Research Practice and Methods

What Can Digital Disease Detection Learn from (an External Revision to) Google Flu Trends?*

Mauricio Santillana, Ph.D., M.S.[1]; D. Wendong Zhang M.A.[1]; Benjamin M. Althouse, Ph.D., Sc.M.[2] John W. Ayers, Ph.D. M.A.[3]

[1] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

[2] Santa Fe Institute, Santa Fe, NM, USA

[3] Graduate School of Public Health, San Diego State University, San Diego, CA, USA

Abstract

Background: Google Flu Trends (GFT) claimed to generate real-time and valid predictions of population influenza-like illness (ILI) using search queries, heralding acclaim and replication across public health. However, recent studies have questioned the validity of GFT.

Purpose: We propose an alternative methodology that better realizes the potential of GFT, with collateral value for digital disease detection broadly.

Methods: Our alterative method automatically selects the specific queries to monitor and autonomously updates the model each week as new information about CDC-reported ILI becomes available, as developed in 2013. Root mean squared errors (RMSE) and Pearson correlations comparing predicted ILI (the proportion of patient visits indicative of ILI) with subsequently observed ILI were used to judge model performance.

Results: During the height of the H1N1 pandemic (2 August to 22 December 2009) and 2012/13 season (30 September 2012 to April 12 2013) GFT's predictions had RMSEs of 0.023 and 0.022 (i.e., hypothetically if GFT predicted 0.061 ILI one week, it is expected to error by 0.023), and correlations of r=0.916 and r=0.927. Our alternative method had RMSEs of 0.006 and 0.009, and correlations of r=0.961 and r=0.919 for the same periods. Critically, during these important periods the alternative method yielded more accurate ILI predictions every week, and was typically more accurate during other influenza seasons.

Conclusion: GFT may be inaccurate, but improved methodological underpinnings can yield accurate predictions with significant public health value. Applying similar methods elsewhere can improve digital disease detection, with broader transparency, improved accuracy, and real-world public health impacts.

Citation: M Santillana et al. **What Can Digital Disease Detection Learn from (an External Revision to) Google Flu Trends?** *American Journal of Preventive Medicine*, Volume 47, Issue 3, 2014.

Introduction

The rapid escalation of digital methods is changing public health surveillance.[1-3] By harvesting web data, investigators claim to validly estimate cholera,[4] dengue,[5,6] influenza,[7,8] kidney stones,[9,10], listerosis,[11] methicillin-resistant staphylococcus aureus,[12] mental health,[13] and tobacco control[14] trends, but are they actually valid?

The novelty of digital data has generally remained the central focus in these studies, while the methods and disinterested interpretations have been overlooked. As a result, studies demonstrating modest associations with ground truth outcomes (*e.g.*, r^2=0.15,[14] r^2=0.25,[4] or r^2=0.62[11]) have been presented as accurate, without further model validation. Most notable is *Google Flu Trends* (GFT),[8] not because it is potentially the most flawed, but because it is oft-cited and many subsequent studies modeled their approach after GFT[6,12,15,16] or even used weaker methods.

Concerns about GFT's accuracy came to light via media reports in 2009 when it misrepresented the epidemic curve and required updating that Autumn.[17] Again during 2012/13 media reports questioned the revised GFT,[18] followed by separate peer-reviewed analyses suggesting GFT was typically inferior to traditional sentinels due to inaccuracies.[19,20] Most recently, Google again updated their model to improve GFT operation, but did not identify their revisions nor describe its performance.[21] Many, unfortunately, are unaware of these problems.

The head of the CDC Influenza Surveillance and Outbreak Response Team told *Nature News* she monitored GFT (and other digital disease detection sentinels) "all the time", likely in the sense that some data are better than no data.[18] Moreover, some investigators are beginning to use GFT as ground truth for epidemiologic studies.[22] But if GFT (and similar systems for other outcomes) are invalid, should public health officials be paying any attention and should investigators rely on these data?

We remain optimistic about the future of GFT and digital disease detection broadly,[23-26] because a methodological problem has a methodological solution. A transparent, external evaluation of GFT, as a case study for the scientific status of digital disease detection, is presented. An alternative methodology capable of outperforming GFT is then proposed, with potential application across digital disease detection.

Methods

The methodology behind the original GFT and the 2009 revision (published in 2011) consisted of building a regression for CDC-reported influenza-like illnesses (ILI) with a *single* explanatory variable. Originally, the single variable was the mean trend for the 45 search terms with the strongest correlation with ILI for September 28, 2003 through March 11, 2007.[8] The revised

GFT single variable was the mean trend for the most correlated search terms (~160, the exact number unknown) for September 28, 2003 through September 13, 2009 after removing terms related to influenza complications and general interest in influenza.[17] Both generated ILI predictions at time (t) using search data from (t) and historic periods, but because ILI is delayed, these estimates are typically available 1 to 2 weeks earlier.

The original and updated GFT methodology is problematic for at least three reasons. First, combining multiple queries into a single variable ignores the variability in individual search query tendencies over time, how certain unique queries may be better predictors.[8] Second, the exclusion of search queries in the revised GFT relies on investigator opinion rather than any empirical evidence.[17] Third, the model is static, assuming that queries predicting ILI at time (t) will equally predict ILI at time (t + x years). The language of search undoubtedly changes over time (e.g., swine flu, H1N1, H1N9, etc.) and must be accounted for any prediction model.

Our alternative approach, inspired by data-assimilation techniques,[27,28] supervised machine learning,[29] and artificial intelligence,[30] expands upon (a) their single explanatory-variable approach, by allowing multiple individual queries to contribute independently to the prediction, (b) their quasi-non empirical search query selection, by empirically choosing search queries that maximize predictive accuracy in real-time, and (c) their use of manual revisions, by dynamically updating how individual queries predict influenza each week to ensure accurate prediction across a changing search and influenza landscape. All improve the transparency of GFT.

These revisions are executed in a multivariable linear model with different coefficients for each specific search query trend. Each query is prescribed a different level of importance based on its coefficient, determined by a LASSO for the best predicting and most parsimonious model.[31] The coefficients change each week based on refitting the model at time (t) to ILI through (t – 2), representing the latest available CDC ILI estimates if the system was running in real-time. The equation is given by the simplified regression notation below:

$$\text{logit}\,(I(t)) = \sum_{i=1}^{n} a_i(t)\ \text{logit}\,(Q_i(t)) + e, \qquad \text{(eq. 1)}$$

where I(t) is the percentage of ILI physician visits, $Q_i(t)$ is the query fraction for term *i* at time t, and $a_i(t)$ is the multiplicative coefficient associated to such term at time t, e is the normally-distributed error term, and logit(p) = ln(p/(1 - p)).

The criterion is the weekly percent of confirmed influenza-like illness (ILI) related physician visits (*e.g.*, fever > 100*F and cough or sore throat as a percentage of all outpatient healthcare provider visits nationally to over 2,900 reporting clinics) publicly available at (*cdc.gov/flu/weekly/*). Predictions were made for the entire US, following the strategy used in media critiques and Google's revisions.[17,21]

For the explanatory variables, our alternative conceptually relies on the same initial selection strategy used by GFT: time-trends for the Google queries most strongly correlated with ILI. *Google Correlate (google.org/trend/correlate)* returned the weekly z-scores of the query fractions spanning January 2004 to May 2013 of the 100 Google terms most correlated with CDC-reported ILI from September 28, 2003 to March 22, 2009, as accessed October 17, 2013 and still available today. These 100 terms were then the fixed inputs in the model, where the query fraction of each term is the total count of a query term in the US aggregated weekly and then scaled by the total count of all queries issued in the same week.[8] This choice simulates the optimal selection process that could occur if this methodology operated in real-time and had been implemented since March 22, 2009. Unlike Google, our alternative method did not filter this list of queries based on either the intensity of the correlation or the query content, but used all 100 most-correlated queries anticipating the model would then select queries to maximize prediction.

For performance benchmarks our alternative model was implemented for March 22, 2009 through May 2013, fitting to early trends (since January 1 2004) as a training period. Pearson correlation coefficients and root mean squared errors (RMSE) comparing predicted ILI with subsequently observed ILI were used to judge model performance. The latter was added to accurately assess performance when correlation may not (i.e., two trends may have r=1.00 correlation, but differ by a constant factor).

For comparisons to GFT we relied on the estimates made by GFT as events unfolded. Since GFT was updated after the H1N1 season in 2011 the GFT webpage returns those updated results, but the national predictions originally made by GFT were downloaded through December 27, 2009 from Figure 1 in Google revision.[17] For later periods, GFT predictions were simply downloaded in the summer of 2013 (google.org/flutrends). All analyses were conducted in R 2.15.3.

Results

Figure 1 presents GFT's and our alternative model predictions alongside the subsequently observed ILI trends, where it is readily apparent that the alternative produced more accurate predictions.

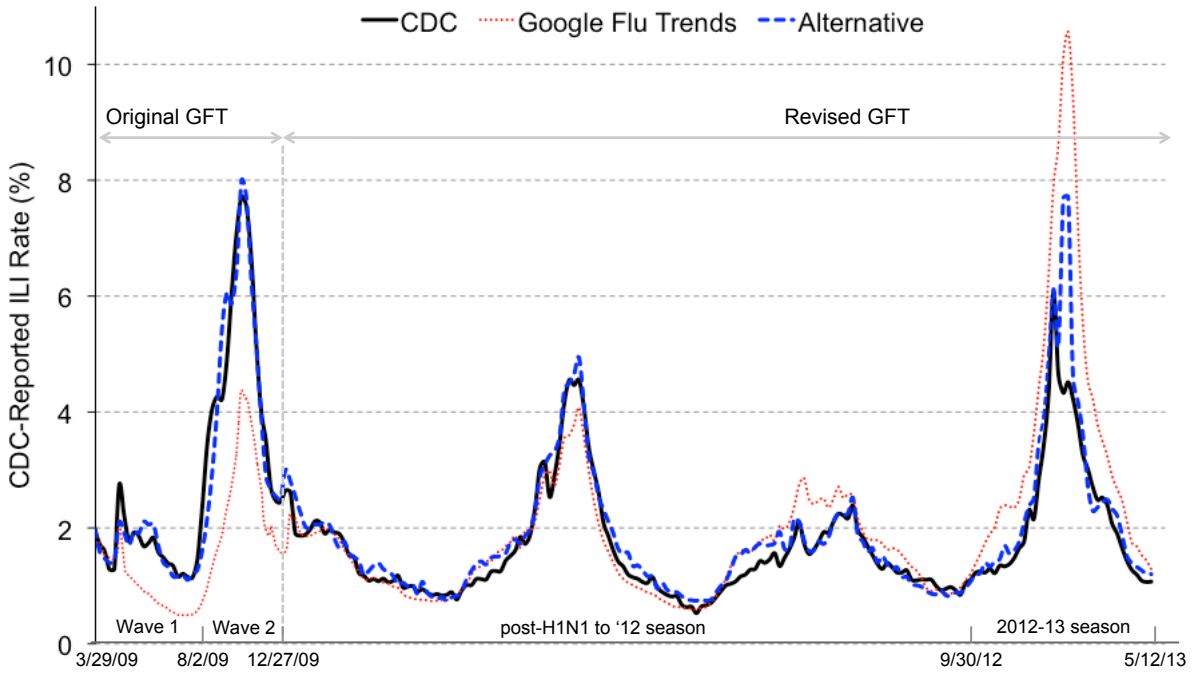

Figure 1. The alternative model outperforms Google Flu Trends
Both GFT (dot) and the proposed alternative model (dash) are shown against the criterion (solid) measure of national CDC-reported ILI (the weekly percent of confirmed influenza-like illness (ILI) related physician visits (*e.g.*, fever > 100*F and cough or sore throat as a percentage of all outpatient healthcare provider visits nationally to over 2,900 reporting clinics). Both GFT and the alternative model generated predictions for ILI at time (t) using search data from the same week (t) and historic periods, but because ILI is delayed, these estimates were typically available 1 to 2 weeks earlier than CDC-reported ILI.

During Wave 1 (March 29 through August 2, 2009) and Wave 2 (August 3 through December 27, 2009) of the H1N1 outbreak, particularly important periods of ILI surveillance, the RMSEs were 0.008 and 0.023 (i.e., if GFT predicted 0.061 ILI, it would be have a usual error of 0.008 or 0.023 each week) with correlations of r=0.290 and r=0.916 for GFT (**Table 1**). In contrast, our alternative model had RMSEs 0.002 and 0.006 with correlations of r=0.887 and r=0.961. In practical terms, the alternative model yielded more accurate predictions 17 out of 18 weeks during Wave 1 and every week (21 of 21) during Wave 2. For example, the original GFT predicted an estimated 4.3% peak ILI compared to a 7.7% CDC-reported ILI, an absolute difference of 3.4%; whereas our alternative model predicted an 8.0% peak, an absolute difference of 0.3%.

Our alternative predictions were often better than GFT during other influenza cycles. For Dec 28, 2009 to September 29, 2012, the alternative predictions had smaller error (RMSE=0.002, r=0.978 versus GFT RMSE=0.003, r=0.912), suggesting a relative 33% reduction in error (i.e.,

(0.003-0.002)/0.003). Again, in practical terms, the alternative model was more accurate 126 of 176 weeks (72%), with several periods where the alternative model produces predictions that mirror ILI when GFT is mis-predicting ILI by +/- 1%.

During the 2012/13 season the GFT's RMSE was 0.022 versus 0.009 for our alternative model, with correlations of r=0.927 and r=0.919. In practical terms, the alternative yielded more accurate predictions every single week. For example, GFT predicted a peak ILI of 10.6% versus 6.1% CDC-reported ILI, an absolute difference of 4.5%, compared to a 7.7% peak estimate from our alternative method, an absolute difference of 1.6%.

The autonomous and dynamic nature of our alternative appears to be a key component to improved predictions **(Figure 2)**. First, the model automatically excludes or reduces the predictive influence of many non-influenza terms. For example, "chaos tour" is typically given zero or little weight in the predictions without relying on human opinion as in prior revisions to GFT. Second, there is great variability in the coefficient estimates for each term by each week, where the alternative is automatically updating in response to how terms are indicative of ILI or not from the recent past. For example, terms indicative of influenza concern are given diminishing predictive value overtime and those indicative of treatment/complications are given more predictive value (i.e., "is flu contagious" versus "expectorant").

Most important our alternative does not guarantee accurate predictions, but when predictions do go astray, the alternative learns from its mistake in just 2 weeks, rather than waiting 2 years for a manual revision.[17] At the height of the 2012/13 season the alternative mis-predicted the peak ILI proportion by a nontrivial amount, but the model self-corrected changing the queries included/excluded just 2 weeks later (when the model was first aware of the error due to the delay in CDC-reported ILI). For example, in response to the model error "flu incubation" queries were discounted and used to predict lower ILI where "flu symptoms" were given more weight, used to predict higher ILI, with large changes in how dozens of other queries where included in the model as indicated by a clear shift in coefficient values in the latter portion of Figure 2.

Discussion

Our alternative methodology is capable of producing more accurate predictions of influenza activity than GFT, and does so autonomously with dynamic updating of the model each week. With 3-5 million infected and 250-500 thousand killed by influenza worldwide each year,[32] influenza surveillance is of tremendous importance, providing necessary intelligence for hospitals facing staffing decisions, physicians facing active and accurate diagnoses, employers with workers at risk for infection, and public health officials making recommendations for protecting unvaccinated individuals. Yet, these results have even greater implications as a case study for digital disease detection broadly.

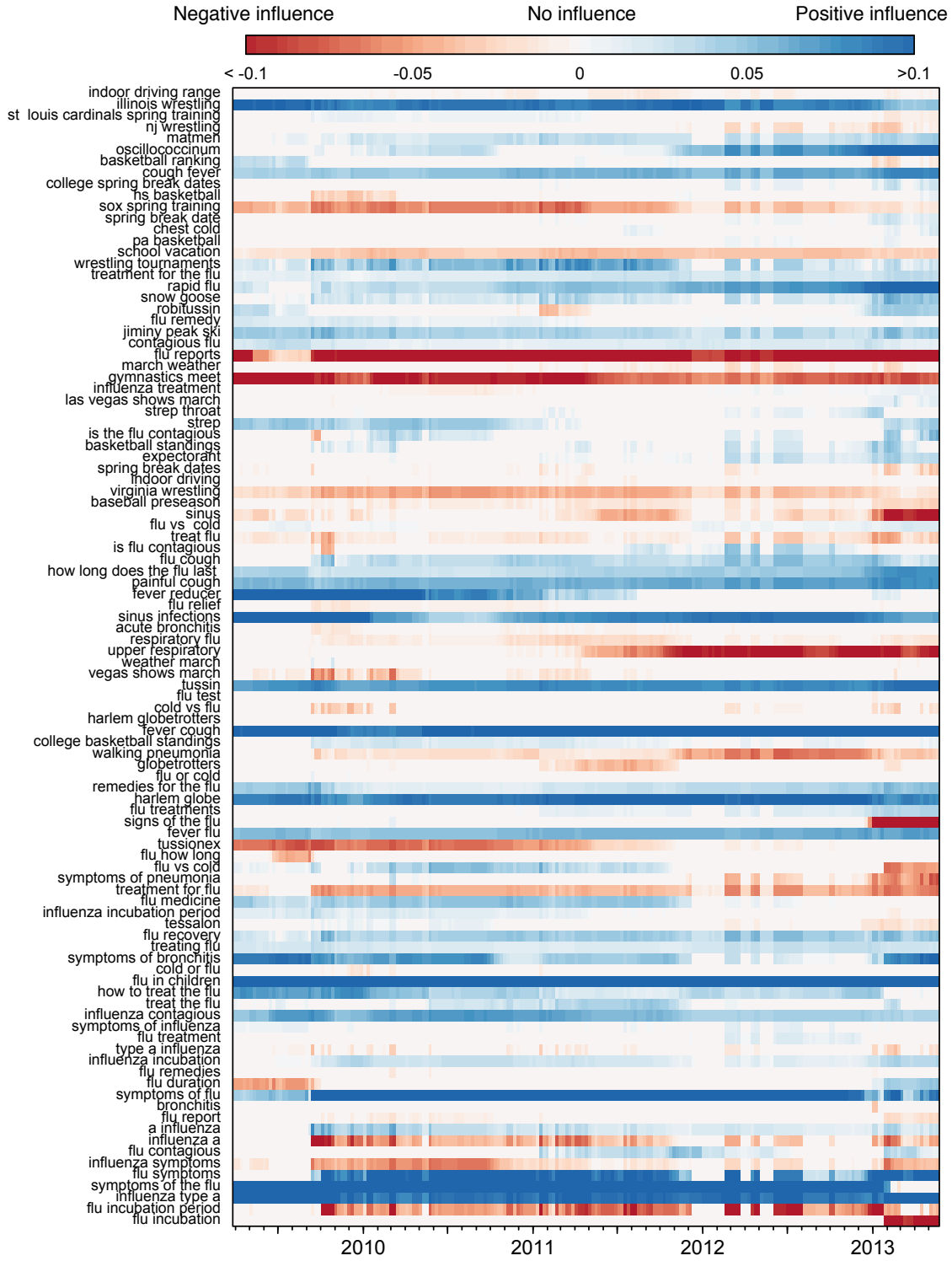

Figure 2. Search term inclusion and dynamic updating over time in the alternative model
Each line is an indicator of the estimated coefficient for a specific search term (y axis) updated each week (x axis). Colors indicate the relative importance of the search term (positive in blue, negative in red) on prediction over time. For example "flu in children" has a uniformly positive

prediction and "st louis cardinals spring training" has a nearly uniform zero effect. This figure shows how the alternative model is autonomously selecting queries and updating the relative value assigned to these queries each week.

Implications for the next Google Flu Trends

In a brief working paper, Google recently described the need to revise GFT.[21] In that they acknowledged that a multivariable approach (one of the improvements implemented here) would enhance the accuracy of GFT, but that paper also shared many of the weaknesses inherit in the original and first revision to GFT that our alternative may overcome. First, the methods lacked transparency as the working paper did not identify the model they were implementing.[21] Second, the predictive validity of the revised GFT remains unknown, since Google.org only included 5 weeks of data in the paper estimating the predictive accuracy. Using this small sample of data, however, our alternative method appears to be a better predictor. Last, their revision still relied on investigator opinion to select/omit some queries and failed to incorporate automatic updating, as we added herein. As a result, our alternative method may serve as the foundation for another revision to GFT. Specifically, because much of the alternative is automated, it can be scaled up (e.g., Google could apply it to thousands of strongly correlated search terms instead of just 100 as herein). Our study is just an initial step toward improving GFT, as the structure around our model can be further refined to yield even greater accuracy. Moreover, by making the inner workings of GFT and the data behind GFT more public, such improvements may be more quickly realized by other external teams.[19,20]

Implications for Digital Disease Detection

It is important to appraise how the leading system in the field, GFT, produced questionable predictions, while investigators mimicked the methods behind GFT[6,12,15,16] and few levied significant criticisms,[19,20,33,34] and how this omission can serve as a call for action within the field.

One European study found GFT predictions before and during the H1N1 pandemic only crudely associated with influenza (rho=0.39 and 0.52, respectively), but concluded GFT provided accurate detection.[35] Analyses from Australia,[36] China,[37] Japan,[38] New Zealand,[11] and the United States,[39] suggest GFT predicted larger incidences and missed the timing of outbreaks. Yet, these studies concluded GFT was valid, with one calling the accuracy of GFT "remarkable,"[36] as also quoted in a Google led publication.[17] This disparity in results and praise is indicative of a larger problem in digital disease detection.

Many of the current studies are the first of their kind and deserve praise as such, but to move this potentially very important field forward, investigators and public health leaders need to exercise caution and become discerning scientific consumers. Claims need to be carefully critiqued, as

nearly all of the literature on digital disease detection[4,6,9-14,40] relies on weak methodological approaches or patterns of association similar to GFT, with rare exception.[5,41-44] As a result, investigators should turn to more sophisticated approaches for the development and evaluation of digital systems. The stakes are high in public health surveillance and the methodological bar must be raised higher accordingly.

Yet, we remain optimistic about digital disease detection, because as with all adolescent fields there needs to be periodic methodological critique and revision. Building on this study and the work of others,[19,20] digital disease detection may better realize its aims.

	H1N1 (Wave 1) (3/29/09 - 8/2/09)	**H1N1 (Wave 2)** (8/2/09 - 12/27/09)	**post-H1N1 to '12 season** (12/27/09 - 9/30/12)	**2012-13 season** (9/30/12 - 5/12/13)
Correlation				
Alternative	**0.887**	**0.961**	**0.978**	**0.919**
GFT (Original)	0.290	0.916	-	-
GFT (Updated)	-	-	0.912	0.927
RMSE				
Alternative	**0.002**	**0.006**	**0.002**	**0.009**
GFT (Original)	0.008	0.023	-	-
GFT (Updated)	-	-	0.003	0.022

Table 1. Predictive Accuracy of the Alternative Model and Google Flu Trends
Both GFT and th alternative method generated predictions for ILI at time (t) using search data from the same week (t) and historic periods, but because ILI is delayed, these estimates were typically available 1 to 2 weeks earlier than CDC-reported ILI.

Limitations

Though advancing the underpinnings of GFT and digital disease detection, the alternative method is not without limitations. As usual, the alternative method was only evaluated at 1 geographic resolution, and model fits may vary in other geographies, especially geographies for which there is no ground-truth to compare the model predictions against. In addition, CDC-reported ILI estimates (the ground truth) sometimes are later revised as the flu season progresses,[45] thus introducing uncertainties in the (or any) prediction methodology. In this regard, the dynamic alternative approach is capable of incorporating information as it becomes available, and thus, will automatically produce an improved model every time CDC ILI information is either initially released or revised. Last, Google queries are but one source of information and a multi-sourced model would be preferred, but through appropriate modeling, searches can be leveraged to achieve accurate predictions that better realize the implications of GFT.[8]

Conclusion

The methods behind digital disease detection are wanting, but by outlining a correction that could improve the accuracy of influenza detection we hope public health officials find value in digital disease detection and investigators refine our approach to achieve more transparent and accurate surveillance that can have real-world public health impacts.

References

1. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med. 2009;360:2153-5, 2157.
2. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med. 2011;40:S154-S158.
3. Ayers JW, Althouse BM, Dredze M. Could Behavioral Medicine Lead the Web Data Revolution? JAMA. 2014;Early Release Online.
4. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak. Am J Trop Med Hyg. 2012;86:39-45.
5. Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis. 2011;5:e1258.
6. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis. 2011;5:e1206.
7. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc. 2006:244-248.
8. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457:1012-1014.

9. Breyer BN, Sen S, Aaronson DS, Stoller ML, Erickson BA, Eisenberg ML. Use of Google Insights for Search to track seasonal and geographic kidney stone incidence in the United States. Urology. 2011;78:267-271.
10. Willard SD, Nguyen MM. Internet Search Trends Analysis Tools Can Provide Real-time Data on Kidney Stone Disease in the United States. Urology. 2011;81:37-42.
11. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. CMAJ. 2009;180:829-831.
12. Dukic VM, David MZ, Lauderdale DS. Internet Queries and Methicillin-Resistant Staphylococcus aureus Surveillance. Emerg Infect Dis. 2011;17:1068-1070.
13. Yang AC, Tsai SJ, Huang NE, Peng CK. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord. 2011;132:179-184.
14. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, et al. Monitoring of non-cigarette tobacco use using Google Trends. Tob Control. 2014.
15. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One. 2013;8:e64323.
16. Ocampo AJ, Chunara R, Brownstein JS. Using search queries for malaria surveillance, Thailand. Malar J. 2013;12:390.
17. 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:e23610.
18. Butler D. When Google got flu wrong. Nature News. 2013;http://www.nature.com/news/when-google-got-flu-wrong.
19. 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:e1003256.
20. Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. Science. 2014;In Press.
21. Copeland P, Romano R, Zhang T, Hecht G, Zigmond D, Stefansen C. Google Disease Trends: An Update. Google.org. ;http://patrickcopeland.org/papers/isntd.pdf.
22. Nsoesie EO, Buckeridge DL, Brownstein JS. Guess who's not coming to dinner? Evaluating online restaurant reservations for disease surveillance. J Med Internet Res. 2014;16:e22.
23. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med. 2011;40:448-453.
24. Ayers JW, Althouse BM, Allem JP, et al. Novel surveillance of psychological distress during the great recession. J Affect Disord. 2012;142:323-330.
25. Ayers JW, Althouse BM, Noar SM, Cohen JE. Do celebrity cancer diagnoses promote primary cancer prevention? Prev Med. 2014;58:81-84.
26. Ayers JW, Althouse BM, Johnson M, Cohen JE. Circaseptan (weekly) rhythms in smoking cessation considerations. JAMA Intern Med. 2014;174:146-148.

27. Ghil M, Malanotte-Rizzoli P. Data assimilation in meteorology and oceanography. Adv. Geophys. 1991;33:141-266.
28. Wang B, Zou X, Zhu J. Data assimilation and its applications. Proceedings of the National Academy of Sciences. 2000;97:11143-11144.
29. Hastie T, Tibshirani R, Friedman J, Franklin J. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer. 2005;27:83-85.
30. Russell SJ, Norvig P, Canny JF, Malik JM, Edwards DD. Artificial intelligence: a modern approach. Prentice hall Englewood Cliffs; 1995.
31. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996:267-288.
32. WHO. World Health Organization. Influenza fact sheet. . www.who.int/mediacentre/ factsheets/fs211/en/index.html. 2009.
33. Patwardhan A, Bilkovski R. Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. PLoS One. 2012;7:e43611.
34. Liu F, Lv B, Peng G, Li X. Influenza Epidemics Detection Based on Google Search Queries. Recent Progress in Data Engineering and Internet Technology. 2012:371-376.
35. Valdivia A, Monge-Corella S. Diseases tracked by using Google trends, Spain. Emerg Infect Dis. 2010;16:168.
36. Kelly H, Grant K. Interim analysis of pandemic influenza (H1N1) 2009 in Australia: surveillance trends, age of infection and effectiveness of seasonal vaccination. Euro Surveill. 2009;14.
37. Kang M, Zhong H, He J, Rutherford S, Yang F. Using google trends for influenza surveillance in South china. PLoS One. 2013;8:e55205.
38. Wada K, Ohta H, Aizawa Y. Correlation of "Google Flu Trends" with Sentinel Surveillance Data for Influenza in 2009 in Japan. Open Public Health Journal. 2011;4:17-20.
39. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. PLoS One. 2011;6:e18687.
40. Dugas AF, Hsieh YH, Levin SR, et al. Google Flu Trends: Correlation With Emergency Department Influenza Rates and Crowding Metrics. Clin Infect Dis. 2012;54:463-469.
41. Paul MJ, Dredze M. You are what you tweet: Analyzing Twitter for public health. Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011). Barcelona, Spain, 2011.
42. Althouse BM, Allem J-P, Childers MA, Dredze M, Ayers JW. Population Health Concerns During the United States' Great Recession. Am J Prev Med. 2014;46:166-170.
43. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on google. Am J Prev Med. 2013;44:520-525.

44. Ayers JW, Althouse BM, Ribisl KM, Emery S. Digital Detection for Tobacco Control: Online Reactions to the United States' 2009 Cigarette Excise Tax Increase. Nicotine Tob Res. 2013.
45. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. Proc Natl Acad Sci U S A. 2012;109:20425-20430.