Examining Patterns of Influenza Vaccination in Social Media

Xiaolei Huang[1], Michael C. Smith[2], Michael J. Paul[1*], Dmytro Ryzhkov[1], Sandra C. Quinn[3], David A. Broniatowski[2], Mark Dredze[4]

[1] University of Colorado, Boulder, CO 80304

[2] George Washington University, Washington, DC 20052

[3] University of Maryland, College Park, MD 20742

[4] Johns Hopkins University, Baltimore, MD 21211

[*] Corresponding author: mpaul@colorado.edu

Abstract

Traditional data on influenza vaccination has several limitations: high cost, limited coverage of underrepresented groups, and low sensitivity to emerging public health issues. Social media, such as Twitter, provide an alternative way to understand a population's vaccination-related opinions and behaviors. In this study, we build and employ several natural language classifiers to examine and analyze behavioral patterns regarding influenza vaccination in Twitter across three dimensions: temporality (by week and month), geography (by US region), and demography (by gender). Our best results are highly correlated official government data, with a correlation over 0.90, providing validation of our approach. We then suggest a number of directions for future work.

Introduction

Effective vaccination programs require the collection of detailed information about a population's vaccine-related beliefs and behaviors (Zell et al. 2000). Understanding vaccination adherence and refusal, and the motivations thereof, is especially critical for creating effective systems of health communication (Downs, de Bruin, and Fischhoff 2008). A range of annual surveys in the United States aim to capture information to improve our understanding of population beliefs toward vaccination, primarily through panels and telephone surveys (Parker et al. 2013). However, these methods are too slow to be used in real-time, and traditional surveys can underrepresent young, urban participants and minorities (Keeter et al. 2006).

Mining social media can potentially address these concerns, as data can be analyzed in real-time and reflect populations that are difficult to reach with traditional surveys (Dredze et al. 2015). In this study, we employ natural language classifiers to infer vaccine-related intentions from Twitter messages, focusing specifically on the influenza (flu) vaccine, which is delivered to the population annually. Analyzing a dataset spanning three flu seasons (2013–2016), we seek to measure levels of flu vaccine uptake aggregated by time (week or month), geography (US region), and demographic group (gender), where geographic and demographic attributes are inferred from user profiles. This information is a starting point for understanding vaccination behavior in near real time.

We compare our Twitter findings to published government survey data about vaccination from the US Centers for Disease Control and Prevention (CDC). We find strong agreement between social media-derived statistics and gold standard data, with temporal correlations as high as 0.90 and geographic correlations as high as .67. These findings suggest opportunities to use social media to improve traditional surveys (for example, by computing statistics at finer temporal and geographic resolutions), and we also discuss challenges and directions for future improvements.

Related Work

A large body of work has used social media, specifically Twitter, to monitor population health (Abbasi et al. 2014; Paul et al. 2016). Most relevant are works exploring influenza and vaccination, summarized below. To the best of our knowledge, this is the first study to look specifically at the geographic and demographic patterns of the flu vaccine in social media.

Many researchers have used Twitter data to monitor influenza prevalence (Culotta 2010; Signorini, Segre, and Polgreen 2011), with the best systems using natural language processing methods to identify relevant tweets (Aramaki, Maskawa, and Morita 2011; Doan, Vo, and Collier 2012; Lamb, Paul, and Dredze 2013). Beyond population-level surveillance, research has also shown that tweets can predict disease transmission between individuals (Sadilek, Kautz, and Silenzio 2012), can estimate crowding in hospitals (Broniatowski et al. 2015), and can forecast future prevalence (Paul, Dredze, and Broniatowski 2014).

Less work has used social media to study vaccination patterns, though some research has analyzed attitudes and sentiment toward vaccination using Twitter (Salathe and Khandelwal 2011; Salathé et al. 2013; Dunn et al. 2015; Dredze, Broniatowski, and Hilyard 2016). The work of (Salathe and Khandelwal 2011) found correlations between sentiment and vaccination rates across geography in the United States.

Data Collection and Classification

We built a tweet classifier to track flu vaccinations over time, as well as by geography and gender. We compared the extracted patterns to official government data to validate our models. Our analysis covers three flu vaccination seasons beginning with the 2013–14 season.

Vaccine Data

We collected official government data from the US Centers for Disease Control and Prevention (CDC) on influenza vaccination. The data includes vaccination coverage by month, by geographic regions defined by the US Department of Health and Human Services (HHS), and by demographic group. The data can be downloaded from the CDC's FluVaxView system.[1] The CDC's estimates come from several large national surveys: the Behavioral Risk Factor Surveillance System (BRFSS, which targets adults), the National Health Interview Survey (NHIS), and the National Immunization Surveys (NIS, which targets children).

Twitter Data

We have continuously collected tweets containing a set of health-related keywords (including flu-related words) using the Twitter streaming API since 2012, described in our prior work (Paul and Dredze 2014). Our goal was to build vaccination behavior classifiers on a labeled dataset, a sample of the data, and analyze temporal, demographic, geographic patterns on the rest of data. For this study, we filtered this large dataset for tweets containing at least one flu-related term (*flu*, *influenza*) and at least one vaccine-related term (*shot(s)*, *vaccine(s)*, *vaccination(s)*).

We removed retweets and non-English tweets,[2] although we did not filter tweets specifically for US tweets except for our comparisons by geographic region (where each region is defined by a set of US states). We inferred the US state for tweets using the Carmen geolocation system (Dredze et al. 2013). The final dataset contained 1,007,582 tweets.

Data Annotation

We collected annotations for a random sample of 10,000 tweets from our collection to be used as training data. Annotations were obtained from Amazon Mechanical Turk (Callison-Burch and Dredze 2010), with three independent annotations per tweet. Tweets were labeled with the following:

- Does this message indicate that someone received, or intended to receive, a flu vaccine? (yes or no)
  - If yes: has the person already received a vaccine, or do they intend to receive the vaccine in the future.

We rejected annotators whose agreement was anomalously low (percentage agreement was $\leq 60\%$). Three bad annotators were removed from our final dataset. We took a majority vote on the remaining 29,970 annotations to obtain the final labels. If there was not a majority label, then we defaulted to the 'no' label.

The final dataset contained 10,000 tweets, with 67.2% labeled as positive for intent, with a kappa score of 0.793, using Fleiss' kappa (Fleiss 1973) to measure the inter-annotator agreement.

[1] `http://www.cdc.gov/flu/fluvaxview/`

[2] Using the `langid` package: `https://github.com/saffsd/langid.py`

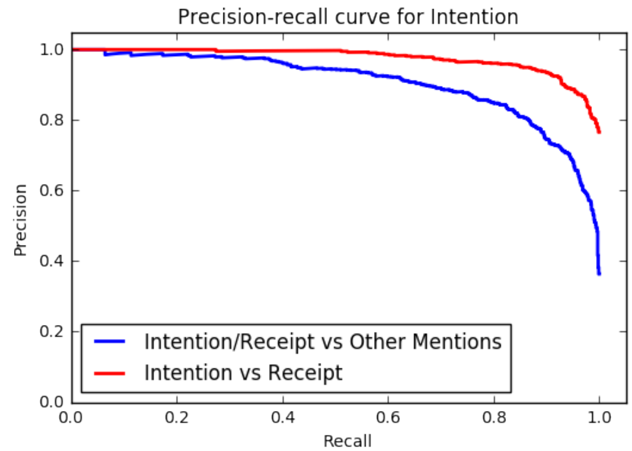


Figure 1: Precision-recall curves for the classification tasks related to vaccine intention.

	Prec.	Rec.	F1
Received/Intends vs Other	.84	.80	.82
Received vs Intends	.90	.95	.93

Table 1: Classifier performance from 5-fold cross-validation.

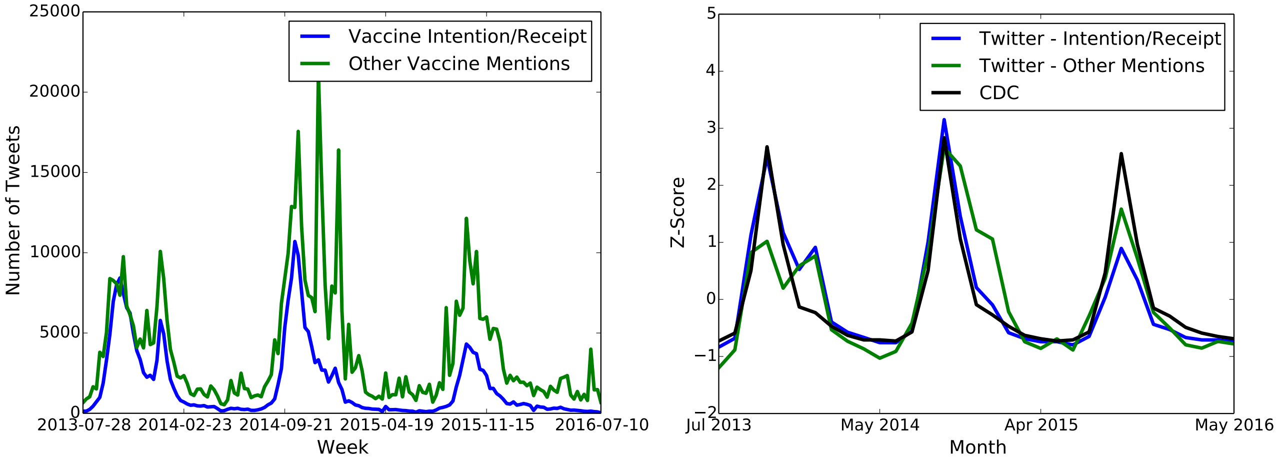
Data Classification

In order to analyze vaccine patterns in the dataset, we trained and built classifiers with the following steps. First, we preprocessed the tweets by removing both URLs and stop words. We initially experimented with n-gram features (unigram, bigram, trigram) and different classification models (SVM, Multinomial Naive Bayes, RandomForest) with default parameters. Cross-validation (5-fold) and F1-measure were used to evaluate each classifier's performance. We then chose the best-performing classifier, logistic regression, in our further experiments.

Our classifiers were implemented using `sklearn` (Pedregosa et al. 2011). We used $\ell 2$ regularization with default parameters. The classifiers used TF-IDF weighted n-gram features, as well as part-of-speech counts from TweeboParser (Gimpel et al. 2011), and emoji and emoticon features derived from two open lexicons (Kralj Novak et al. 2015; Mohammad and Turney 2013). Feature counts were normalized to sum to 1 within each tweet. Cross-validation (5-fold) results for the logistic regression classifier are shown in Table 1.

The precision and recall curves are shown in Figure 1. While we used the default classification threshold for our analysis (probability of positive label ≥ 0.5), the curves show that we can adjust the threshold to achieve desired tradeoffs in precision and recall.

30.5% of tweets were classified as indicating the receipt or intention to receive a flu vaccine. Of positively classified tweets, 87.9% indicated that someone had received a flu vaccine (in contrast to only intending to).



(a) Weekly (b) Monthly

Figure 2: Twitter and CDC vaccination prevalence estimates by week (left) and month (right). CDC data are only available by month.

Analysis of Vaccination Patterns

In this section, we analyze patterns of vaccination behavior, specifically whether a tweeter received or intended to receive a flu vaccine. We ran the intention classifier over the entire dataset (three years of tweets) to identify tweets indicating vaccine intention or receipt, and then computed the volume of classified tweets within different groups of interest. Our analysis was applied across three dimensions: time, geography, and demography, giving an in-depth characterization of the temporal, geographic, and demographic patterns of flu vaccine intentions in social media. We validated our methods by comparing our results to CDC data.

By Time

Figure 2(a) shows the weekly counts of tweets classified as receiving or intending to receive a flu vaccine (blue) and the counts of all other tweets in the flu vaccine data (green). That is, the blue line represents positively classified tweets while the green line represents negatively classified tweets. It is visually apparent that the positively classified tweets in blue provide a smoother and more consistent curve. There are seasonal peaks every October (when flu vaccines are distributed in the US), with relatively few bumps in the curves outside of that peak when using the classified tweets. The other tweets in the dataset, in contrast, have very high week-to-week variability, with numerous spikes that do not fit the seasonal trends. This is strong evidence that our classifier is reducing the noise and improving our identification of vaccine behaviors in our original dataset.

To evaluate the temporal trends against gold standard data, we compared our extracted tweet counts to the CDC's data on vaccination coverage. Specifically, the CDC provides the percentage of American adults who received a flu vaccination in a given month. The monthly counts from all data sources are shown in Figure 2(b). Rather than raw counts, we show standardized counts (z-scores) so that the Twitter and CDC counts are comparable. We see that the positively classified tweet counts in blue are a closer fit to the CDC data ($r = .903$) than the negatively classified tweets in green ($r = .816$), although the difference between the two correlations is not statistically significant ($p = .187$). However, we believe the performance difference between the two Twitter trends is understated when viewing monthly counts, as much of the noise that is seen in the weekly counts is smoothed out in the monthly counts.

Finally, we compared the temporal trends of tweets classified as having received a vaccine versus intending to receive a vaccine. Intention tweets have a weaker correlation with the CDC data ($r = .869$) than tweets expressing vaccine receipt ($r = .911$) which is what we would expect, although we do not find major differences between them. One reason is that many of the intention tweets indicate that the tweeter will receive a vaccine in the near future (e.g., "I need to get my flu shot today"), so such tweets would still be accurate for counting vaccine coverage in that week or month.

By Geography

We also explored whether geographic variability can be accurately captured by the Twitter counts. We aggregated tweet counts for each of the 10 HHS regions. Because Twitter usage varies by location, it is important to normalize location-specific counts so that they can be compared. To do this, we divided the vaccine-related counts per region by the total number of tweets from that region, using a random sample of tweets from the Twitter streaming API.

The region-specific counts of positively classified Twitter have a strong correlation with the CDC's regional percentages, at $r = .674$. This is significantly higher than the correlation using negatively classified tweets ($r = .420$).

By Demographics

Finally, we examined how vaccine tweet counts vary across one demographic attribute, gender. We inferred the gender of each Twitter user in the dataset using the Demographer tool[3]

[3]https://bitbucket.org/mdredze/demographer

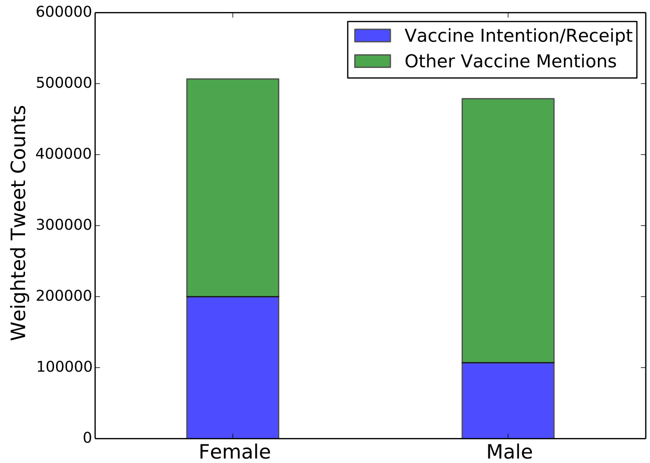


Figure 3: Twitter counts by gender.

(Knowles, Carroll, and Dredze 2016). We then grouped the Twitter counts by gender, shown in Figure 3. To adjust for the fact that Twitter users are not evenly balanced by gender, we weighted the counts, dividing by the proportion for that gender. Surveys have estimated that 53% of Twitter users are male (Pew Research Center 2015), while our Demographer statistics put this number at 59%. We used the median of 56% for weighting.

We find that both genders tweet about flu vaccines to roughly the same degree (with slightly more tweets by female users after weighting), but female users are substantially more likely to tweet about receiving or intending to receive a vaccine, while male users are more likely to tweet about vaccines in other ways. The difference in proportions is significant with $p \ll .01$.

This finding is consistent with CDC data. For example, in 2011 the CDC reports that among American adults, 42.0% of women were vaccinated for flu, compared to 35.4% of men. Thus, it makes sense that women are more likely to tweet about receiving a flu vaccine.

Future Direction: Sentiment Classification

In this project, we also attempted to classify flu vaccine tweets by sentiment, with the goal of examining attitudes toward vaccination. However, the classification performance was not strong enough to include in the study. Specifically, we asked the Mechanical Turk annotators to label each tweet with positive, negative, or neutral sentiment. We trained classifiers using the same process as above, but only achieved F1-scores of 0.42 and 0.62 for positive and negative sentiment, respectively.

A primary reason for the poor performance seems to be poor annotation quality, as the sentiment labels had low annotator agreement ($\kappa = 0.401$). This is likely in part due to the ambiguity in what is meant by sentiment. For example, consider the message, "That shot hurt me :(stupid flu shot nurse!" This message expresses negative sentiment toward that particularly experience, but not toward vaccination in general, so it is unclear what the appropriate label is without additional guidance. Thus, we may need to collect new annotations with finer-grained categorizations and more explicit instructions on what constitutes positive and negative sentiment in the context of vaccination.

Discussion and Conclusion

These experiments represent preliminary findings which lay the groundwork for an in-depth analysis of how we can track vaccine attitudes and behaviors on Twitter. We plan to extend this initial work to other demographic categories, such as age and race/ethnicity. While these early experiments have focused on validating against existing CDC statistics, we plan to next conduct analyses that would provide new insights not captured by existing research.

Another direction is to work on improving the classifiers, through more extensive parameter tuning and better features, such as word embeddings (Mikolov et al. 2013), as well as other classifiers including neural networks. We also intend to explore the precision-recall tradeoff in more depth, to understand how this affects the correlations with CDC data.

Acknowledgements

This research was supported in part by the National Institute of General Medical Sciences grant (R01GM114771).

References

Abbasi, A.; Adjeroh, D.; Dredze, M.; Paul, M. J.; Zahedi, F. M.; Zhao, H.; Walia, N.; Jain, H.; Sanvanson, P.; Shaker, R.; Huesch, M. D.; Beal, R.; Zheng, W.; Abate, M.; and Ross, A. 2014. Social Media Analytics for Smart Health. *IEEE Intelligent Systems* 29(2):60–80.

Aramaki, E.; Maskawa, S.; and Morita, M. 2011. Twitter catches the flu: Detecting influenza epidemics using Twitter. In *EMNLP*.

Broniatowski, D. A.; Dredze, M.; Paul, J. M.; and Dugas, A. 2015. Using social media to perform local influenza surveillance in an inner-city hospital: A retrospective observational study. *JMIR Public Health Surveill* 1(1):e5.

Callison-Burch, C., and Dredze, M. 2010. Creating speech and language data with amazon's mechanical turk. In *Workshop on Creating Speech and Language Data With Mechanical Turk at NAACL-HLT*.

Culotta, A. 2010. Towards detecting influenza epidemics by analyzing Twitter messages. In *ACM Workshop on Soc.Med. Analytics*.

Doan, S.; Vo, B. K. H.; and Collier, N. 2012. An analysis of Twitter messages in the 2011 Tohoku Earthquake. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, volume 91 LNICST, 58–66.

Downs, J. S.; de Bruin, W. B.; and Fischhoff, B. 2008. Parents' vaccination comprehension and decisions. *Vaccine* 26(12):1595 – 1607.

Dredze, M.; Paul, M. J.; Bergsma, S.; and Tran, H. 2013. Carmen: A Twitter geolocation system with applications to public health. In *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*.

Dredze, M.; Broniatowski, D. A.; Smith, M.; and Hilyard, K. M. 2015. Understanding vaccine refusal: Why we need social media now. *American Journal of Preventive Medicine*.

Dredze, M.; Broniatowski, D. A.; and Hilyard, K. M. 2016. Zika vaccine misconceptions: A social media analysis. *Vaccine*.

Dunn, A. G.; Leask, J.; Zhou, X.; Mandl, K. D.; and Coiera, E. 2015. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: An observational study. *J. Med. Internet Res.* 17(6):e144.

Fleiss, J. 1973. *Statistical methods for rates and proportions Rates and proportions*. Wiley.

Gimpel, K.; Schneider, N.; O'Connor, B.; Das, D.; Mills, D.; Eisenstein, J.; Heilman, M.; Yogatama, D.; Flanigan, J.; and Smith, N. A. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Association for Computational Linguistics (ACL)*.

Keeter, S.; Kennedy, C.; Dimock, M.; Best, J.; and Craighill, P. 2006. Gauging the impact of growing nonresponse on estimates from a national rdd telephone survey. *Public Opinion Quarterly* 70(5):759–779.

Knowles, R.; Carroll, J.; and Dredze, M. 2016. Demographer: Extremely simple name demographics. In *EMNLP Workshop on Natural Language Processing and Computational Social Science*.

Kralj Novak, P.; Smailović, J.; Sluban, B.; and Mozetič, I. 2015. Sentiment of emojis. *PLoS ONE* 10(12):1–22.

Lamb, A.; Paul, M. J.; and Dredze, M. 2013. Separating Fact from Fear: Tracking Flu Infections on Twitter. *NAACL*.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*.

Mohammad, S. M., and Turney, P. D. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29(3):436–465.

Parker, A. M.; Vardavas, R.; Marcum, C. S.; and Gidengil, C. A. 2013. Conscious consideration of herd immunity in influenza vaccination decisions. *American Journal of Preventive Medicine* 45(1):118–121.

Paul, M. J., and Dredze, M. 2014. Discovering health topics in social media using topic models. *PLoS ONE* 9(8):e103408.

Paul, M. J.; Sarker, A.; Brownstein, J. S.; Nikfarjam, A.; Scotch, M.; Smith, K. L.; and Gonzalez, G. 2016. Social media mining for public health monitoring and surveillance. In *Pacific Symposium on Biocomputing (PSB)*.

Paul, M. J.; Dredze, M.; and Broniatowski, D. 2014. Twitter improves influenza forecasting. *PLOS Currents Outbreaks*.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12(Oct):2825–2830.

Pew Research Center. 2015. Social media update: Twitter users. `http://www.pewinternet.org/2015/01/09/social-media-update-2014/pi_2015-01-09_social-media-new_01/`.

Sadilek, A.; Kautz, H.; and Silenzio, V. 2012. Predicting disease transmission from geo-tagged micro-blog data. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Salathe, M., and Khandelwal, S. 2011. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Comput Biol* 7(10):e1002199.

Salathé, M.; Vu, D. Q.; Khandelwal, S.; and Hunter, D. R. 2013. The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science* 2(1).

Signorini, A.; Segre, A. M.; and Polgreen, P. M. 2011. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE* 6(5):e19467.

Zell, E. R.; Ezzati-Rice, T. M.; Battaglia, M. P.; and Wright, R. A. 2000. National Immunization Survey: The methodology of a vaccination surveillance system. *Public Health Reports* 115(1):65–77.