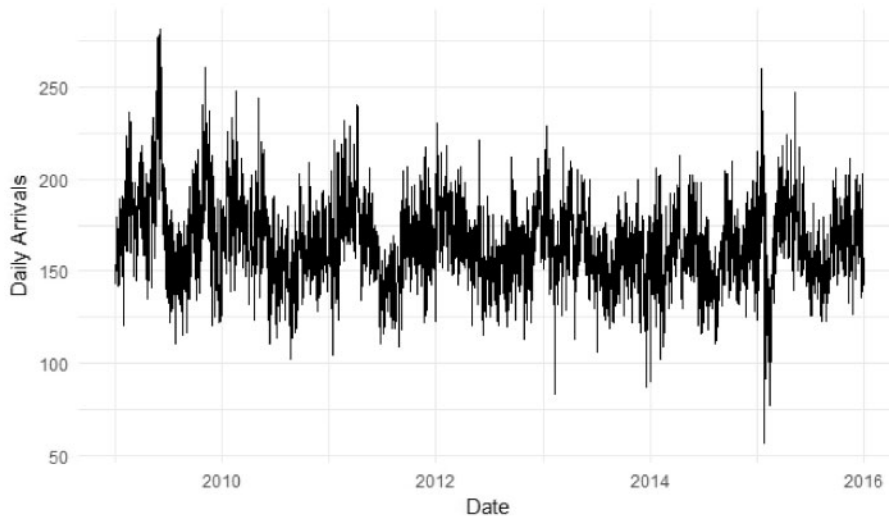


Figure 2. Daily emergency department arrivals at Boston Children’s Hospital for the whole study period.

Boston Public Schools calendar, indicator variables were created for summer vacation, winter vacation, first spring vacation, second spring vacation, holidays, Monday after a long weekend, Tuesday after a long weekend, other day after a long weekend or vacation, and early release days. While most previous studies created indicator variables for each month, we created a sine function with period 365.25 and a cosine function with period 365.25, starting on January 1, 2009. Due to the properties of trigonometric identities, regressing a sine and cosine function with the same period will pro-

duce a single sine function with the amplitude and shift that corresponds to the least-squares fit to the outcome variable.

$$a * \sin(x) + b * \cos(x) = c * \sin(x + d)$$

$$c = \sqrt{a^2 + b^2}d = atan2(a, b)$$

First, the advantage of this approach over the month indicator variable approach is that a maximum of 2 parameters are

Table 1. Comparison of data in training and test sets

		Training set				Test set			
		Mean	SD	Min	Max	Mean	SD	Min	Max
ED visits	Max	166	24.6	83	281	162.5	23.1	57	260
Temperature	Max	60.2	18	13	103	59.1	19.1	14	96
	Mean	53.2	17	6	92	51.8	18.3	8	84
	Min	45.6	16.4	-2	81	43.8	17.9	-3	74
Barometer	Max	30.1	0.2	29.2	30.8	30.1	0.2	29.6	30.9
	Mean	30	0.2	29	30.7	30	0.2	29.1	30.8
	Min	29.9	0.3	28.7	30.6	29.9	0.2	28.8	30.6
Visibility	Max	10	0.5	0	10	9.9	0.6	2	10
	Mean	8.8	2.1	0	10	8.9	2	1	10
	Min	6.5	4	0	10	6.7	4	0	10
Wind speed	Max	20.8	6.2	6	48	20.5	5.9	8	45
	Mean	10.7	3.8	2	31	10.6	3.6	3	26
Snowfall	Total	0.1	0.5	0	10.8	0.1	1	0	13.6
Fever	Vol	2374.1	547.4	798.8	6471.7	2674	517.1	1240.2	4612.4
Flu shot	Vol	555.7	933.8	0	16443.1	486.7	558.4	0	3682.5
Flu	Vol	3557	8722.6	140.2	163310.8	2169	1477.3	209.5	8462.8
Pneumonia	Vol	987.4	399.4	265.2	2960.7	1064	369.9	194.6	2262.4
Bronchitis	Vol	554	242	0	2908	537.2	217.4	136.7	1293.2
Snow	Vol	9964.8	12114.7	2371.6	214082.3	13351.6	19133.6	2628.8	219681.7
Allergies	Vol	920.2	423	166.3	3079.6	982.2	600	186.2	4687.4
Blizzard	Vol	930.4	2901.1	152.2	84259.9	1112.2	4133.1	139.6	72836.5
Forecast	Vol	6680	3111.1	1207.4	31667.9	7699.6	3803.6	2132.9	33665
Humidifier	Vol	612.1	446.7	0	3036	644.1	485.8	121.5	2929.9

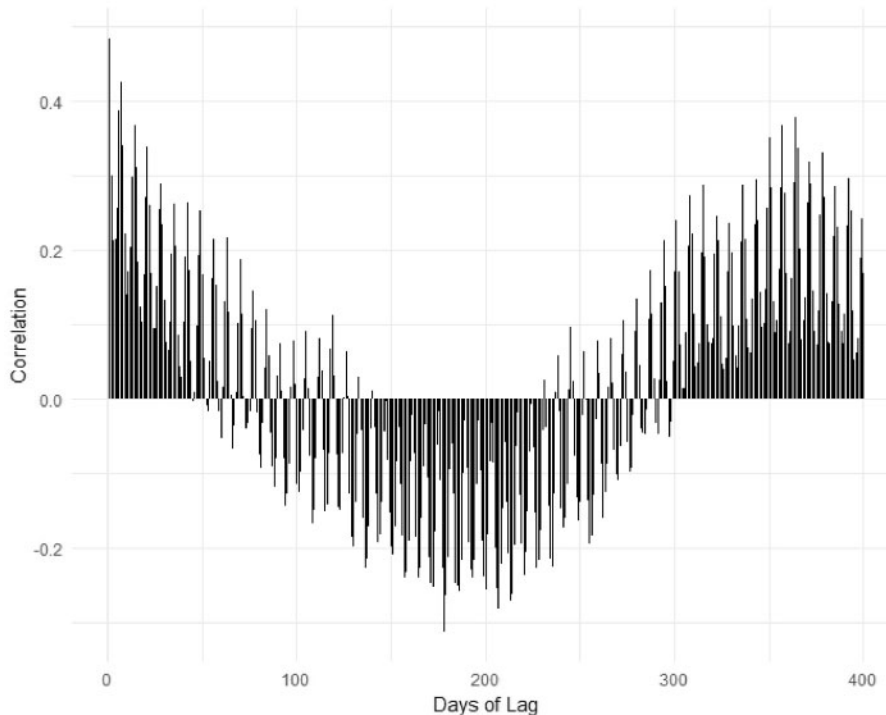


Figure 3. Autocorrelation function plot for training range.

being estimated, instead of up to 11. Second, this method demonstrates continuous change throughout the month and into the next month, instead of an abrupt change between the last day of a month and the first day of the next month. One disadvan-

tage is that this model implies a linear relationship between the outcome variable and the annual cycle, while indicator variables for months allow for nonlinear relationships with the annual cycle.

Weather variables

Weather data were downloaded from National Oceanic and Atmospheric Administration⁴⁸ for the Boston Logan Airport weather station. Daily data were collected on minimum, mean, and maximum temperature, barometer, and visibility. Visibility was measured on a 10-point scale with 10 being the clearest visibility. Daily maximum and mean wind speeds were also recorded, but not minimum wind speed, as this was almost always zero. Daily snowfall was also collected. Daily occurrences of weather warnings, watches, and advisories for blizzards, heat, floods, freezing rain, high wind, hurricanes, severe thunderstorms, tornadoes, wind chills, winter storms, and winter weather were represented as binary indicator variables. Table 1 shows that the weather data were very similar between the testing and training sets, with the noticeable exception of snowfall. The average snowfall was more than twice as high in the testing set as in the training set. Furthermore, the maximum snowfall was almost 3 inches higher in the testing set than in the training set. During January and February of 2015, Boston experienced record levels of snowfall (colloquially labeled by the locals as “snowmageddon”).

Google search volume

The daily Google search fraction for a set of weather- and health-related search terms were retrieved through the Google Health application programming interface (API). The structure of the dataset retrieved from the API was the indicator value of the fraction of daily searches containing each selected term for the Google-defined Boston metro area (Google search location code: “eUS-MA-506”). We chose 10 search terms to test in the model: *allergies*, *blizzard*, *bronchitis*, *fever*, *flu*, *flu shot*, *forecast*, *humidifier*, *pneumonia*, and *snow*. We chose to limit the number of terms to 10 so as to not overwhelm the model selection process. The terms were chosen based on our theory that disease, symptom, and weather-related terms would be most predictive. Google Correlate is a publicly available tool that finds search terms that are most correlated with a weekly time series. We were unable to use Google Correlate to find terms whose Boston-area search volume was correlated with the daily ED visit count at BCH, as Google Correlate is only available for analyzing search volumes at the national level. We did include *humidifier* in our list of search terms because Google Correlate suggested that it was highly correlated with searches for *flu* at the national level.

Table 1 shows that the characteristics of the training and testing sets are broadly similar. The most notable exception appears to be *flu*. The average daily search volume for *flu* is much higher in the training period, with the maximum indicator value of 163 310 searches on April 30, 2009. This corresponds with the beginning of the media coverage of the swine flu epidemic of 2009. However, the number of ED visits during this period is not particularly high, and the peak number of visits related to flu did not occur until September 2009. This time period serves as a cautionary note that local search volume for terms such as *flu* can sometimes become disconnected from the conditions generating the local demand for the ED. This may reflect the fact that local search activity is not exclusively related to the actual local prevalence of the disease being searched, but can also reflect general interest in the subject. Some search terms have much higher volume than others: *snow*, *forecast*, and *fever* are the 3 highest in the testing period. Search terms with higher average volumes had lower variance relative to their search volumes, making it easier to identify potentially predictive signals compared with terms with lower volume and higher variance relative to their search volumes. While Yang et al⁴¹ found that dynamic

reweighting of search terms was needed to take into account the changing search behavior, our own modeling experiments showed that a dynamic reweighting approach did not significantly increase the forecasting accuracy in comparison to a statically trained model.

To verify that the weather and search volume datasets contain independent information, we performed pairwise correlations and a principal component analysis. In Figure 4, the highest correlation is between recorded snowfall and searches for *snow*. Searches for *allergies*, *bronchitis*, *humidifier*, and *pneumonia* also show positive and negative correlation with the temperature variables. This is likely due to the seasonal nature of all these variables. However, most of the variables have little or no pairwise correlation. The principal component analysis yielded similar results. There is some shared information in the weather and search volume datasets, but most of the variables are independent and cannot be reduced.

Accuracy metrics

A number of accuracy metrics were used to evaluate the performance of each of the models in our study. The formulas associated for these quantities can be found in Yang et al,⁴¹ such as MAPE, a metric that reflects the average prediction error in a percent scale. One hundred percent error suggests over- or underprediction by as much as the observed quantity and zero percent error suggests a perfect prediction. This metric is used frequently and thus, allows for comparison across different studies and departments. Root mean squared error is the cost function minimized during the model fitting process and has the scale of the outcome variable. This metric cannot be used to compare with other studies conducted in different EDs with heterogeneous daily visits values. R^2 , or the coefficient of determination, can be interpreted as the percent of the variance in the predicted variable explained by the model. This is a useful measure, as it can be compared across studies and departments. Its limitation is that it does not have a practical application when trying to understand the magnitude of uncertainty in the forecasts. Finally, we used an ad hoc metric, namely, the percent absolute percent error >20%, as a way to measure the percentage of days in which the model prediction error was >20%. After discussing the staffing needs with subject matter experts in the ED, we were informed that on days when staffing was mismatched by >20%, they often had to bring in extra on-call staff. Therefore, we designed this measure as a way to assess how our predictions could have affected staffing strategies in the ED.

Formulation of prediction models

For each of our models, we followed a 2-stage approach that included a variable selection stage and the development of an optimal model.

Variable selection

We used a regularized multivariable linear regression, LASSO, to select the strongest predictors in our set of input variables, and as a way to minimize the use of redundant information in our final models. The benefit of LASSO regression is its ability to zero-out redundant input variables (or predictors), therefore serving as a model selection tool.³⁶ With nearly 90 covariates in our full model, LASSO provides a systematic form of variable selection. Before performing LASSO regression, all the covariates were normalized so that the coefficients from the LASSO regression could be interpreted as reflecting the parameters' importance in the model. The optimal parameter was chosen from a series of possible values as the one with the minimum root mean squared error in 10-fold cross validation on

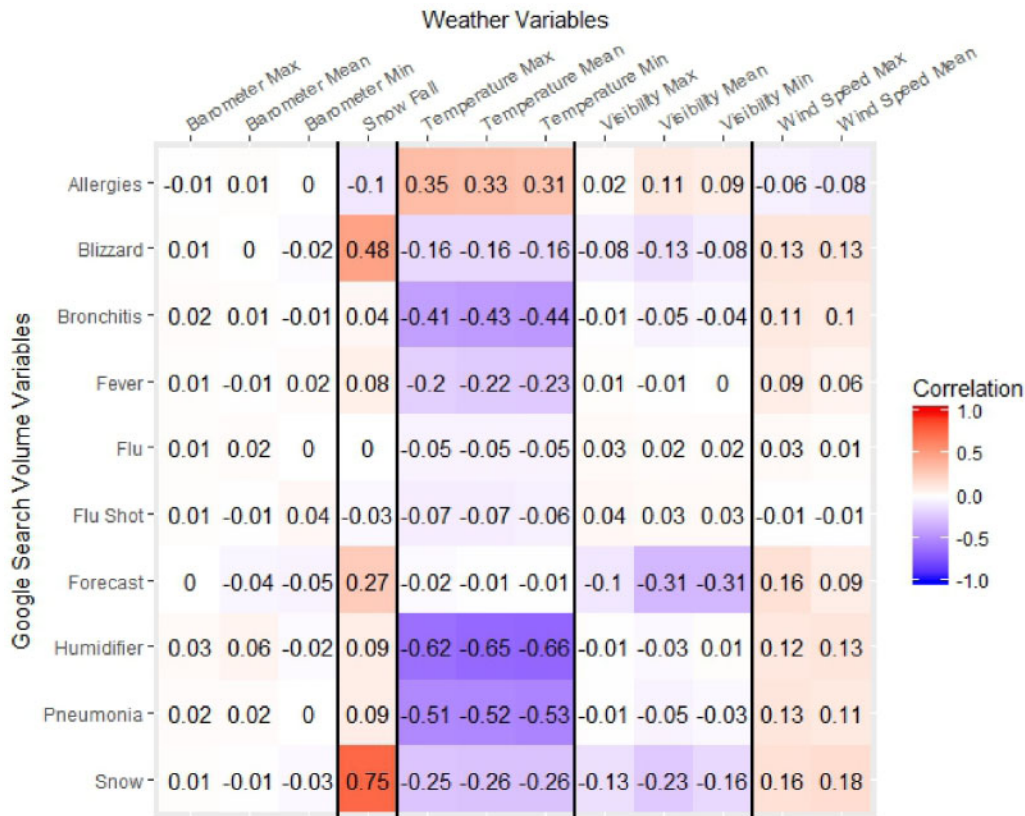


Figure 4. Pairwise correlation of weather and search volume data.

the training set. A known downside of using LASSO regression is that the variable selection may be unstable; thus, zeroing out the contribution of useful variables in the optimization process.⁴⁹

Optimal model

To build an optimal model, our second stage consisted of selecting the variables consistently chosen by LASSO and fitted a standard linear regression without regularization.

Gradual incorporation of data sources

To test the contribution of each data source, we classified our variables into 5 groups: day of the week, autoregressive, calendar, weather, and search volume. We used our modeling approach on all 5 groups, adding each data source sequentially to determine its contribution to the forecasting accuracy. We started with day of the week as a baseline, since the BCH ED currently uses long-term, day-of-the-week averages to determine staffing. We then produced retrospective predictions using each resulting model for each of the subsets of data sources. We chose to order the data source categories according to how commonly they were used in the previous literature on the subject²⁻⁸ with the search volume data being tested last.

RESULTS

Retrospective simulated predictions for daily visit counts were produced for the testing period using all the data available before the predicted day. Five models were fit on the training set period using different data sources and tested for accuracy on 4 different metrics,

as shown in Table 2. Each model also included the data sources listed in the models above it.

Variables from each data source were selected and added to the model to improve the out-of-sample accuracy without overfitting the model to the in-sample dataset. Each model with an added data source performed better across all measures of accuracy than the models with fewer data sources. Calendar variables appeared to provide more forecasting improvement in the later models because more calendar terms were included once the model accounts for possible confounding due to weather. Weather variables provided a large improvement in accuracy in an already crowded model reducing MAPE from 8.41% to 7.71%. Weather variables, such as maximum temperature and mean visibility, are among the most important variables in the model.

Even in a model with 66 other covariates, the 7 selected search variables provided additional out-of-sample forecasting accuracy, lowering MAPE from 7.71% to 7.58% and root mean squared error from 12.29 to 12.07 (Figure 5). The meaningful impact of adding search variables can be assessed with the metric percent absolute percent error >20%. This corresponds to preventing 2 improperly staffed days a year in comparison with the full model without search volume data.

To further compare the relative value of the weather and search volume datasets, we compared a model using everything but weather to a model using everything but the search volume data. Remarkably, these 2 models achieved the exact same MAPE of 7.71% in the test set. This suggests that the value of these 2 data sources in very similar even though the search volume dataset only contains 10 terms.

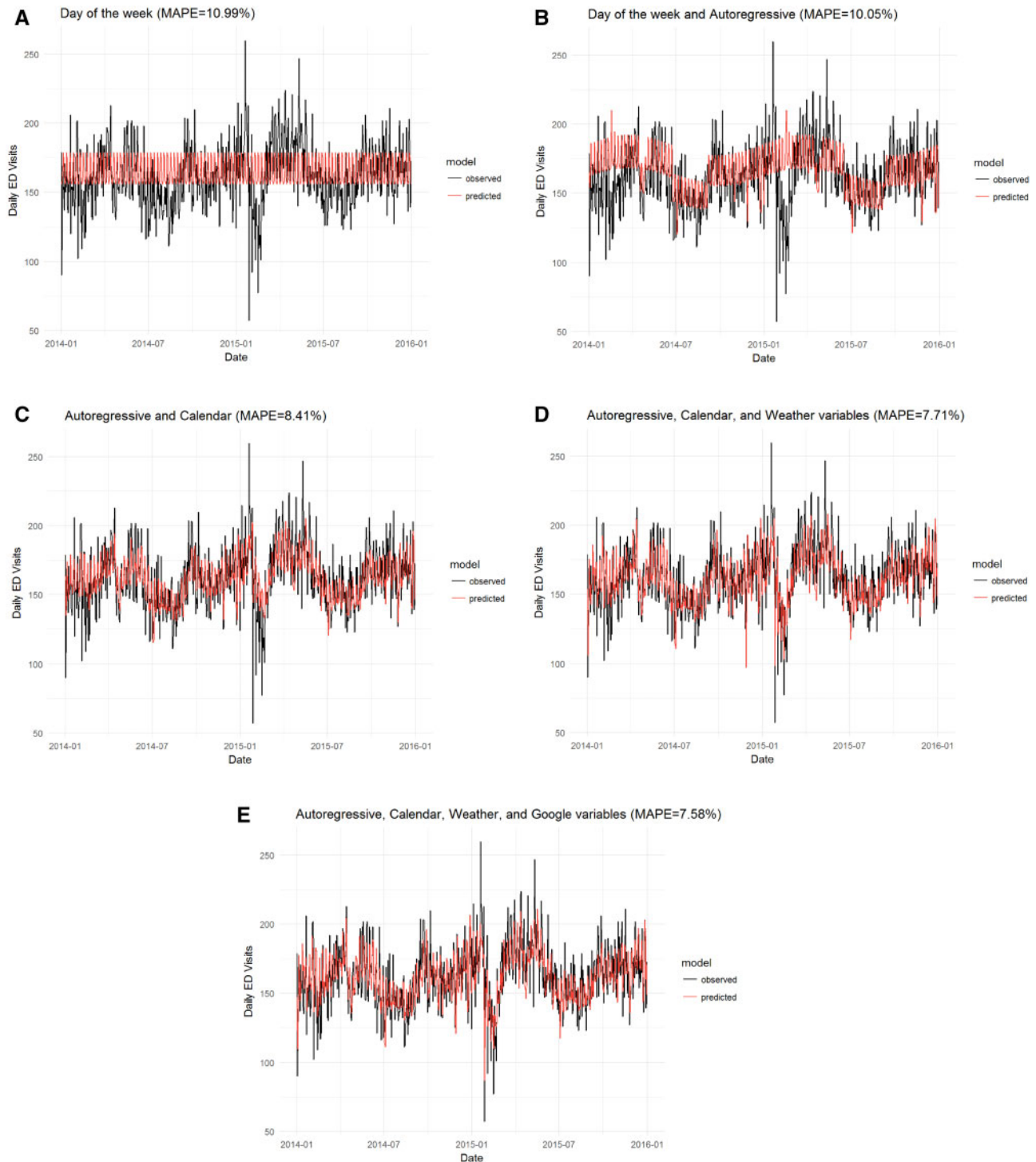


Figure 5. Observed and predicted values for different models. ED: emergency department; MAPE: mean absolute percent error.

Among the search terms included in the final model are terms relating to weather, diseases, and specific symptoms. The word *snow* was the most important search term (Table 3). Because the outcome is log transformed, the coefficient can be interpreted as the percent change in the estimated daily ED visit volume associated with a 1-SD increase in search volume with all other variables held equal. On January 27, 2015 Boston experienced 19.7 inches of snowfall. This corresponded to an 18.7-SD increase in search volume relat-

ing to the word *snow*. Accordingly, the final model lowered its estimate by 29.8 visits. While the models without search volume data also had very high estimates for the effect of snowfall, the model with search volume data has higher accuracy. We understand that this demonstrates that search volume data measure the public's reaction to, and awareness of, these snow storms that weather data alone cannot capture. On May 10, 2015 during the height of the spring pollen season, searches containing the word

Table 2. Comparison of model accuracy

Model name	MAPE (%)	RMSE	R ²	PAPE > 20% (%)
Day of the week	10.99	16.73	0.135	11.23
Calendar	10.05	15.12	0.213	8.36
Autoregressive	8.41	12.93	0.443	5.21
Weather	7.71	12.29	0.551	3.84
Search volume	7.58	12.08	0.569	3.29

MAPE: mean absolute percent error; PAPE: percent absolute percent error; RMSE: root mean square error.

Table 3. Coefficient estimates and importance rank of search terms in final model

Search term	Coefficient	Rank
<i>snow</i>	-0.00967	13
<i>allergies</i>	0.00557	23
<i>blizzard</i>	-0.00464	28
<i>flu shot</i>	0.00461	29
<i>fever</i>	0.00452	31
<i>forecast</i>	-0.00354	41
<i>pneumonia</i>	0.00329	42
<i>flu</i>	0	N/A
<i>bronchitis</i>	0	N/A
<i>humidifier</i>	0	N/A

N/A: not available.

allergies surged 7.1 SDs above average. Accordingly, the final model raised its estimate by 7.1 visits. This demonstrates how search volume data can capture local population level health information and translate that into associated estimates of its effect on ED visits.

DISCUSSION

Based on 5 years of historical data, the forecasting model provides good accuracy in predicting daily ED visit volume. In addition to weather data and calendar variables, search volume provides additional useful information that improves the accuracy of the forecasts. Search terms related to weather and allergies were the most useful in our model, suggesting that search volume data provide additional information that can be used to forecast ED visits than weather and calendar data can alone. Our results suggest that over our 2-year test period, the best model would have only 24 instances in which the model was off by more than 20%, probably leading to improperly staffed days in the ED. This is compared with 28 days in the model without search volume data and 82 days in our baseline day-of-the-week model. While in our study we only chose 10 search terms to investigate, using additional search terms could provide further prediction accuracy. Our side-by-side comparison showed that a model with search volume data and without weather data performed equally well as a model with weather data and no search volume data. With the use of the API, Google search volume data could prove to be easier and more valuable than weather data for ED volume prediction.

The 2-step modeling approach of using LASSO regression for variable selection and log-linear regression for model fitting can successfully handle variable selection and model calibration for datasets of many variables. This system reliably chose useful variables from se-

quentially added data sources and improved out-of-sample forecasting accuracy while reducing the change of overfitting to the new data.

This methodology can be applied widely at other EDs to improve forecasting of patient volumes to improve staff and resource planning and reduce overcrowding. Any ED could attempt to build a similar prediction model with any combination of data sources available. The variables that are important, the relationship between the variables and the outcome, as well as the accuracy, are likely to be unique to each location. To implement this procedure a model would need to be fit on local historical data. This framework can also be used to fit predictions to any amount of time into the future that is desired. In the master's thesis on which much of this article is based, it was shown that predictions for greater time horizons into the future led to decreases in accuracy.⁵⁰ The number of days into the future at which the model is calibrated is a choice that each location would need to make after considering the trade-off between the preparation time needed to take meaningful action and the accuracy of those predictions. Once a model is trained, a data pipeline is needed to send clean, up-to-date data to the model to make predictions, most likely on a daily basis. Future work should include validating the methodology at additional diverse sites, incorporating additional novel sources of data, and prospectively testing the models in real time.

CONCLUSION

This study demonstrates that Google search volume data can meaningfully improve forecasting of ED visit volume. Additionally, we describe a broad and flexible methodology that can be used to test at other departments and other possible data sources. This framework could be implemented to help EDs with resource planning to help reduce problems of overcrowding and improve quality and cost.

CONTRIBUTIONS

ST, MS, and BR conceived research. JB collected institutional data. ST collected external data and normalized all datasets. ST, MS, and BR contributed to the design of predictive models. ST implemented and analyzed predictive models. ST wrote the first draft of the manuscript. All authors contributed and approved to the final version of the manuscript.

ETHICS APPROVAL

This research was approved by the Institutional Review Board at Boston Children's Hospital (Protocol number IRB-P00023195) and was determined to not represent human subjects research and informed consent was waived because the data was deidentified.

ACKNOWLEDGMENTS

The datasets generated and analyzed during the current study are available in a GitHub repository (https://github.com/sat157/BCH_ED_Prediction).

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Kharbanda AB, Hall M, Shah SS, *et al.* Variation in resource utilization across a national sample of pediatric emergency departments. *J Pediatr* 2013; 163 (1): 230–6.
2. Gonzalez Morganti K, Bauhoff S, Blanchard JC, *et al.* *The Evolving Role of Emergency Departments in the United States*. Santa Monica, CA: Rand Corporation; 2013.
3. Jones SS, Thomas A, Evans RS, *et al.* Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008; 15 (2): 159–70.
4. Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med* 2010; 17 (9): 970–8.
5. Boyle J, Jessup M, Crilly J, *et al.* Predicting emergency department admissions. *Emerg Med J* 2012; 29 (5): 358–65.
6. Sun Y, Heng BH, Seow YT, Seow E. Forecasting daily attendance at an emergency department to aid resource planning. *BMC Emerg Med* 2009; 9 (1): 1.
7. Marcilio I, Hajat S, Gouveia N. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Acad Emerg Med* 2013; 20 (8): 769–77.
8. Kam HJ, Sung JO, Park RW. Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthc Inform Res* 2010; 16 (3): 158–65.
9. Moskop JC, Sklar DP, Geiderman JM, *et al.* Emergency department crowding, part 1—concept, causes, and moral consequences. *Ann Emerg Med* 2009; 53 (5): 605–11.
10. Institute of Medicine. *Hospital-Based Emergency Care: At the Breaking Point*. Washington, DC: National Academies Press; 2007.
11. Miró O, Antonio MT, Jiménez S, *et al.* Decreased health care quality associated with emergency department overcrowding. *Eur J Emerg Med* 1999; 6: 105–7.
12. Pines JM, Hollander JE. Emergency department crowding is associated with poor care for patients with severe pain. *Ann Emerg Med* 2008; 51 (1): 1–5.
13. Sun BC, Hsia RY, Weiss RE, *et al.* Effect of emergency department crowding on outcomes of admitted patients. *Ann Emerg Med* 2013; 61 (6): 605–11.
14. Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emerg Med J* 2003; 20 (5): 402–5.
15. McCarthy ML, Zeger SL, Ding R, *et al.* Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients. *Ann Emerg Med* 2009; 54 (4): 492–503.
16. Schull MJ, Vermeulen M, Slaughter G, *et al.* Emergency department crowding and thrombolysis delays in acute myocardial infarction. *Ann Emerg Med* 2004; 44 (6): 577–85.
17. Pines JM, Iyer S, Disbot M, *et al.* The effect of emergency department crowding on patient satisfaction for admitted patients. *Acad Emerg Med* 2008; 15 (9): 825–31.
18. Rondeau KV, Francescutti LH. Emergency department overcrowding: the impact of resource scarcity on physician job satisfaction. *J Healthc Manag* 2005; 50 (5): 327–40.
19. Zibners LM, Bonsu BK, Hayes JR, *et al.* Local weather effects on emergency department visits: a time series and regression analysis. *Pediatr Emerg Care* 2006; 22 (2): 104–6.
20. Friede KA, Osborne MC, Erickson DJ, *et al.* Predicting trauma admissions: the effect of weather, weekday, and other variables. *Minn Med* 2009; 92 (11): 47–9.
21. Tai C-C, Lee C-C, Shih C-L, Chen S-C. Effects of ambient temperature on volume, specialty composition and triage levels of emergency department visits. *Emerg Med J* 2007; 24 (9): 641–4.
22. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med* 2001; 8 (1): 48–53.
23. Attia MW, Edward R. Effect of weather on the number and the nature of visits to a pediatric ED. *Am J Emerg Med* 1998; 16 (4): 374–5.
24. Olson DR, Heffernan RT, Paladini M, Konty K, Weiss D, Mostashari F. Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Med* 2007; 4 (8): e247.
25. Mathes RW, Ito K, Lane K, Matte TD. Real-time surveillance of heat-related morbidity: Relation to excess mortality associated with extreme heat. *PLoS One* 2017; 12 (9): e0184364.
26. Marx MA, Rodriguez CV, Greenko J, *et al.* Diarrheal illness detected through syndromic surveillance after a massive power outage: New York City, August 2003. *Am J Public Health* 2006; 96 (3): 547–53.
27. Reis BY, Brownstein JS, Mandl KD. Running outside the baseline: Impact of the 2004 major league baseball postseason on emergency department use. *Ann Emerg Med* 2005; 46 (4): 386–7.
28. Mathes RW, Ito K, Matte T. Assessing syndromic surveillance of cardiovascular outcomes from emergency department chief complaint data in New York City. *PLoS One* 2011; 6 (2): e14677.
29. Xu M, Wong TC, Chin KS. Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network. *Decis Support Syst* 2013; 54 (3): 1488–98.
30. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak* 2003; 3 (1): 1.
31. Ekström A, Kurland L, Farrokhnia N, Castrén M, Nordberg M. Forecasting emergency department visits using internet data. *Ann Emerg Med* 2015; 65 (4): 436–42.
32. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis* 2008; 47 (11): 1443–8.
33. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009; 49 (10): 1557–64.
34. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457 (7232): 1012–4.
35. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis. *Science* 2014; 343 (6176): 1203–5.
36. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011; 6 (8): e23610.
37. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013; 9 (10): e1003256.
38. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) google flu trends? *Am J Prev Med* 2014; 47 (3): 341–7.
39. Davidson MW, Haim DA, Radin JM. Using networks to combine “big data” and traditional surveillance to improve influenza predictions. *Sci Rep* 2015; 5: 8154.
40. Stefansen C. Google flu trends gets a brand new engine. Google Research Blog; 2014. <https://ai.googleblog.com/2014/10/google-flu-trends-gets-brand-new-engine.html> Accessed June 21, 2019.
41. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. *Proc Natl Acad Sci U S A* 2015; 112 (47): 14473–8.
42. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS Negl Trop Dis* 2017; 11 (1): e0005295.
43. Yang S, Kou SC, Lu F, Brownstein JS, Brooke N, Santillana M. Advances in using Internet searches to track dengue. *PLoS Comput Biol* 2017; 13 (7): e1005607.
44. Lu FS, Hou S, Baltrusaitis K, *et al.* Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston Metropolitan. *JMIR Public Health Surveill* 2018; 4 (1): e4.
45. Lampos V, Miller AC, Crossan S, Stefansen C. Advances in nowcasting influenza-like illness rates using search query logs. *Sci Rep* 2015; 5 (1): 12760.
46. Yang S, Santillana M, Brownstein JS, Gray J, Richardson S, Kou SC. Using electronic health records and Internet search information for accurate influenza forecasting. *BMC Infect Dis* 2017; 17 (1): 332.

-
47. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol* 2015; 11 (10): e1004513.
 48. National Oceanographic and Atmospheric Association. Weather. <http://www.noaa.gov/weather> Accessed August, 3, 2016.
 49. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996; 58 (1): 267–88.
 50. Tideman S. *Internet Search Query Data Improves Forecasts of Daily Emergency Department Volume* [master's thesis]. Cambridge, MA, Harvard University; 2016.