

1-1-2016

Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits.

Joseph Jeffrey Klembczyk

Mehdi Jalalpour

Scott Levin

Raynard E Washington

Jesse M. Pines
George Washington University

See next page for additional authors

Follow this and additional works at: https://hsrc.himmelfarb.gwu.edu/smhs_emerg_facpubs

 Part of the [Clinical Epidemiology Commons](#), [Emergency Medicine Commons](#), [Influenza Humans Commons](#), and the [Trauma Commons](#)

APA Citation

Klembczyk, J., Jalalpour, M., Levin, S., Washington, R., Pines, J. M., Rothman, R., & Dugas, A. (2016). Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. *Journal of Medical Internet Research*, 18 (6). <http://dx.doi.org/10.2196/jmir.5585>

This Journal Article is brought to you for free and open access by the Emergency Medicine at Health Sciences Research Commons. It has been accepted for inclusion in Emergency Medicine Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Joseph Jeffrey Klembczyk, Mehdi Jalalpour, Scott Levin, Raynard E Washington, Jesse M. Pines, Richard E Rothman, and Andrea Freyer Dugas

Original Paper

Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits

Joseph Jeffrey Klembczyk¹, MPH, MD; Mehdi Jalalpour², PhD; Scott Levin³, PhD; Raynard E Washington⁴, PhD; Jesse M Pines⁵, MD, MBA, MSCE; Richard E Rothman³, MD, PhD; Andrea Freyer Dugas³, MD, PhD

¹Johns Hopkins University, School of Medicine, Hampstead, NC, United States

²Cleveland State University, Department of Civil and Environmental Engineering, Cleveland, OH, United States

³Johns Hopkins University, Department of Emergency Medicine, Baltimore, MD, United States

⁴Agency for Healthcare Research and Quality, Rockville, MD, United States

⁵Departments of Emergency Medicine and Health Policy, George Washington University, Washington, DC, United States

Corresponding Author:

Joseph Jeffrey Klembczyk, MPH, MD

Johns Hopkins University

School of Medicine

128 S Belvedere Dr

Hampstead, NC, 28443

United States

Phone: 1 518 573 2045

Fax: 1 910 937 1802

Email: jjklem@gmail.com

Abstract

Background: Influenza is a deadly and costly public health problem. Variations in its seasonal patterns cause dangerous surges in emergency department (ED) patient volume. Google Flu Trends (GFT) can provide faster influenza surveillance information than traditional CDC methods, potentially leading to improved public health preparedness. GFT has been found to correlate well with reported influenza and to improve influenza prediction models. However, previous validation studies have focused on isolated clinical locations.

Objective: The purpose of the study was to measure GFT surveillance effectiveness by correlating GFT with influenza-related ED visits in 19 US cities across seven influenza seasons, and to explore which city characteristics lead to better or worse GFT effectiveness.

Methods: Using Healthcare Cost and Utilization Project data, we collected weekly counts of ED visits for all patients with diagnosis (International Statistical Classification of Diseases 9) codes for influenza-related visits from 2005-2011 in 19 different US cities. We measured the correlation between weekly volume of GFT searches and influenza-related ED visits (ie, GFT ED surveillance effectiveness) per city. We evaluated the relationship between 15 publically available city indicators (11 sociodemographic, two health care utilization, and two climate) and GFT surveillance effectiveness using univariate linear regression.

Results: Correlation between city-level GFT and influenza-related ED visits had a median of .84, ranging from .67 to .93 across 19 cities. Temporal variability was observed, with median correlation ranging from .78 in 2009 to .94 in 2005. City indicators significantly associated ($P < .10$) with improved GFT surveillance include higher proportion of female population, higher proportion with Medicare coverage, higher ED visits per capita, and lower socioeconomic status.

Conclusions: GFT is strongly correlated with ED influenza-related visits at the city level, but unexplained variation over geographic location and time limits its utility as standalone surveillance. GFT is likely most useful as an early signal used in conjunction with other more comprehensive surveillance techniques. City indicators associated with improved GFT surveillance provide some insight into the variability of GFT effectiveness. For example, populations with lower socioeconomic status may have a greater tendency to initially turn to the Internet for health questions, thus leading to increased GFT effectiveness. GFT has the potential to provide valuable information to ED providers for patient care and to administrators for ED surge preparedness.

(*J Med Internet Res* 2016;18(6):e175) doi:[10.2196/jmir.5585](https://doi.org/10.2196/jmir.5585)

KEYWORDS

influenza; surveillance; emergency department; google flu trends; infoveillance

Introduction

Background

Influenza accounts for up to 294,000 hospitalizations and 30,000 deaths per year in the United States and costs an estimated US \$12 billion annually [1-3]. Seasonal influenza patterns result in sudden increases in emergency department (ED) volume, further straining an already stressed health care safety net [4-8]. Increased influenza patient volume exacerbates ED crowding, which is linked to delays in critical treatments and increased morbidity and mortality [9-12]. Beyond seasonal influenza, the potential for a pandemic influenza outbreak is a well-recognized and serious threat to the US health care infrastructure [5,8]. Therefore, accurate and timely influenza surveillance is critical for diagnosis and treatment, as well as public health and hospital preparedness.

The Centers for Disease Control and Prevention (CDC) publicly releases weekly influenza surveillance information aggregated from diagnostic laboratories, reports from outpatient providers, and mortality and hospitalization data [4]. Although widely relied upon, the CDC surveillance information is released with a 1-2 week delay [9]. In order to provide a more timely estimate of influenza activity, Google developed Google Flu Trends (GFT), an algorithm assessing billions of Internet search queries from Google users at various geographic levels. GFT was trained with CDC regional data to estimate the proportion of outpatient visits that were related to influenza-like illness (ILI) [13]. GFT time series data can be obtained down to the city level for 122 large metropolitan areas in the United States. Although the exact algorithm calculating these estimates is proprietary, this geographically focused, publicly available data is a potential source for timely surveillance information [13].

Prior Work

Since the original validation of GFT in 2008, numerous independent evaluations have shown variable results [13]. Many studies of GFT have shown close correlation between GFT and either ILI or confirmed influenza cases in broad geographic areas and individual cities [14-16]. GFT has also been successfully included in numerous influenza forecasting models,

at both the local and the national level [17-20]. Others have identified challenges for GFT estimates. Specifically, the H1N1 pandemic in 2009 was predicted late and underestimated by GFT. This was attributed to its unusual timing and altered Internet search habits following increased media coverage of the pandemic [21,22]. Consequently, GFT's algorithm was updated to include more direct influenza-related terms rather than complications of the disease [21,22]. Even with the updated algorithm, GFT underestimated the moderately severe 2012-2013 influenza season [21]. The GFT algorithm was subsequently updated twice more in 2013 and 2014. The value of GFT and the settings in which it is most effective are not well understood.

Objective

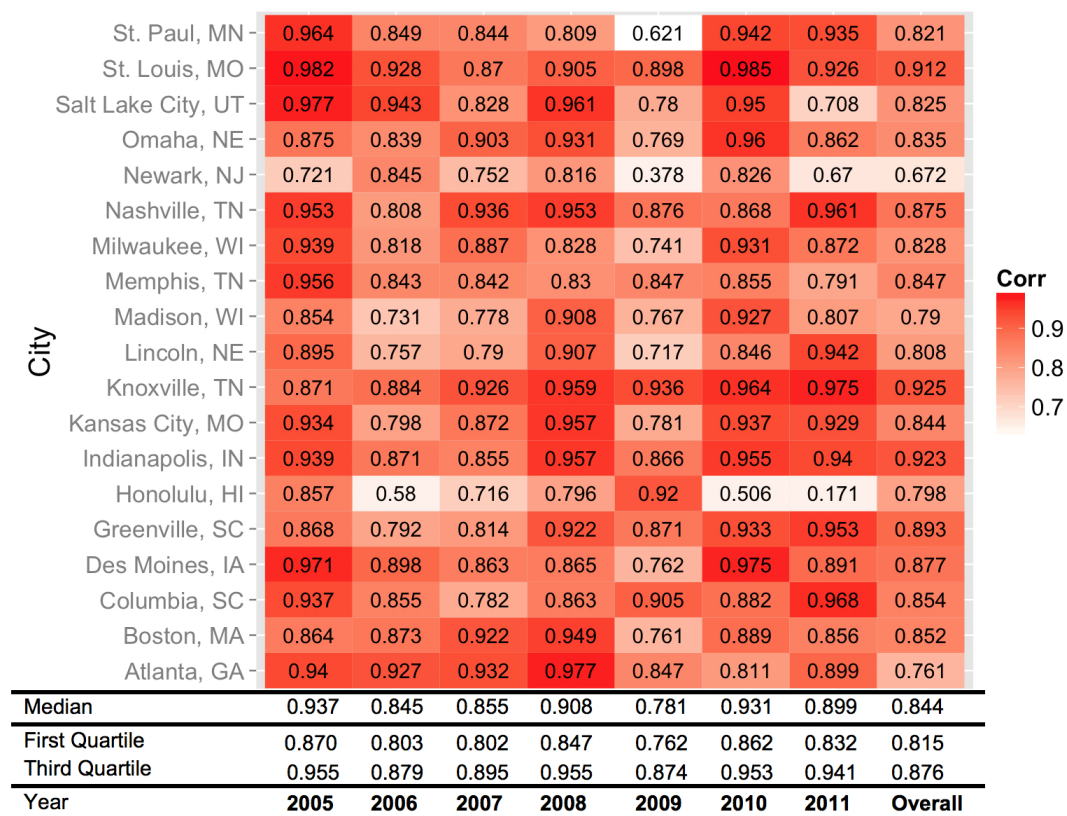
Although there have been promising single center validations, expansion to broader geographic locations is required to fully evaluate the potential role for this alternate or complementary source of influenza surveillance. This study is the first to examine the effectiveness of GFT simultaneously in several geographically distinct regions throughout the United States. Additionally, we explored the correlation of several sociodemographic factors with GFT effectiveness. This was completed to determine the factors associated with GFT effectiveness and increase our understanding of the tool. We hypothesize that GFT will be validated as a geographically robust early predictive signal for ED influenza.

Methods

Study Population and Setting

This study, in collaboration with the Agency for Healthcare Research and Quality (AHRQ), used data from the (HCUP) State Emergency Department Databases (SEDD) to estimate influenza-related ED visits in 19 US cities from 2005-2011. The SEDD are a set of databases that include nearly all ED visits from non-rehabilitation community hospitals in participating states [23]. The 19 cities were selected based on availability of both HCUP and GFT city-level data. The cities evaluated are listed in [Figure 1](#).

Figure 1. Correlation coefficients between Google Flu Trends and influenza-related emergency department visits.



Data Collection

We obtained the weekly number of ED visits for influenza-related illnesses among selected cities from the HCUP databases for January 1, 2005, through December 31, 2011 [23]. This contained all ED visits to community hospitals located within the designated city area: both visits that resulted in a treatment and discharge, as well as visits that resulted in hospital admission. We defined influenza-related illness using International Statistical Classification of Diseases and Related Health Problems (ICD-9-M) codes representing diagnoses related to pneumonia or influenza (480-487, 488.1), as described by Rubison et al [24]. The addition of select pneumonia diagnoses has been validated for accurately characterizing influenza [24]. The date of the ED visit was used to create weekly totals of ED encounters for influenza-related visits for each city. Because this de-identified data was collected for another purpose, this research was exempt from the Institutional Review Board.

City-level GFT data were downloaded from the Google Flu Trends website in June 2014 for each of the 19 cities and corresponded to the 2009 update of GFT. Output consists of a local weekly parameter estimating the proportion of outpatient visits for ILI [25].

A total of 15 city indicators hypothesized to explain GFT efficacy was collected. They comprised 11 sociodemographic, two health care utilization, and two climate city-based characteristics. These measurements were most often available annually or occasionally through less-frequent surveys. The most appropriate available discrete measurement or average over our study period of the indicator was used for analysis.

The 15 sociodemographic characteristics collected for each city from the US Census Bureau (2010) included the following:

1. Population density.
2. Proportion of the population female.
3. Proportion of the population <18 years of age.
4. Proportion of the population ≥65 years of age.
5. Proportion of the population Caucasian.
6. Proportion of the population African/American.
7. Proportion of the population Hispanic/Latino [26].
8. Proportion of the population uninsured, which was collected from the 2008 Small Area Health Insurance Estimates project [27].
9. Proportion of the population with Medicare in 2008, which was collected from the Centers for Medicare and Medicaid [28].
10. Availability of Internet services (relevant to Google searches) for each city, measured by the number of Internet connections per household, was collected from the Federal Communications Commission (FCC). However, the data were binned into groups of 200 such that only categories of 0-200, 200-400, etc, per 1000 households in a given county were available [29]. Thus, we used the midpoint of each bin to provide a household-weighted average among counties for each city.
11. A collective measure of socioeconomic status (SES) was created by combining four separate indicators: household median income (US Census Bureau 2010), proportion with high school degree (American Community Survey 2007-2009), proportion

with college degree (American Community Survey 2007-2009), and proportion employed (Bureau of Labor Statistics 2008, collected by county and population-weighted) [26,30,31]. These individual indicators were highly correlated and thus considered proxies for socioeconomic status. The four indicators were normalized along the 19 observations to produce the SES variable with a mean of zero (SD 3.15).

12. Medicaid-reimbursed hospital inpatient days per 1000 person-years, collected from the American Hospital Association [32].

13. Total ED visits per person-year, retrieved for 2011 from HCUP [23]. Because no significant time variation in total ED visits was observed, the temporal average was used for each city. These two health care utilization measures were available only by county, so population-weighted averages of the counties composing each city were calculated.

14. City climate conditions were included. Air pollution (particulate matter 2.5) was collected from the CDC for 2008 at the county level and was also population-weighted [33].

15. Seasonality of climate was estimated using daily historical temperature readings for each city collected from Weather Underground [34]. Average monthly temperatures along the entire time series were calculated, and the standard deviation of these averages was taken as a measure of seasonality of temperature as described by Legates and Willmott [35].

Google Flu Trends Effectiveness

Pearson correlation coefficients between GFT and ED visits for pneumonia and influenza for each city were calculated both for individual seasons and the entire time series. Each season included data from August 1 of the prior year to July 31 of the stated year with the exception of 2005, which began at January 1 and ended July 31 due to data availability. For example, the

2006 season includes data from August 1, 2005, to July 31, 2006.

We used two separate methods to identify potential outliers with respect to GFT effectiveness. First, we used the traditional box and whisker method in which cities with a correlation coefficient the distance of 1.5 times the interquartile range (IQR) outside of the IQR were considered outliers. We also applied the median absolute deviation method of outlier identification [36,37].

City Indicators

Univariate linear regression was performed along the 19 cities using each of the 15 sociodemographic variables as independent variables and the correlations between GFT and ED visits for pneumonia and influenza as the dependent variable. Trend lines were displayed only for those sociodemographic factors for which regression yielded $P < .10$.

Results

Google Flu Trends Effectiveness

Overall, GFT is highly correlated with ED visits for pneumonia and influenza, with a median correlation of .844 (range .672-.925) across the 19 cities included in this analysis (Figure 1). However, there is temporal variability (Figure 2), with yearly median correlations ranging from .781 during the 2009 H1N1 pandemic to .937 in 2005. There is additional geographic variability, as shown in Figure 3, with a trend of higher correlations between GFT and ED visits for influenza-related visits in the midwest and southeast regions including Des Moines, IA; St. Louis, MO; Indianapolis, IN; Nashville, TN; Knoxville, TN; and Greenville, SC. Figure 4 displays a time series comparison of GFT and weekly influenza-related visits for the three cities with the lowest, median, and highest correlation.

Figure 2. Correlation between Google Flu Trends and influenza-related emergency department visits for individual cities, by year (outliers are marked by red +, including Honolulu, HI [Hnl] and Newark, NJ [Nrk]).

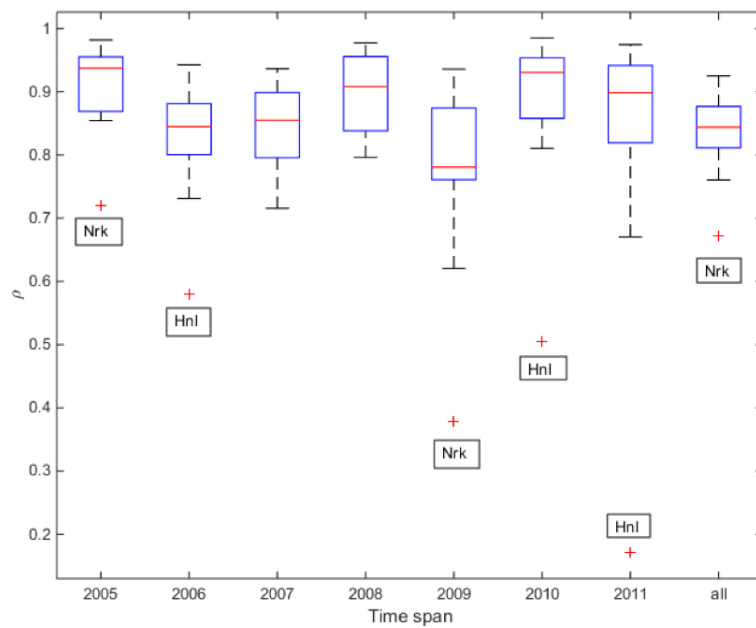


Figure 3. Correlation coefficients between Google Flu Trends and influenza-related emergency department visits for individual cities over the total time series (2005-2011). Correlations range from .672 (yellow) to .925 (red).

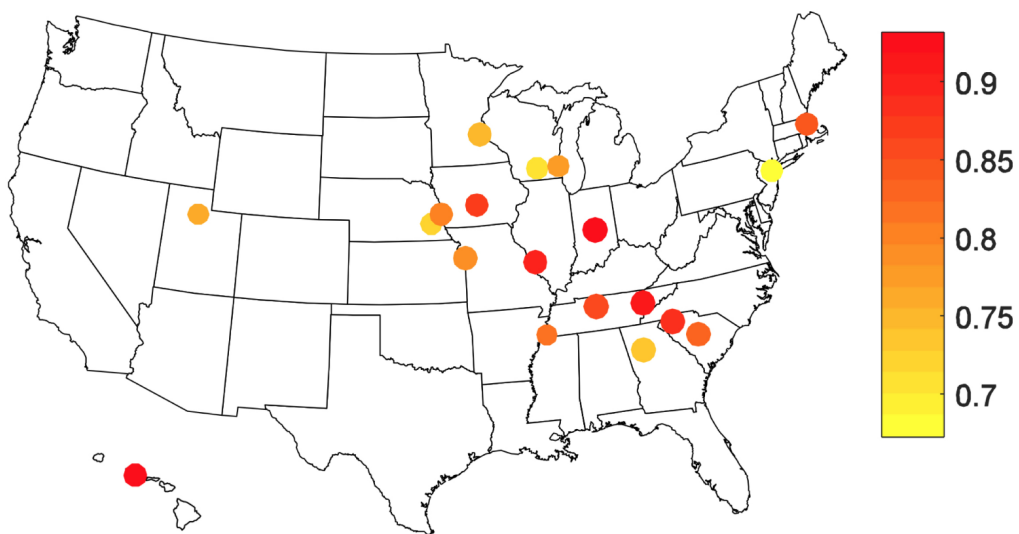
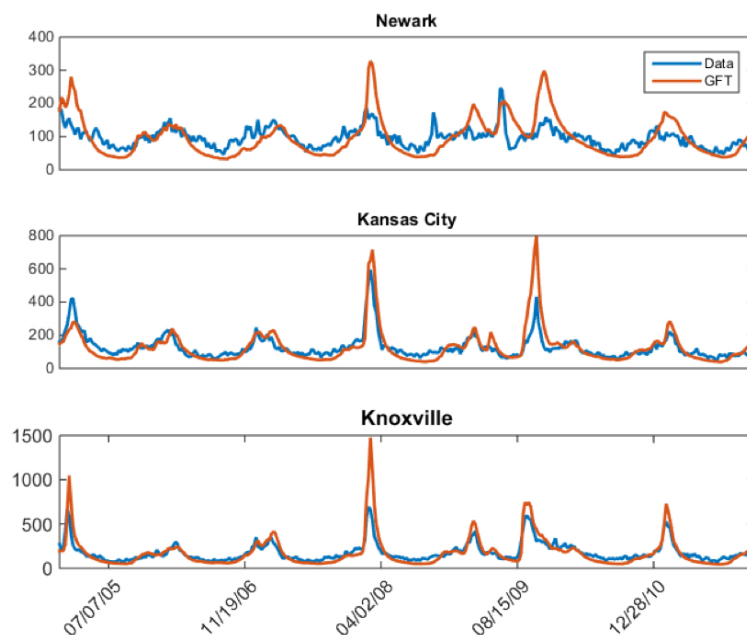


Figure 4. Time series comparing Google Flu Trends and influenza-related emergency department visits for individual cities over the total time series (2005-2011) demonstrating the lowest (Newark, NJ $P=.672$), median (Kansas City, MO $P=.844$), and highest (Knoxville, TN $P=.925$) correlation coefficients.

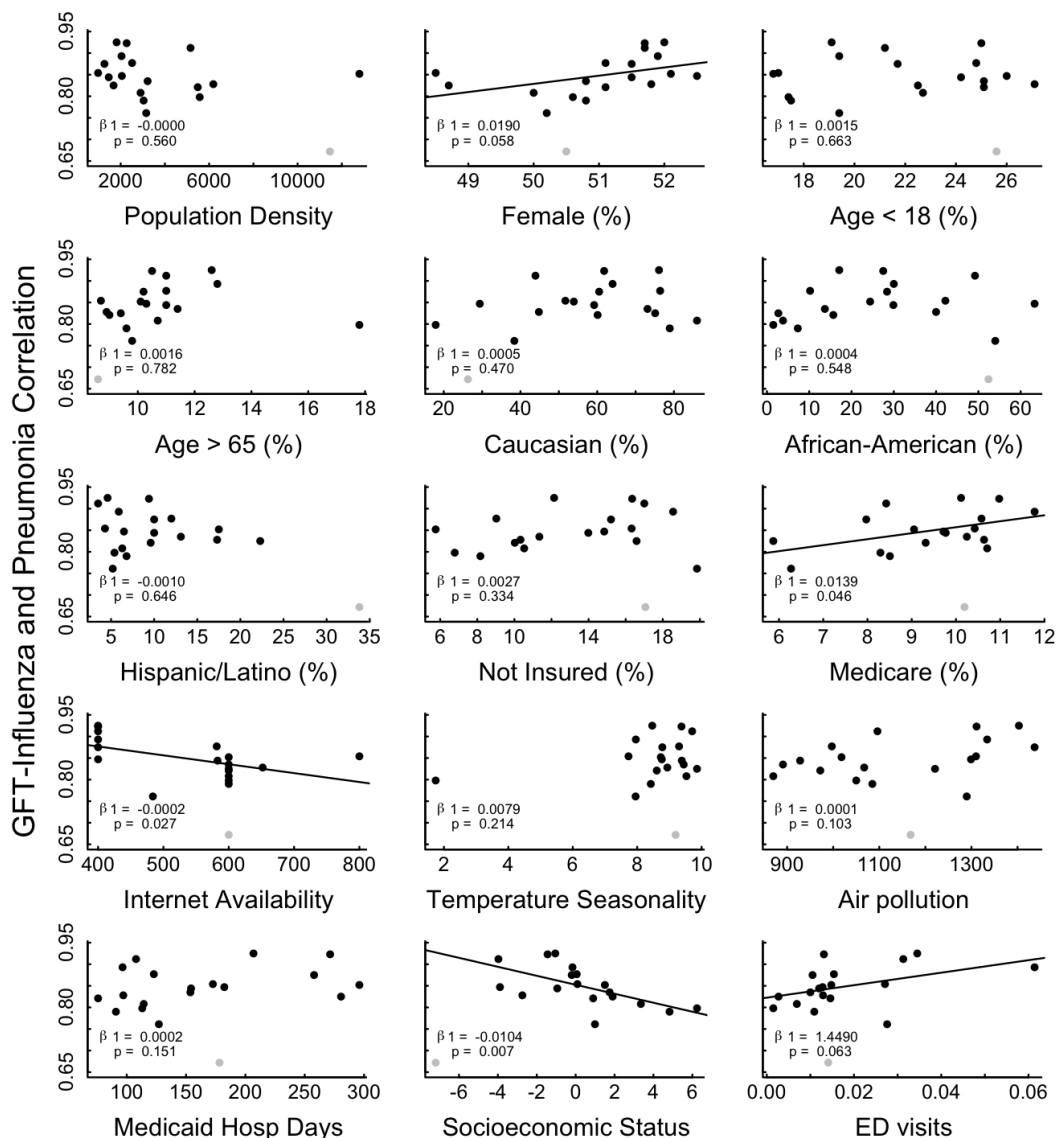


City Indicators

Newark, NJ, was found to be an outlier with respect to GFT effectiveness. This was based on consensus of the two independent outlier-identification techniques as well as expert opinion. We believe it carried undue influence in our analysis of city-based indicators and therefore removed Newark from these calculations. Honolulu appears outside the IQR for the distribution of cities in Figure 2 in three different years, but it was not quantitatively identified as an outlier over the whole time series and thus was included in the analysis.

Fifteen sociodemographic indicators, collected for each of the 19 cities, were evaluated for their potential association with the correlation between GFT and ED visits for pneumonia and influenza (Figure 5). Of the indicators evaluated, Internet availability and socioeconomic status were negatively correlated with GFT effectiveness (decrease in these variables was associated with an increase in GFT effectiveness). Proportion of the population that is female, proportion with Medicare insurance, and number of ED visits per person were positively correlated with GFT effectiveness.

Figure 5. Correlation between Google Flu Trends and emergency department visits for pneumonia and influenza for individual cities over the total time series (2005-2011) plotted against 15 different city-level indicators. Trend lines are plotted for variables with a *P* value of less than .10. Univariate regression coefficients and *P* values are displayed for each indicator. The outlier city (Newark) was not included in the analysis but is still displayed in gray.



Discussion

Principal Results and Prior Work

Although GFT is a promising new source of real-time influenza surveillance, there is conflicting evidence regarding its accuracy. Previous studies have validated GFT at the national level or in a specific local setting, but this is the first to evaluate GFT across multiple cities simultaneously with local clinical outcomes [13-16,21-22]. We sought to more fully understand the geographical and temporal correlations between GFT and influenza-related ED visits by evaluating 19 different US cities over 7 influenza seasons.

Overall, we found GFT to be a valuable tool that provides useful surveillance in a variety of settings. However, there remains some geographic and temporal variability. Cities in the Southeast and Midwest appeared to have stronger correlations between GFT and influenza-related ED visits compared to cities in other

regions. Similar to our results, temporal variability in GFT effectiveness has been observed in past studies [15-20,22]. This may be due to a combination of outbreak timing, outbreak severity, media coverage, public health awareness, and other unpredictable sources of variability. GFT has been updated in the past in an attempt to reduce some of these problems [21]. Characterization and minimization of this temporal variability is critical when incorporating GFT into influenza surveillance systems.

We further explored the geographic findings by evaluating characteristics of individual cities that may impact the relationship between GFT and influenza-related ED visits. The only basic city demographic variable that correlated with effective GFT was proportion of the population that is female. Per-capita health care use tends to be higher among females, which may explain this trend [38,39]. More notable was that several factors including age and ethnicity did not correlate with GFT. We hypothesized that proportionally older populations

(age >65 years) may be less likely to access health information on the Internet; however, the proportion of those populations in a city did not impact GFT effectiveness. Additionally, there was no change in GFT surveillance effectiveness in cities with a large Hispanic or Latino population despite the hypothesis that primary language differences may limit search queries counted by the GFT algorithm, which uses English search terms only.

Other indicators are more difficult to interpret. Internet connections per household was associated with decreased GFT effectiveness, while we hypothesized that greater connectivity would lead to more predictive GFT. Internet availability was only available as data binned into 5 levels from the FCC. The granularity in measurement of this variable may have limited its utility in accurately distinguishing differences between the cities. Furthermore, our hypothesis would be best tested by a measure of Internet use, rather than availability, but a consistent indicator of use was not readily available.

Lower SES was associated with more effective GFT. This may be because lower SES populations may disproportionately use the ED for non-urgent conditions (eg, ILI) due to limited access to other community health services such as primary care [40-44]. This SES effect is likely more than a reflection of the health insurance status of the populations, as the correlation with proportion uninsured was insignificant. Further, lower SES populations may be more likely to consult the Internet for health care questions, resulting in more accurate GFT predictions.

In evaluating the correlation between health utilization with GFT effectiveness, both the proportion of the population insured by Medicare and the per capita number of ED visits had a positive correlation with GFT effectiveness. Given that we are evaluating GFT effectiveness through correlation with ED visits, it is expected that cities more dependent upon ED care would have stronger correlation between a marker of influenza and ED visits for potential influenza. Therefore, we would expect GFT to be most useful as an ED and hospital surveillance tool in populations with lower SES and higher ED utilization.

Markers of local climate, such as air pollution or seasonality of climate, did not correlate with the effectiveness of GFT. Several influenza forecast models have included temperature to predict severity of influenza [45,46]. Cities with increased variation in temperature by season may have more severe and predictable influenza. However, the insignificance of the climate variable, and the determination that warmer cities in the southeast United States had increased GFT effectiveness, fail to support this hypothesis.

Similarly, we hypothesized that cities with increased air pollution might have poorer baseline lung function and thus more severe influenza pathology. This would cause heightened influenza awareness and diagnostic rates, leading to improved GFT effectiveness. However, this effect was also not supported by our data. Our analysis suggests that GFT effectiveness may not be driven by severity of disease.

Our results support the conclusion that traditional surveillance models can benefit from the addition of Internet search query

data. However, temporal and geographic variability exists, which should be considered when generalizing results from a single influenza season or single hospital or region. This study specifically demonstrates the magnitude of variability that may be expected across different cities in the United States. Further, our results suggest that a population-based measure of SES may be useful to understand and modulate confidence in GFT effectiveness. Regardless, before incorporating GFT or other Internet query-based data into local public health surveillance systems, it is important to account for GFT performance in that specific location.

Limitations

Limitations of our study include a small sample size of 19 cities, which may have hindered our ability to detect trends in city characteristics. The sample size also constrained us from carrying out multivariate regression analyses. Additionally, historical GFT data were available only in weekly intervals, limiting the temporal resolution of our analyses. As previously mentioned, Internet access and usage was difficult to quantify. Health care access and utilization was also difficult to capture at the local level, and more available variables in this category may have yielded further insight. Additionally, Newark was excluded as an outlier from the sociodemographic factors analysis. While we justified the decision to remove Newark, it did affect the significance of some trends: proportion Hispanic/Latino became insignificant, while SES became significant. The sensitivity of our results is a function of both the small sample size as well as the extreme values for Newark in GFT effectiveness and some of the city-level indicators. Next, while we validated GFT's correlation with influenza-related ED visits, GFT is more broadly designed to correlate with outpatient ILI visits. Therefore, our inferences of the factors driving GFT effectiveness may not be generalizable to GFT as used in settings outside of the emergency department. Moreover, the study used ED visits data up to 2011 and the corresponding 2009 GFT model of the era, which limits generalizing the conclusions about GFT to recent trends. Finally, while GFT access is currently limited by Google to only research institutions, our results are still relevant to future iterations of GFT and other Internet search query-based surveillance tools.

Conclusions

As a whole, our results indicate that GFT is a sensitive surveillance tool that can add value to our current surveillance systems. Because of its spatio-temporal variability in effectiveness, GFT is likely most useful as an additional, early signal to influenza prediction models, rather than as a stand-alone approach. Furthermore, our results help explain where GFT may be most effective, specifically in higher percent female populations with lower socioeconomic status and high ED use. This can help inform the most useful settings for further GFT study and implementation. Effective, real-time influenza surveillance is useful both for emergency medicine providers on a patient-to-patient basis and for ED crowding preparedness. Characterizing geographic effectiveness and variability of GFT and Internet search query data is crucial for the continued progress of influenza surveillance.

Acknowledgments

The authors would like to thank the 14 state data organizations that contributed data to HCUP used in this study: Georgia Hospital Association, Hawaii Health Information Corporation, Iowa Hospital Association, Indiana Hospital Association, Massachusetts Center for Health Information and Analysis, Missouri Hospital Industry Data Institute, New Jersey Department of Health, South Carolina Revenue and Fiscal Affairs Office, Missouri Hospital Industry Data Institute, Minnesota Hospital Association, Tennessee Hospital Association, Wisconsin Department of Health Services, Utah Department of Health, and Nevada Department of Health and Human Services. The authors would also like to acknowledge Ryan Mutter, PhD, for his contributions while affiliated with AHRQ to the development of this study and Social and Scientific Systems, Inc. for data support.

This work was supported by the Cooperative Agreement IDSEP130014-01-00 from The Assistant Secretary for Preparedness and Response within the US Department of Health and Human Services. Contents do not necessarily represent the official views of the US Department of Health and Human Services. Use of trade names and commercial sources is for identification only and does not imply endorsement by the US Department of Health and Human Services.

This publication was made possible by the Johns Hopkins Institute for Clinical and Translational Research (ICTR), which is funded in part by Grant Number TL1 TR001078 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the Johns Hopkins ICTR, NCATS, or NIH.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Seasonal influenza. 2011. URL: <http://www.cdc.gov/flu/about/qa/disease.htm> [accessed 2016-01-28] [WebCite Cache ID 6erU01w1f]
2. Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, et al. Influenza-associated hospitalizations in the United States. *JAMA* 2004 Sep 15;292(11):1333-1340. [doi: [10.1001/jama.292.11.1333](https://doi.org/10.1001/jama.292.11.1333)] [Medline: [15367555](https://pubmed.ncbi.nlm.nih.gov/15367555/)]
3. Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* 2003 Jan 8;289(2):179-186. [Medline: [12517228](https://pubmed.ncbi.nlm.nih.gov/12517228/)]
4. McDonnell WM, Nelson DS, Schunk JE. Should we fear “flu fear” itself? Effects of H1N1 influenza fear on ED use. *Am J Emerg Med* 2012 Feb;30(2):275-282. [doi: [10.1016/j.ajem.2010.11.027](https://doi.org/10.1016/j.ajem.2010.11.027)] [Medline: [21208765](https://pubmed.ncbi.nlm.nih.gov/21208765/)]
5. Institute of Medicine Committee on the Future of Emergency Care in the U.S. Health System. The future of emergency care in the United States health system. *Ann Emerg Med* 2006 Aug;48(2):115-120. [doi: [10.1016/j.annemergmed.2006.06.015](https://doi.org/10.1016/j.annemergmed.2006.06.015)] [Medline: [16857459](https://pubmed.ncbi.nlm.nih.gov/16857459/)]
6. Glaser CA, Gilliam S, Thompson WW, Dassey DE, Waterman SH, Saruwatari M, et al. Medical care capacity for influenza outbreaks, Los Angeles. *Emerg Infect Dis* 2002 Jun;8(6):569-574 [FREE Full text] [doi: [10.3201/eid0806.010370](https://doi.org/10.3201/eid0806.010370)] [Medline: [12023911](https://pubmed.ncbi.nlm.nih.gov/12023911/)]
7. Schull MJ, Mamdani MM, Fang J. Community influenza outbreaks and emergency department ambulance diversion. *Ann Emerg Med* 2004 Jul;44(1):61-67. [doi: [10.1016/S0196064403013167](https://doi.org/10.1016/S0196064403013167)] [Medline: [15226710](https://pubmed.ncbi.nlm.nih.gov/15226710/)]
8. Osterholm MT. Preparing for the next pandemic. *N Engl J Med* 2005 May 5;352(18):1839-1842. [doi: [10.1056/NEJMp058068](https://doi.org/10.1056/NEJMp058068)] [Medline: [15872196](https://pubmed.ncbi.nlm.nih.gov/15872196/)]
9. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, Society for Academic Emergency Medicine, Emergency Department Crowding Task Force. The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med* 2009 Jan;16(1):1-10 [FREE Full text] [doi: [10.1111/j.1553-2712.2008.00295.x](https://doi.org/10.1111/j.1553-2712.2008.00295.x)] [Medline: [19007346](https://pubmed.ncbi.nlm.nih.gov/19007346/)]
10. Pines JM, Hollander JE, Localio AR, Metlay JP. The association between emergency department crowding and hospital performance on antibiotic timing for pneumonia and percutaneous intervention for myocardial infarction. *Acad Emerg Med* 2006 Aug;13(8):873-878 [FREE Full text] [doi: [10.1197/j.aem.2006.03.568](https://doi.org/10.1197/j.aem.2006.03.568)] [Medline: [16766743](https://pubmed.ncbi.nlm.nih.gov/16766743/)]
11. Pines JM, Pollack CV, Diercks DB, Chang AM, Shofer FS, Hollander JE. The association between emergency department crowding and adverse cardiovascular outcomes in patients with chest pain. *Acad Emerg Med* 2009 Jul;16(7):617-625 [FREE Full text] [doi: [10.1111/j.1553-2712.2009.00456.x](https://doi.org/10.1111/j.1553-2712.2009.00456.x)] [Medline: [19549010](https://pubmed.ncbi.nlm.nih.gov/19549010/)]
12. Pines JM, Shofer FS, Isserman JA, Abbuhl SB, Mills AM. The effect of emergency department crowding on analgesia in patients with back pain in two hospitals. *Acad Emerg Med* 2010 Mar;17(3):276-283 [FREE Full text] [doi: [10.1111/j.1553-2712.2009.00676.x](https://doi.org/10.1111/j.1553-2712.2009.00676.x)] [Medline: [20370760](https://pubmed.ncbi.nlm.nih.gov/20370760/)]
13. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]

14. Dugas AF, Hsieh Y, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clin Infect Dis* 2012 Feb 15;54(4):463-469 [FREE Full text] [doi: [10.1093/cid/cir883](https://doi.org/10.1093/cid/cir883)] [Medline: [22230244](https://pubmed.ncbi.nlm.nih.gov/22230244/)]
15. Thompson LH, Malik MT, Gumel A, Strome T, Mahmud SM. Emergency department and 'Google flu trends' data as syndromic surveillance indicators for seasonal influenza. *Epidemiol Infect* 2014 Nov;142(11):2397-2405. [doi: [10.1017/S0950268813003464](https://doi.org/10.1017/S0950268813003464)] [Medline: [24480399](https://pubmed.ncbi.nlm.nih.gov/24480399/)]
16. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. *Am J Emerg Med* 2014 Sep;32(9):1016-1023. [doi: [10.1016/j.ajem.2014.05.052](https://doi.org/10.1016/j.ajem.2014.05.052)] [Medline: [25037278](https://pubmed.ncbi.nlm.nih.gov/25037278/)]
17. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. *PLoS One* 2013;8(2):e56176 [FREE Full text] [doi: [10.1371/journal.pone.0056176](https://doi.org/10.1371/journal.pone.0056176)] [Medline: [23457520](https://pubmed.ncbi.nlm.nih.gov/23457520/)]
18. Pervaiz F, Pervaiz M, Abdur RN, Saif U. FluBreaks: early epidemic detection from Google flu trends. *J Med Internet Res* 2012;14(5):e125 [FREE Full text] [doi: [10.2196/jmir.2102](https://doi.org/10.2196/jmir.2102)] [Medline: [23037553](https://pubmed.ncbi.nlm.nih.gov/23037553/)]
19. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. *Proc Natl Acad Sci U S A* 2012 Dec 11;109(50):20425-20430 [FREE Full text] [doi: [10.1073/pnas.1208772109](https://doi.org/10.1073/pnas.1208772109)] [Medline: [23184969](https://pubmed.ncbi.nlm.nih.gov/23184969/)]
20. Preis T, Moat HS. Adaptive nowcasting of influenza outbreaks using Google searches. *R Soc Open Sci* 2014 Oct;1(2):140095 [FREE Full text] [doi: [10.1098/rsos.140095](https://doi.org/10.1098/rsos.140095)] [Medline: [26064532](https://pubmed.ncbi.nlm.nih.gov/26064532/)]
21. 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 [FREE Full text] [doi: [10.1371/journal.pcbi.1003256](https://doi.org/10.1371/journal.pcbi.1003256)] [Medline: [24146603](https://pubmed.ncbi.nlm.nih.gov/24146603/)]
22. 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 [FREE Full text] [doi: [10.1371/journal.pone.0023610](https://doi.org/10.1371/journal.pone.0023610)] [Medline: [21886802](https://pubmed.ncbi.nlm.nih.gov/21886802/)]
23. Agency for Healthcare Research and Quality. September. 2014. Healthcare Cost and Utilization Project (HCUP) Overview of the State Emergency Department Databases URL: http://www.hcup-us.ahrq.gov/db/state/sedddist/SEDD_Introduction.jsp [accessed 2016-01-28] [WebCite Cache ID 6erUqnDGv]
24. Rubinson L, Mutter R, Viboud C, Hupert N, Uyeki T, Creanga A, et al. Impact of the fall 2009 influenza A(H1N1)pdm09 pandemic on US hospitals. *Med Care* 2013 Mar;51(3):259-265. [doi: [10.1097/MLR.0b013e31827da8ea](https://doi.org/10.1097/MLR.0b013e31827da8ea)] [Medline: [23295577](https://pubmed.ncbi.nlm.nih.gov/23295577/)]
25. Google Flu Trends. 2014. URL: <https://www.google.org/flutrends/about/data/flu/historic/us-historic-v2.txt> [accessed 2014-06-05] [WebCite Cache ID 6erVyQWjB]
26. United States Census Bureau. United States Census Bureau Quick Facts. 2010. Washington: GPO URL: <http://www.census.gov/quickfacts/> [accessed 2016-06-15] [WebCite Cache ID 6iRvTQTF]
27. United States Census Bureau; Small Area Health Insurance Estimates. 2008. Washington: GPO URL: <http://www.census.gov/did/www/sahie/data/20082013/index.html> [accessed 2016-01-28] [WebCite Cache ID 6erWFeNES]
28. Centers for Medicare and Medicaid. Medicare Geographic Variation. 2008. URL: <http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/> [accessed 2016-06-15] [WebCite Cache ID 6iIS48vSo]
29. Federal Communications Commission. 2008. County Data on Internet Access Services URL: <https://www.fcc.gov/general/form-477-county-data-internet-access-services> [accessed 2016-01-28] [WebCite Cache ID 6erWq9Ncw]
30. United States Census Bureau. American Community Survey 2007-2009. 2009. URL: <http://factfinder.census.gov/> [WebCite Cache ID 6erWzT3ik]
31. Bureau of Labor Statistics. 2008. Unemployment County Data URL: <http://www.bls.gov/lau/> [accessed 2016-01-28] [WebCite Cache ID 6erXY8Dbh]
32. American Hospital Association. 2010. AHA Survey (via Area Health Resource File) URL: <http://ahrf.hrsa.gov/download.htm> [accessed 2016-01-28] [WebCite Cache ID 6erXf84tQ]
33. Centers for Disease Control. 2008. WONDER Online Database URL: <http://wonder.cdc.gov/nasa-pm.html> [accessed 2016-01-28] [WebCite Cache ID 6erXoCJVV]
34. Weather Underground. Historical Weather. URL: <http://www.wunderground.com/weather/api> [accessed 2016-01-28] [WebCite Cache ID 6erY0Wjck]
35. Legates DR, Willmott CJ. Mean seasonal and spatial variability in global surface air temperature. *Theor Appl Climatol* 1990;41(1-2):11-21. [doi: [10.1007/BF00866198](https://doi.org/10.1007/BF00866198)]
36. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 2013 Jul;49(4):764-766. [doi: [10.1016/j.jesp.2013.03.013](https://doi.org/10.1016/j.jesp.2013.03.013)]
37. Martinez W, Martinez A. *Computational Statistics Handbook with MATLAB, Second Edition (Chapman & Hall/CRC Computer Science & Data Analysis)*. Boca Raton, FL: Chapman & Hall/CRC; 2007.
38. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. *J Fam Pract* 2000 Feb;49(2):147-152. [Medline: [10718692](https://pubmed.ncbi.nlm.nih.gov/10718692/)]
39. Green CA, Pope CR. Gender, psychosocial factors and the use of medical services: a longitudinal analysis. *Soc Sci Med* 1999 May;48(10):1363-1372. [Medline: [10369437](https://pubmed.ncbi.nlm.nih.gov/10369437/)]

40. DeNavas-Walt C, Proctor B, Smith J. Income, Poverty, and Health Insurance Coverage in the United States. Washington, DC: U.S. Census Bureau; 2013. URL: <https://www.census.gov/prod/2013pubs/p60-245.pdf> [accessed 2016-06-15] [[WebCite Cache ID 6iIS1zWB](#)]
41. Newton MF, Keirns CC, Cunningham R, Hayward RA, Stanley R. Uninsured adults presenting to US emergency departments: assumptions vs data. JAMA 2008 Oct 22;300(16):1914-1924. [doi: [10.1001/jama.300.16.1914](https://doi.org/10.1001/jama.300.16.1914)] [Medline: [18940978](#)]
42. Asplin BR, Rhodes KV, Levy H, Lurie N, Crain AL, Carlin BP, et al. Insurance status and access to urgent ambulatory care follow-up appointments. JAMA 2005 Sep 14;294(10):1248-1254. [doi: [10.1001/jama.294.10.1248](https://doi.org/10.1001/jama.294.10.1248)] [Medline: [16160133](#)]
43. Tang N, Stein J, Hsia RY, Maselli JH, Gonzales R. Trends and characteristics of US emergency department visits, 1997-2007. JAMA 2010 Aug 11;304(6):664-670 [[FREE Full text](#)] [doi: [10.1001/jama.2010.1112](https://doi.org/10.1001/jama.2010.1112)] [Medline: [20699458](#)]
44. Schuur JD, Venkatesh AK. The growing role of emergency departments in hospital admissions. N Engl J Med 2012 Aug 2;367(5):391-393. [doi: [10.1056/NEJMp1204431](https://doi.org/10.1056/NEJMp1204431)] [Medline: [22784039](#)]
45. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog 2007 Oct 19;3(10):1470-1476 [[FREE Full text](#)] [doi: [10.1371/journal.ppat.0030151](https://doi.org/10.1371/journal.ppat.0030151)] [Medline: [17953482](#)]
46. Sloan C, Moore ML, Hartert T. Impact of pollution, climate, and sociodemographic factors on spatiotemporal dynamics of seasonal respiratory viruses. Clin Transl Sci 2011 Feb;4(1):48-54 [[FREE Full text](#)] [doi: [10.1111/j.1752-8062.2010.00257.x](https://doi.org/10.1111/j.1752-8062.2010.00257.x)] [Medline: [21348956](#)]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CDC: Centers for Disease Control and Prevention
ED: emergency department
FCC: Federal Communications Commission
GFT: Google Flu Trends
HCUP: Healthcare Cost and Utilization Project
ILI: influenza-like illness
IQR: interquartile range
SEDD: State Emergency Department Databases
SES: socioeconomic status

Edited by G Eysenbach; submitted 30.01.16; peer-reviewed by F Pervaiz, R Zheng; comments to author 18.02.16; revised version received 05.05.16; accepted 10.05.16; published 28.06.16

Please cite as:

*Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE, Dugas AF
Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits
J Med Internet Res 2016;18(6):e175
URL: <http://www.jmir.org/2016/6/e175/>
doi: [10.2196/jmir.5585](https://doi.org/10.2196/jmir.5585)
PMID: [27354313](https://pubmed.ncbi.nlm.nih.gov/27354313/)*

©Joseph Jeffrey Klembczyk, Mehdi Jalalpour, Scott Levin, Raynard E Washington, Jesse M Pines, Richard E Rothman, Andrea Freyer Dugas. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.06.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.