Original Paper

Integrating Smart Health in the US Health Care System: Infodemiology Study of Asthma Monitoring in the Google Era

Amaryllis Mavragani[1], BSc, MSc; Alexia Sampri[1], DipEng, MSc; Karla Sypsa[2], MPharm; Konstantinos P Tsagarakis[3], DipEng, PhD

[1]Department of Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom

[2]Department of Pharmacy and Forensic Science, King's College London, University of London, London, United Kingdom

[3]Business and Environmental Technology Economics Lab, Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece

Corresponding Author:
Amaryllis Mavragani, BSc, MSc
Department of Computing Science and Mathematics
Faculty of Natural Sciences
University of Stirling
University Campus
Stirling, FK9 4LA
United Kingdom
Phone: 44 752 378 2711
Email: amaryllis.mavragani1@stir.ac.uk

Abstract

Background: With the internet's penetration and use constantly expanding, this vast amount of information can be employed in order to better assess issues in the US health care system. Google Trends, a popular tool in big data analytics, has been widely used in the past to examine interest in various medical and health-related topics and has shown great potential in forecastings, predictions, and nowcastings. As empirical relationships between online queries and human behavior have been shown to exist, a new opportunity to explore the behavior toward asthma—a common respiratory disease—is present.

Objective: This study aimed at forecasting the online behavior toward asthma and examined the correlations between queries and reported cases in order to explore the possibility of nowcasting asthma prevalence in the United States using online search traffic data.

Methods: Applying Holt-Winters exponential smoothing to Google Trends time series from 2004 to 2015 for the term "asthma," forecasts for online queries at state and national levels are estimated from 2016 to 2020 and validated against available Google query data from January 2016 to June 2017. Correlations among yearly Google queries and between Google queries and reported asthma cases are examined.

Results: Our analysis shows that search queries exhibit seasonality within each year and the relationships between each 2 years' queries are statistically significant ($P<.05$). Estimated forecasting models for a 5-year period (2016 through 2020) for Google queries are robust and validated against available data from January 2016 to June 2017. Significant correlations were found between (1) online queries and National Health Interview Survey lifetime asthma ($r=-.82$, $P=.001$) and current asthma ($r=-.77$, $P=.004$) rates from 2004 to 2015 and (2) between online queries and Behavioral Risk Factor Surveillance System lifetime ($r=-.78$, $P=.003$) and current asthma ($r=-.79$, $P=.002$) rates from 2004 to 2014. The correlations are negative, but lag analysis to identify the period of response cannot be employed until short-interval data on asthma prevalence are made available.

Conclusions: Online behavior toward asthma can be accurately predicted, and significant correlations between online queries and reported cases exist. This method of forecasting Google queries can be used by health care officials to nowcast asthma prevalence by city, state, or nationally, subject to future availability of daily, weekly, or monthly data on reported cases. This method could therefore be used for improved monitoring and assessment of the needs surrounding the current population of patients with asthma.

(JMIR Public Health Surveill 2018;4(1):e24) doi:10.2196/publichealth.8726

XSL•FO
RenderX

KEYWORDS

asthma; big data; forecasting; Google trends; health care; health informatics; internet behavior; nowcasting; online behavior; smart health

Introduction

Health informatics is the field where information technology, computer science, social sciences, and health care meet [1]. Recently, with the use of big data (ie, large data volumes characterized by high speed and wide dataset variety [2-4]) being all the more applied in research in general, health informatics provides fertile ground for big data applications.

According to Gu et al [5], big data health care research consists of 3 research stages: disease, life and health, and nursing. Focus is being given to various aspects of diseases, technology, and health care services in areas such as epidemics, data mining, machine learning, and customized service [5]. Big data is being increasingly integrated in health care informatics [5-6] and has been used in the past in smart city management.

Over the last few years during the integration of the health pillar in smart cities, where big data is being continuously gathered and analyzed [7], the concept of smart health has been rising [8-10]. Smart health as a concept is derived from the intersection of medical informatics, public health, and business, where large volumes of social media data, payer-provider big data, genomic-driven big data, and biomedical data are being used for the monitoring and evaluation of patients' conditions [10]. As life expectancy increases, so does the cost of health care, and thus innovative methods are required to achieve improved cost-effective quality services. The use of big data in smart health can assist in P4 medicine (preventive, participatory, predictive, and personalized) [8], in the detection, prediction, and prevention of diseases [5], and in the health industry in general [10] while also taking into account the cost, data sources and quality, and population [4].

What has been of notable popularity in big data analytics is the analysis of online search queries [11-12], mainly using Google Trends [13], a popular open tool that has been widely integrated in scientific research over the course of the past decade, mainly focused on health-related topics [6]. Examples include analysis of online interest in multiple sclerosis [14], epilepsy [15-16], silicosis [17], dementia [18], urinary tract infection [19], Ebola [20], the flu [21-23], tobacco and lung cancer [24], epidemics [25-26], and even in illegal drugs such as dabbing [27], krokodil [28], and methamphetamine [29]. This use of big data has formed the cornerstone of a new concept, the science of infodemiology, which uses the vast variety of data available on the internet such as online queries, publications, or posts on blogs and websites for real-time data analysis with the aim of informing public health and public policy, thus providing a viable alternative to the time-consuming traditional methods of gathering health care data such as population surveys and registries. The use of infodemiology data for surveillance purposes is called infoveillance and could potentially allow for more timely and targeted health care interventions [30].

In this study, online queries for the term "asthma" in the United States were analyzed in order to explore the possibility of nowcasting (ie, forecasting the present) asthma prevalence using Google Trends. Asthma was selected because it is a common chronic respiratory disease characterized by exacerbations, also known as asthma attacks; therefore, the reported cases are bound to show seasonality as well as constant interest.

Asthma is a chronic condition characterized by airway inflammation and hyper-responsiveness that causes airways to constrict in response to exercise, infection, exposure to allergens, and occupational exposures [31]. In 2014, it was estimated that approximately 7.4% of the adult US population and 8.6% of US children lived with asthma [32]. During childhood, asthma is more prevalent in males, whereas in adulthood prevalence shifts toward females. Black and multirace people also have a higher prevalence than white people [33-34].

Asthma presents with coughing, wheezing, and chest tightness that seem to be worse during the night and early mornings. These symptoms, along with a family history of asthma or atopic dermatitis, can prompt investigations to confirm an asthma diagnosis. Exacerbation of normal asthma symptoms is more common in patients with uncontrolled asthma or in high-risk patients [35]. Certain types of asthma exacerbations are linked to particular seasons of the year with those caused by pollen and mold being truly seasonal [36]. It has been shown that pediatric patients experience a peak of asthma exacerbations during the fall and spring months [37], whereas adult patients experience a peak of asthma exacerbations at year end [38].

The management of asthma usually involves the use of several inhalers, leading to a rather complicated treatment regime that presents difficulties in terms of patient compliance because it interferes with their daily living activities. Poor compliance can lead to increased morbidity as well as increased cost of treatment [39]. Apart from treatment compliance, another important factor that weighs in the success of the treatment is inhaler technique, as improper inhaler use is linked to poor asthma control. Studies have shown that 33% to 94% of patients do not receive any training regarding proper inhaler technique, which leads to a great number of patients using inhalers incorrectly [40]. Asthma self-management education and personalized advice can improve a patient's asthma control and quality of life, along with reducing asthma exacerbations and hospital admissions [41].

Asthma has several social complications such as limiting patients' activity levels [42], which has an economic impact on the country's health care system. It was estimated that in 2007, medical expenses, missed work and school days, and early deaths due to asthma cost the United States $56 billion [43].

Google Trends data have been previously shown to be valid by many studies [44], and work on the subject has shown the tool's contribution to forecasting [45-46] and analysis of online behavior, provided careful selection of the examined terms [47]. The aim of this paper is to examine if nowcasting asthma prevalence in the United States is possible using online search traffic data.

XSL•FO
RenderX

Methods

Monthly time series from Google Trends for the keyword "asthma" from 2004 to 2015 in the United States and by individual state were used. The data were normalized by Google and downloaded in .csv format on July 7, 2017, between 12:47 and 13:02 for the United States and on July 18 between 14:03 and 14:33 for each of the 50 states and the District of Columbia. The data adjustment procedure is reported by Google as follows [48]: "Search results are proportionate to the time and location of a query: Each data point is divided by the total searches of the geography and time range it represents, to compare relative popularity. Otherwise places with the most search volume would always be ranked highest. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. Different regions that show the same number of searches for a term will not always have the same total search volumes."

The seasonality of asthma queries was explored followed by the estimation of the forecasts for the online interest in the term from 2016 through 2020 for the country as well as for each state. The additive method for the Holt-Winters exponential smoothing (using the statistical programming language R) is employed. The Holt-Winters equations [49] can be seen in Figure 1.

In order to further elaborate on the seasonality, the Pearson correlations for Google Trends data for the term "asthma" between each 2 years from 2004 to 2015 in the United States were calculated. Finally, the Pearson correlations between Google queries and the National Health Interview Survey (NHIS) prevalence data [50] from 2004 to 2015 and Behavioral Risk Factor Surveillance System (BRFSS) prevalence data [51] from 2004 to 2014 were examined.

Asthma is not included in the list of diseases with a Centers for Disease Control and Prevention (CDC) surveillance case definition, defined as "a set of uniform criteria used to define a disease for public health surveillance. Surveillance case definitions enable public health officials to classify and count cases consistently across reporting jurisdictions. They provide uniform criteria of national notifiable infectious and non-infectious conditions for reporting purposes" [52]. Thus, nationwide surveys are used to gather information regarding asthma prevalence, including additional information on asthma control, medications, and hospitalizations [53]. The BRFSS is a "state-based, random-digit–dialed telephone survey designed to monitor the prevalence of the major behavioral risks among adults associated with premature morbidity and mortality," and the NHIS is a "multistage probability sample survey designed to solicit health and demographic information about the population, conducted annually with face-to-face interviews in a nationally representative sample of households" [54].

In 2011, the BRFSS changed its weighting methodology in addition to also including mobile phone respondents. Therefore, any comparisons between years before and after 2011 should be carefully interpreted. In this study, no such comparisons are made, as each year's online queries are compared with the respective year's asthma reported cases, thus including no cross-year comparisons. For this study, we used the CDC definition of asthma prevalence, based on affirmative responses to the following NHIS questions: (adults) "Have you ever been told by a doctor or other health professional that you had asthma?" and "Do you still have asthma?" and (children) "Has a doctor or other professional ever told you that [sample child] had asthma?" and "Does [sample child] still have asthma?" [55].

Figure 1. Equations for Holt-Winters exponential smoothing, where y_x and $\hat{y}_x$ denote the initial series and the forecasts, respectively. The l_x, b_x, and s_x denote the level, the trend, and seasonal estimates for month x, respectively, with m denoting the period of the seasonality (ie, 12 in this case), and $h^+_m=\lfloor(h-1) mod\ m\rfloor+1$. The level, trend, and seasonal change smoothing factors are denoted by constants α, β^*, and γ, respectively. The estimated values for the coefficients for the level and trend are denoted by a and b, respectively, while the seasonal coefficients are denoted by $s_1,...,s_{12}$, for month 1,...,12, respectively.

$$(1)\quad \hat{y}_{x+h|x} = l_x + hb_x + s_{x-m+h^+_m}$$

$$(2)\quad l_x = \alpha(y_x - s_{x-m}) + (1-\alpha)(l_{x-1} + b_{x-1})$$

$$(3)\quad b_x = \beta^*(l_x - l_{x-1}) + (1-\beta^*)b_{x-1}$$

$$(4)\quad s_x = \gamma(y_x - l_{x-1} - b_{x-1}) + (1-\gamma)s_{x-m}$$

Results

Online Interest in the United States

Figure 2 shows a heat map of the United States classified into 5 groups of interest in the term "asthma" from 2004 to 2015 (ie, 0 to 20, 21 to 40, 41 to 60, 61 to 80, and 81 to 100; light blue to darker blue).

Out of the 50 states and District of Columbia, 29 fall into the 81 to 100 group, 21 in the 61 to 80 group, only 1 (Oregon) in the 41 to 60 group, and none in the 21 to 40 and 0 to 20 groups. This classification indicates that the examined term is of high interest to the population of the United States. The detailed data for Figure 2 are available in Multimedia Appendix 1, Table A1.

Figures 3 and 4 depict the changes in online interest in the term "asthma" for the period 2004 to 2015 and the seasonal changes

XSL•FO
RenderX

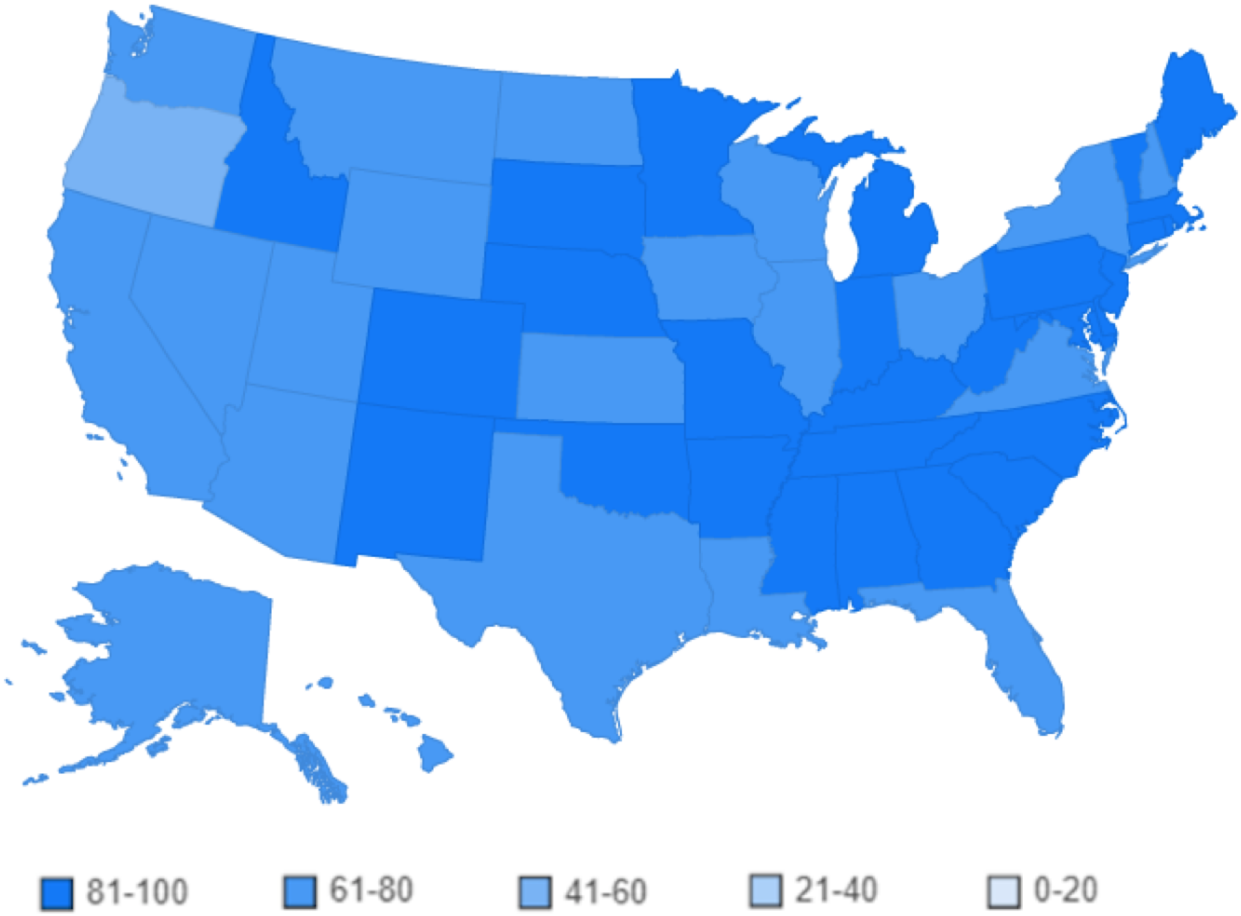
for each year from 2004 to 2015, respectively. As is evident, the data follow a seasonal trend. All years' data, as presented in Figure 4, follow a similar pattern during a full year, supporting our hypothesis that the seasonality of asthma prevalence in the United States is depicted in online searches.

Figure 5 consists of the changes by state in online interest in the term "asthma" by year from 2004 to 2015. All data are available in Multimedia Appendix 1, Table A2.

There has been a significant increase in searches for the term "asthma" in the states from 2004 to 2015, with the lowest count of states in the 81 to 100 group being in 2007 and the highest in 2012. The top asthma-related queries in the United States from January 2004 to December 2015 include "allergy asthma" (100), "asthma symptoms" (45), "asthma attack" (35), "what is asthma" (25), "asthma inhaler" (20), "asthma children" (15), "exercise asthma" (15), "asthma medications" (10), and "allergy and asthma center" (10).

As is evident, online behavioral changes toward the term "asthma" depict behavior toward said disease. The next steps are to examine if forecasting online interest in the United States is possible and identify existing relationships between online search traffic data and reported asthma cases.

Figure 2. Online interest by state in the term "asthma" from 2004 to 2015.



XSL•FO
RenderX

Figure 3. Monthly changes in online interest in the term "asthma" from 2004 to 2015.

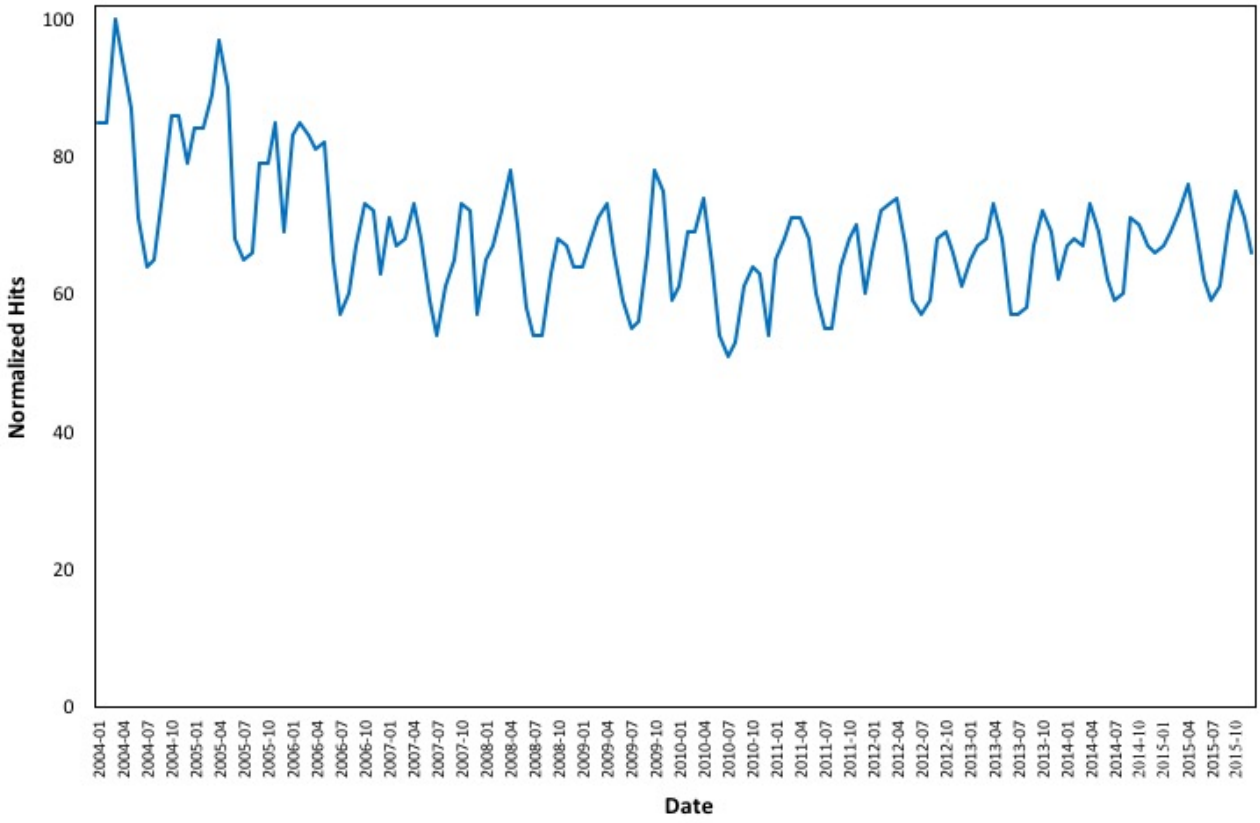
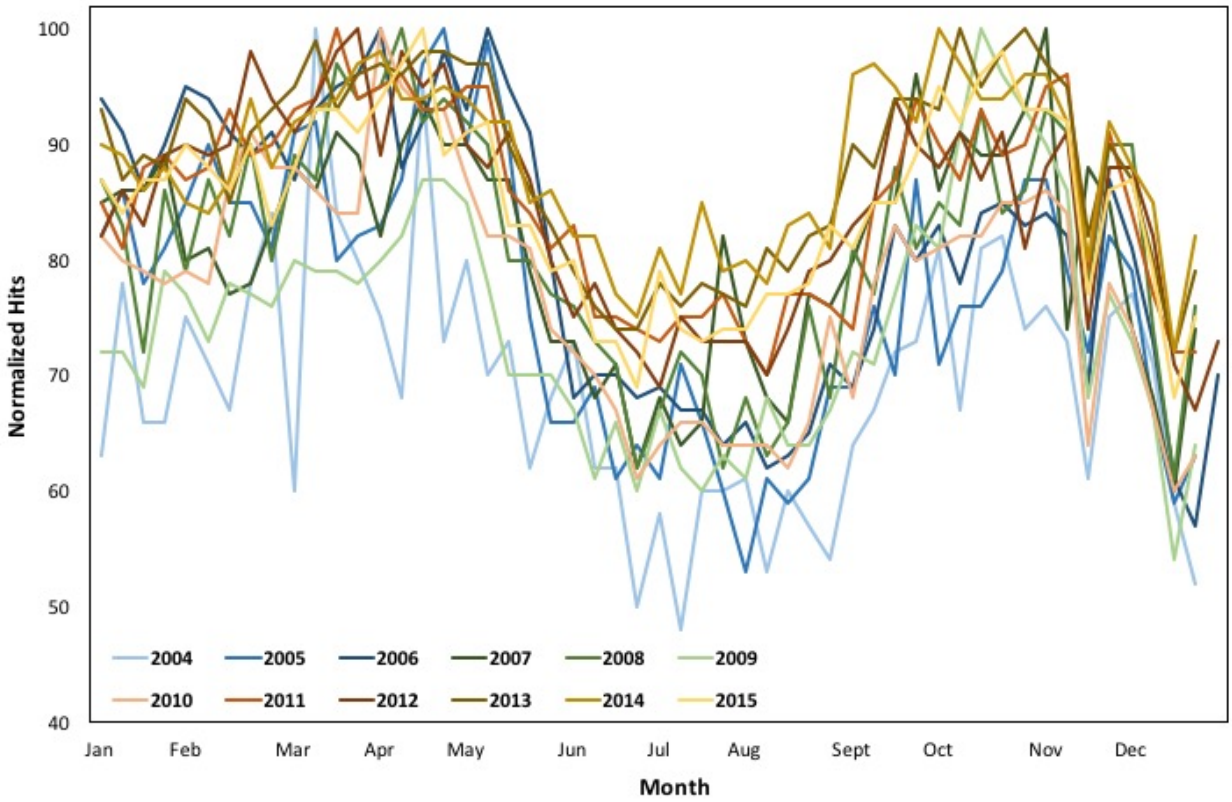
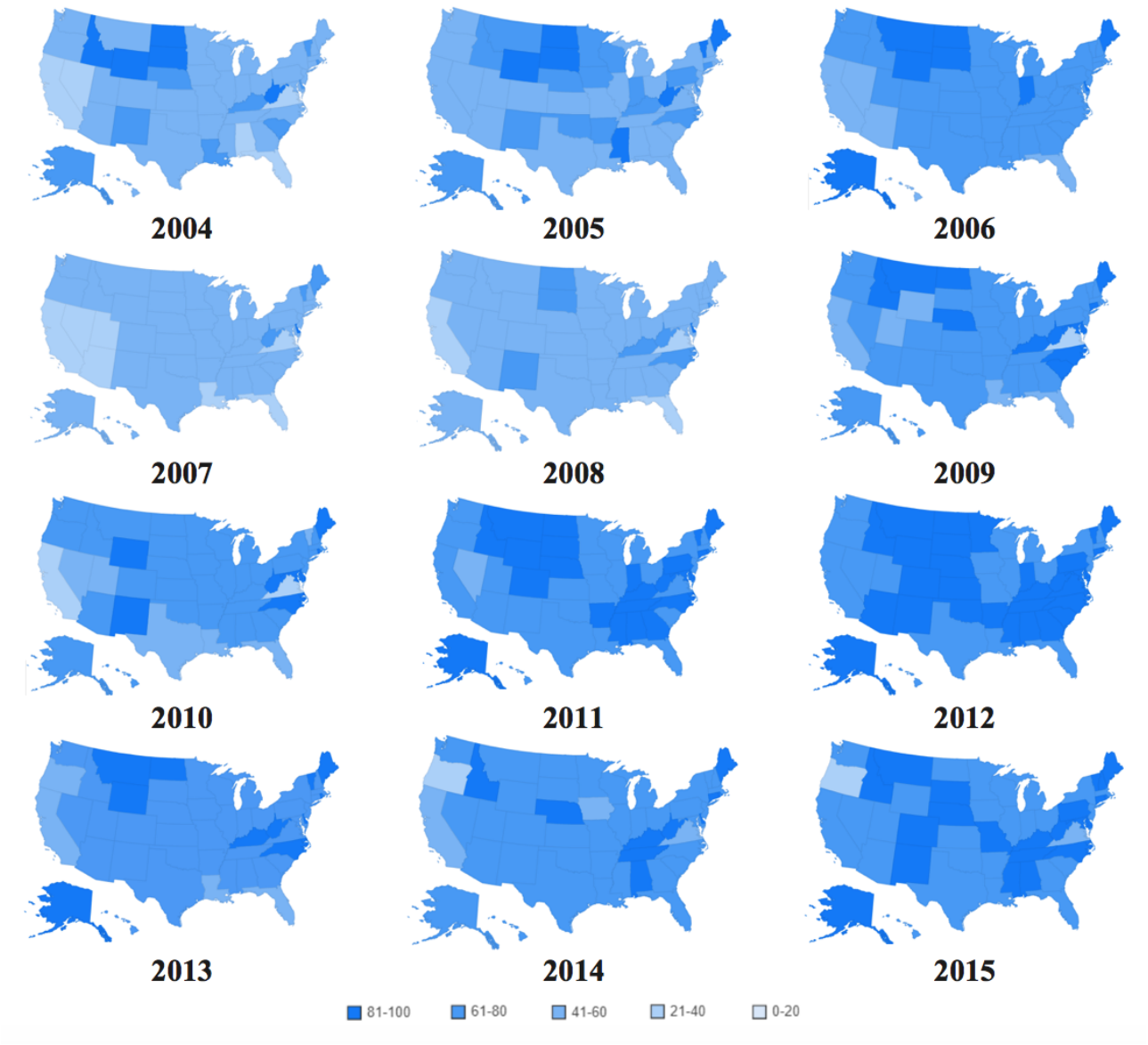


Figure 4. Weekly changes in online interest in the term "asthma" for each year from 2004 to 2015.



XSL•FO
RenderX

Figure 5. Online interest by state in the term "asthma" per year from 2004 to 2015.



Forecasting Online Interest in the United States

Figure 6 depicts changes in online interest over the period 2004 to 2015 and estimated forecasts from 2005 to 2020. The estimated model closely approximates the actual Google queries for the term "asthma" in the United States over the examined period.

The smoothing parameters for the additive Holt-Winters exponential smoothing with trend and additive seasonal component are α=.33, β*=0, and γ=.65. The estimated values for the coefficients for the level, trend and season are as follows: a=69.54, b=–.07, s_1=–.94, s_2=1.44, s_3=3.37, s_4=7.84, s_5=2.51, s_6=–5.68, s_7=–8.51, s_8=–7.20, s_9=1.89, s_{10}=4.67, s_{11}=1.11, and s_{12}=–3.53.

In order to elaborate on the robustness of the forecasting model, the estimated values are validated against the available Google queries for the term "asthma" from January 2016 to June 2017, as is shown in Figure 7. It is evident that the forecasts follow the same curve and well approximate the actual Google Trends data for the aforementioned period.

It is therefore suggested that the online behavior exhibits seasonality and can be predicted. The last step in exploring if nowcasting of asthma prevalence in the United States is possible using Google Trends is to examine the correlations between Google Trends data and reported lifetime and current asthma.

Google Trends Versus Reported Asthma

As shown in Figure 4, each examined year's online interest seems to follow a similar seasonal trend from January to December. To elaborate on the seasonal trend, the Pearson correlations between each 2 years' queries are calculated (Table 1). The monthly Google Trends data between each 2 years from 2004 to 2015 exhibit high correlations, while all comparisons are statistically significant, with P<.05.

XSL•FO
RenderX

Figure 6. Google Trends (2004 to 2015) versus forecasts (2005 to 2020) in the United States.

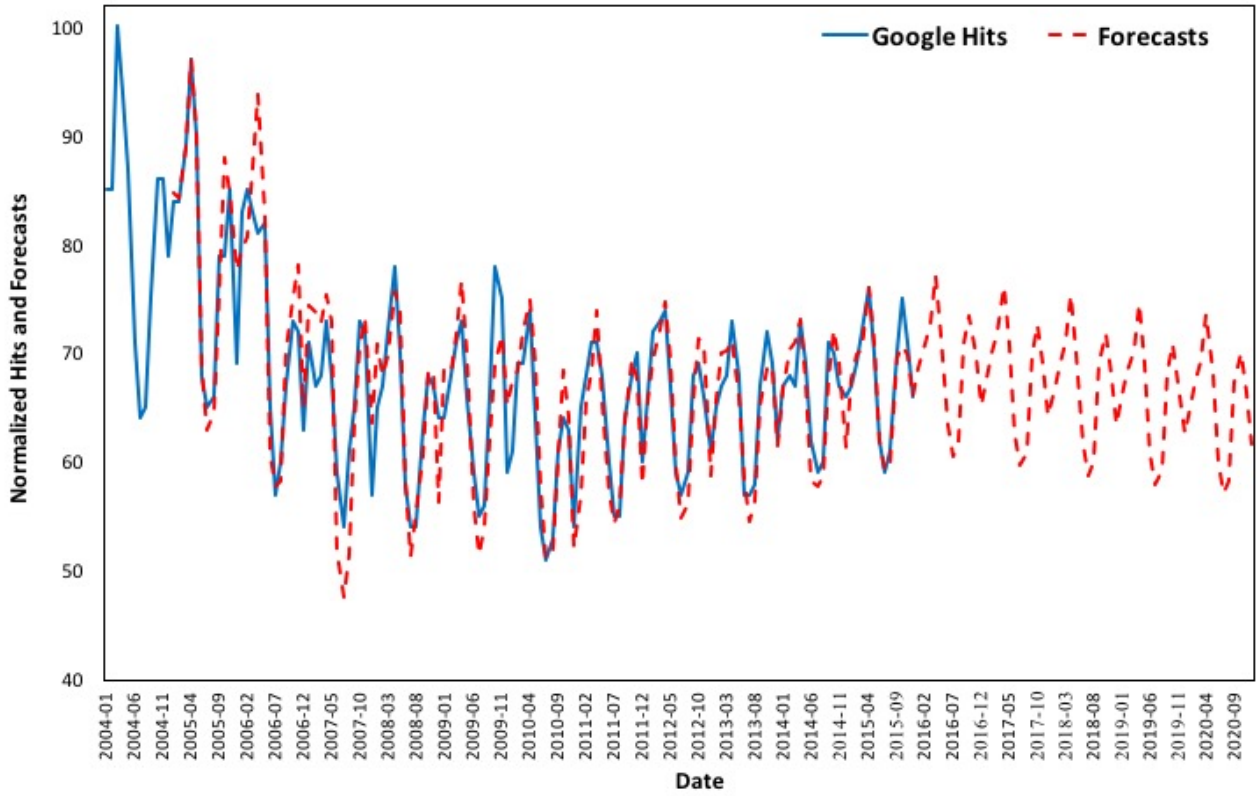
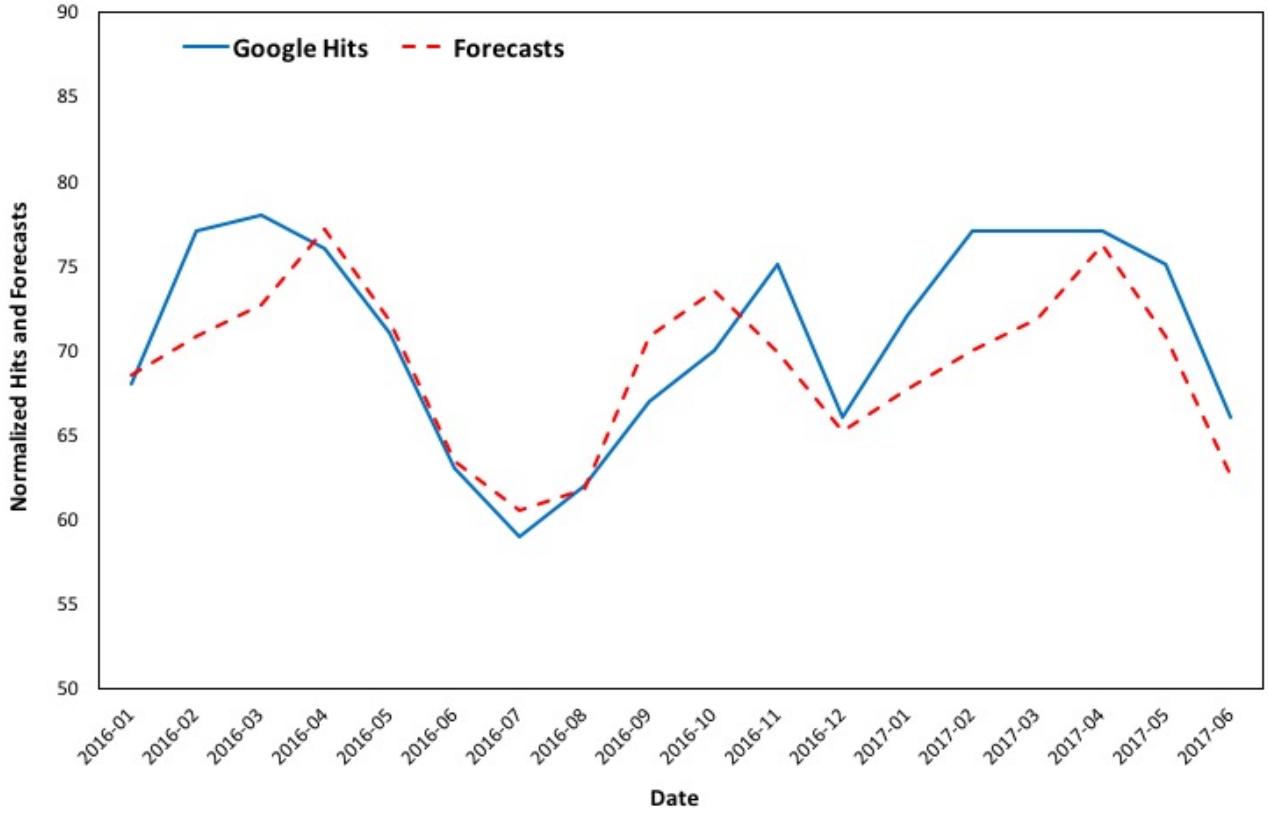


Figure 7. Google Trends (2004 to 2015) versus forecasts (January 2016 to June 2017) in the United States.



XSL•FO
RenderX

Table 1. Pearson correlations between each 2 years' normalized Google asthma queries in the United States from 2004 to 2015.

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
2005	.89	—	—	—	—	—	—	—	—	—	—
2006	.86	.89	—	—	—	—	—	—	—	—	—
2007	.77	.85	.77	—	—	—	—	—	—	—	—
2008	.94	.93	.81	.78	—	—	—	—	—	—	—
2009	.79	.76	.64	.89	.80	—	—	—	—	—	—
2010	.88	.94	.87	.82	.92	.81	—	—	—	—	—
2011	.94	.93	.85	.87	.93	.91	.93	—	—	—	—
2012	.88	.90	.85	.81	.90	.82	.98	.91	—	—	—
2013	.84	.87	.72	.89	.90	.93	.89	.92	.90	—	—
2014	.75	.82	.68	.77	.87	.78	.82	.83	.86	.92	—
2015	.86	.85	.69	.86	.92	.93	.88	.92	.90	.98	.93

Table 2. Total lifetime and current asthma National Health Interview Survey (2004 to 2015) and Behavioral Risk Factor Surveillance System (2004 to 2014) prevalence data.

Year	NHIS[a]			BRFSS[b]		
	Lifetime asthma	Current asthma	Asthma hits[c]	Lifetime asthma	Current asthma	Asthma hits[c]
2004	30,189	20,545	81.41	33,084,183	20,422,385	83.17
2005	32,621	22,227	79.58	30,661,476	19,453,974	80.33
2006	34,132	22,876	72.58	35,107,599	22,853,570	73.92
2007	34,008	22,879	65.66	36,832,798	23,556,048	68.17
2008	38,450	23,333	65.00	38,050,505	24,521,005	66.92
2009	39,930	24,567	65.83	38,033,371	24,051,245	67.92
2010	39,191	25,710	61.41	39,005,338	25,069,373	62.83
2011	39,504	25,943	64.58	34,759,106	22,605,961	66.42
2012	39,982	25,553	65.91	39,085,744	25,954,771	67.67
2013	37,328	22,648	65.25	41,030,777	26,227,484	67.00
2014	40,461	24,009	66.58	40,706,401	26,957,918	68.75
2015	40,153	24,633	68.16	—	—	—

[a]NHIS: National Health Interview Survey.

[b]BRFSS: Behavioral Risk Factor Surveillance System.

[c]Values slightly vary due to the different time frame: 2004 to 2015 for NHIS and 2004 to 2014 for BRFSS.

To further explore the relationships between online searches and asthma prevalence in the United States, data on the yearly cases of lifetime and current asthma for all ages from the NHIS prevalence data from 2004 to 2015 [50] and the BRFSS prevalence data [51] from 2004 to 2014 (Table 2) are used.

The Pearson correlations of the annual NHIS prevalence data with the annual averages of the normalized Google Trends data from 2004 to 2015 show high correlations between lifetime asthma (r=–.82, P=.001) and current asthma (r=–.77, P=.004). BRFSS prevalence data also exhibit high correlations with Google Trends data for lifetime (r=–.78, P=.003) and current asthma (r=–.79, P=.002). The Spearman correlations for the aforementioned pairs of variables all exhibit the same negative relationship, although not all are statistically significant.

Although statistically significant, all Pearson correlations are negative, and lag analysis should be employed to identify the time interval of response between asthma online interest and case reporting or vice versa. Although Google Trends data for the term "asthma" in the United States over the examined period are monthly, the data on lifetime and current asthma are yearly; until weekly or monthly data are available, further analysis cannot by done.

Forecasting Online Interest by State

In order to show that the method of nowcasting asthma prevalence in the United States using Google queries is possible, this methodology is applied in each of the 50 states and the District of Columbia and exhibits good forecasting results. Figures 8 to 11 depict the changes in online interest in the term

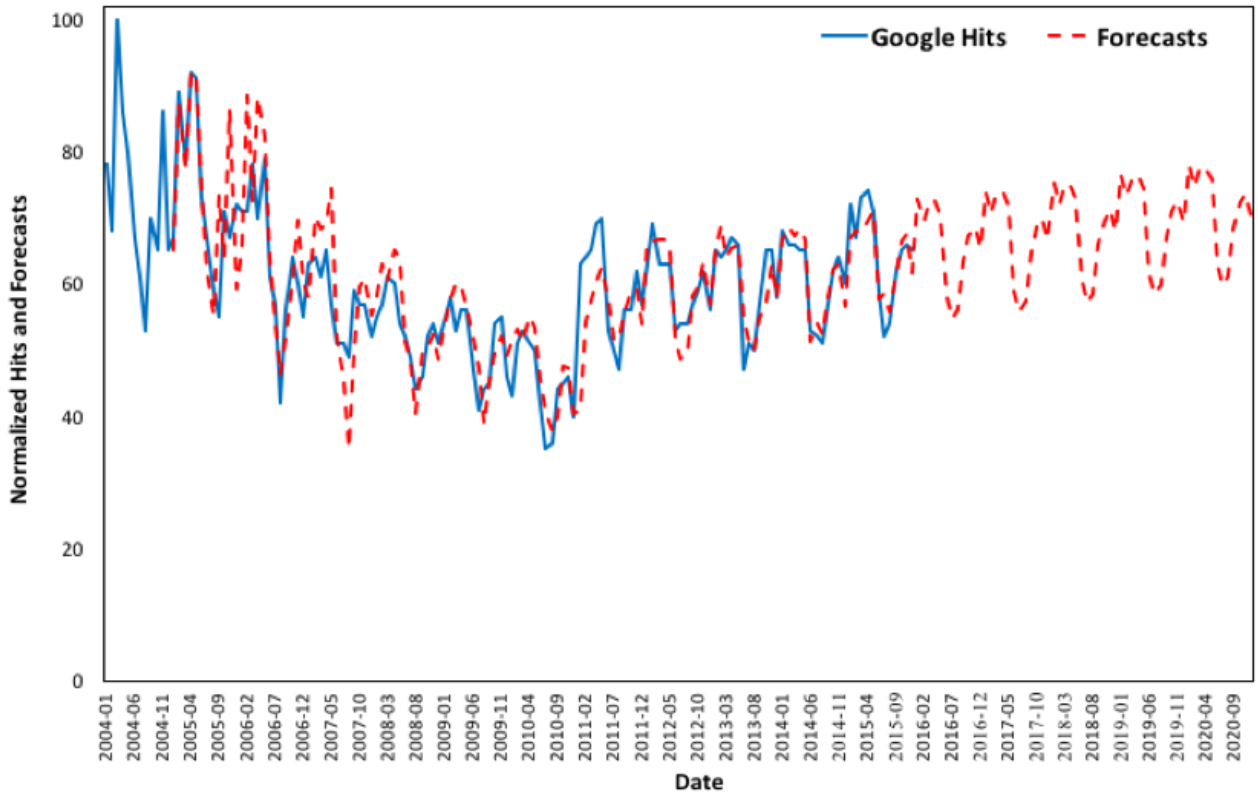
XSL•FO
RenderX

"asthma" from 2004 to 2015 and forecasts from 2016 to 2020 for the 4 most populated states (ie, California, Texas, Florida, and New York), and the graphs for all states can be found in Multimedia Appendix 2, Figures B1-B51. The values of the smoothing parameters α, β^*, and γ and the coefficients for each state's forecasts can be found in Multimedia Appendix 1, Tables A3 and A4, respectively. As online behavioral changes can be predicted and data on asthma cases are correlated with online queries, nowcasting of asthma could be possible provided short-interval data (eg, monthly, weekly, or even daily) are available.

According to the results, online interest in Alaska, Nebraska, New Hampshire, Oklahoma, and Tennessee exhibits increasing forecast trends from 2016 to 2020. On the contrary, online interest in Delaware, Kansas, Oregon, and Virginia exhibits decreasing forecast trends from 2016 to 2020. Overall, the states of Arizona, California, Connecticut, Florida, Georgia, Illinois, Indiana, Maryland, Michigan, Missouri, New Jersey, New York, North Carolina, Pennsylvania, Texas, and Washington show high interest in the term "asthma" throughout the examined period, while in Hawaii and Wyoming, interest is low. Virginia is the only state where online interest exhibits very significant variations from 2004 to 2016.

Our study indicates that analysis of online behavior toward asthma by state can assist with nowcasting asthma prevalence. Since search queries and reporting of asthma are shown to correlate in the United States, if short-interval data (eg, weekly or monthly) were made available, a robust nowcasting model could be developed.

Figure 8. Google Trends (2004 to 2015) versus forecasts (2005 to 2020) in California.



XSL•FO
RenderX

Figure 9. Google Trends (2004 to 2015) versus forecasts (2005 to 2020) in Texas.

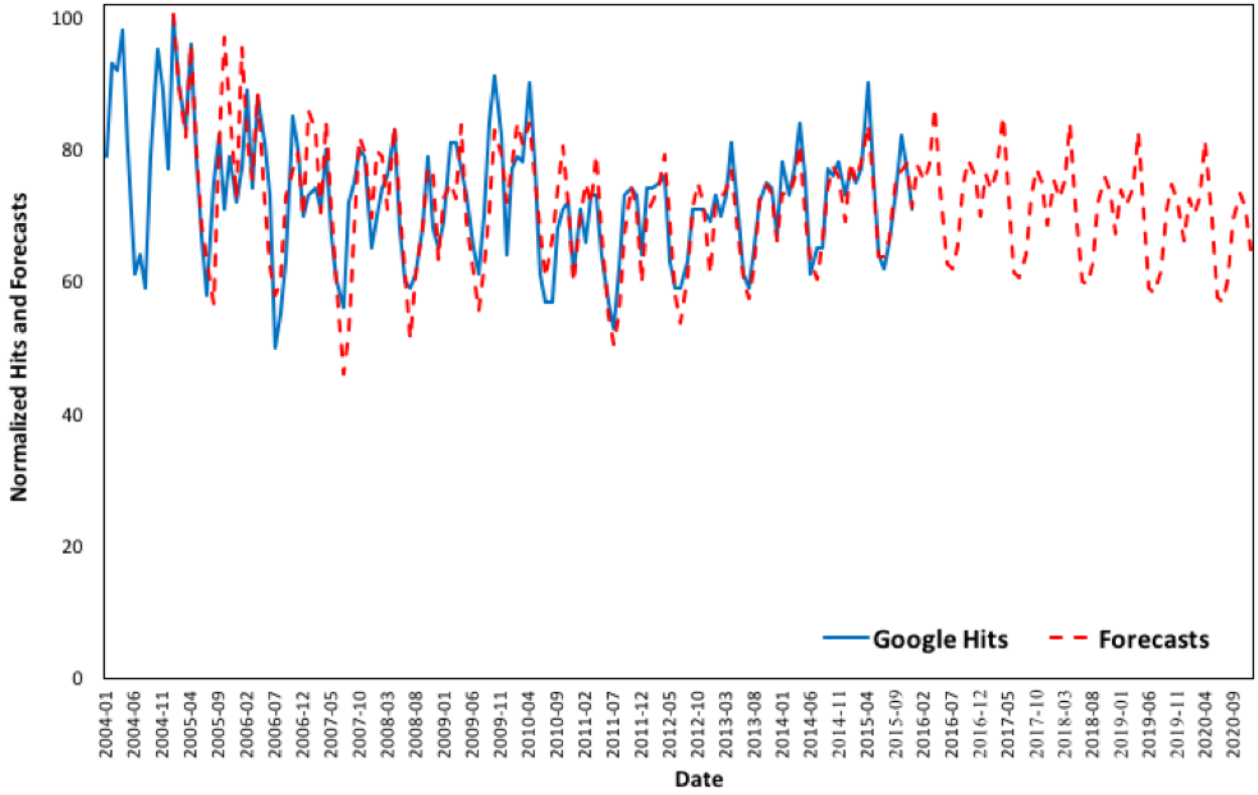
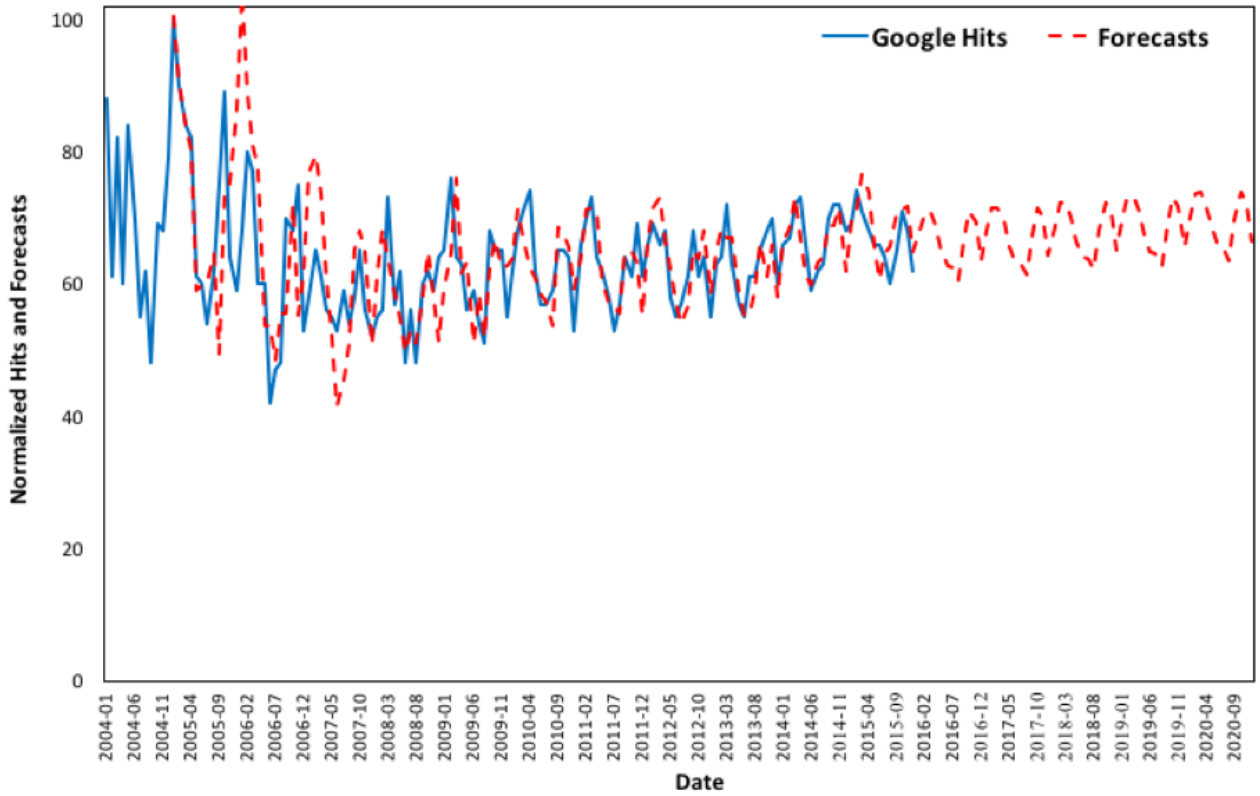
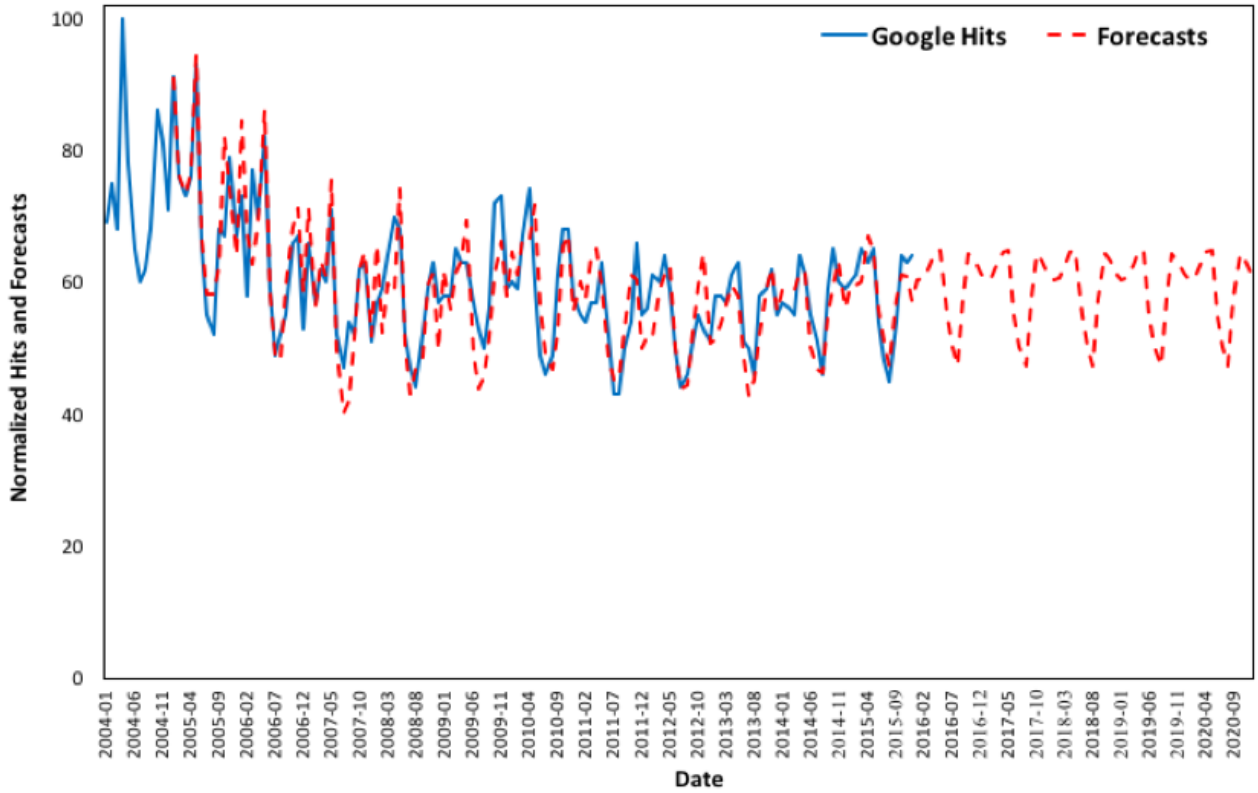


Figure 10. Google Trends (2004 to 2015) versus forecasts (2005 to 2020) in Florida.



XSL•FO
RenderX

Figure 11. Google Trends (2004 to 2015) versus forecasts (2005 to 2020) in New York.



Discussion

Principal Findings

In addressing integration of smart health into smart city management, monitoring of search traffic data could be useful in predictions and nowcastings, as has also been suggested by previous work on the subject. This study shows that online interest can be predicted nationally and by state. Therefore, governments, policy makers, and health care officials have the ability to use these data to better address the responsiveness of the US health care system at national, regional, state, or even city level in order to nowcast asthma prevalence. Google Trends also provides detailed regional US data, and this method can be applied in other countries as well.

Empirical relationships between Google Trends and human behavior have been suggested, therefore nowcasting asthma prevalence in the United States is possible using online search traffic data, subject to availability of daily, weekly, or monthly data. In this study, it was shown that online search traffic data are highly correlated between each 2 years during the examined period and that Google Trends data are correlated with reported cases of lifetime and current asthma in the United States from 2004 to 2015.

After analyzing changes in online interest in the United States over the examined period, the next step was to identify any seasonal similarities between each 2 years' (monthly) search queries. As the hits between each 2 years from 2004 to 2015 on the term "asthma" were highly correlated, the seasonal effect was evident; using Holt-Winters exponential smoothing, 5-year forecasts for online interest in the term from 2016 to 2020 nationally and in each state were estimated. Validated against available data from January 2016 to June 2017, the forecasts were well fitted and accurately approximated the actual Google Trends data for the same period, suggesting seasonal behavioral changes over the course of a year can be accurately predicted using the proposed method. Google Trends data are correlated with reported cases of lifetime and current asthma, and thus nowcasting asthma prevalence in the United States is suggested to be possible using online search traffic data. As the calculated correlations are negative at this point and there is a lag between internet queries and asthma reporting and vice versa, short-interval data (eg, monthly, weekly, and daily—not available at this point) are required in order to identify said lag.

Limitations

This study has limitations. It cannot be assumed that each hit corresponds to an asthma case and vice versa because hits could be also attributed to academic or research reasons or general interest on the subject, and they could be influenced by news reports or social media. Queries related to asthma could be also influenced by factors such as changes of health insurance and weather or environmental conditions that trigger similar symptoms. This is a general limitation when examining online queries, despite the empirical relationships that have been shown to exist between Google Trends and health data.

The sample is not representative, although as internet penetration increases, so does the possibility of higher volumes of online queries being related to asthma cases. Additionally, nowcasting asthma prevalence using online search queries is not possible at this point because the available data on reported lifetime and

XSL•FO
RenderX

current asthma are yearly. If monthly, weekly, or daily data on past asthma prevalence were available and the correlations between search traffic data and reported asthma are validated, the possibility of nowcasting asthma could be further explored.

This study has not accounted for state-by-state confounders that could influence search patterns, such as the socioeconomic status and demographics of different states that might be relevant to asthma prevalence, as this exceeds the scope of this paper. The latter, along with the impact of socioeconomic and cultural differences on asthma reporting and online search patterns, are of interest for further investigation. In addition, more search terms related to asthma symptoms such as "breathlessness" and "wheezing" could be included in future research on asthma monitoring in the United States.

Conclusion

The findings of this study support previous work on the subject and highlight the value of online data in health and medical informatics. Google Trends data have been shown to be useful and valuable in the monitoring, surveillance, or prediction of epidemics and outbreaks [20,25-26,56], as have been various other internet sources such as Twitter [57], medical portals [58], and Baidu [59]. Google queries provide us with the revealed and not the stated user interest contrary to traditional survey methods [60], and the use of Web data will benefit the exploration of behavior in medical issues [61]. Data from traditional sources and big data should be combined in order to take full advantage of all available information [62]. When daily, weekly, or monthly data on reported asthma cases are made available, data from online sources like Google Trends could be used centrally and then applied by state or used by each city or state individually, assisting with the integration of the smart health concept in smart city management.

Internet behavior can be measured by infodemiology metrics as information patterns and population health are related [30]. Surveillance of asthma is mainly assessed through nationwide surveys and interviews, and data on asthma prevalence are only available long after the cases of asthma are reported. Nowcasting Google queries on selected terms related to asthma could assist health officials at both national and state levels to detect any behavioral variations toward the disease, providing time-effective allocation of resources and a more cost-effective approach to asthma assessment. This study suggests a relationship between asthma prevalence and Google Trends data. In the future, analysis of online queries could be valuable in the monitoring and evaluation of the responsiveness of the US health care system to asthma patient admissions and prescription drug needs, as well as assisting with the implementation of targeted health interventions and campaigns during periods when increased asthma admissions are predicted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

State data tables.

[PDF File (Adobe PDF File), 52KB - publichealth_v4i1e24_app1.pdf]

Multimedia Appendix 2

Google Trends (2004 to 2015) versus forecasts (2005 to 2020) by state.

[PDF File (Adobe PDF File), 3MB - publichealth_v4i1e24_app2.pdf]

References

1. Vignesh RP, Sivasankar E, Pitchiah R. Framework for smart health: toward connected data from big data. 2015 Presented at: Intelligent Computing and Applications Proceedings of the International Conference; December 22-24, 2014; New Delhi p. 423-433. [doi: 10.1007/978-81-322-2268-2]
2. Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. Science 2011 Apr 01;332(6025):60-65 [FREE Full text] [doi: 10.1126/science.1200970] [Medline: 21310967]
3. Chen CP, Zhang C. Data-intensive applications, challenges, techniques and technologies: a survey on big data. Information Sciences 2014 Aug;275:314-347. [doi: 10.1016/j.ins.2014.01.015]
4. Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J. Applications of big data to smart cities. J Internet Serv Appl 2015 Dec 1;6(1). [doi: 10.1186/s13174-015-0041-5]
5. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Inform 2017 Feb;98:22-32. [doi: 10.1016/j.ijmedinf.2016.11.006] [Medline: 28034409]
6. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
7. Khan J, Anjum A, Soomro K, Tahir M. Towards cloud based big data analytics for smart future cities. J Cloud Comp 2015 Feb 18;4(1). [doi: 10.1186/s13677-015-0026-8]

XSL•FO
RenderX

8. Holzinger A, Röcker C, Ziefle M. From smart health to smart hospitals. Lect Notes Comput Sc 2015;8700:1-20. [doi: 10.1007/978-3-319-16226-3_1]
9. Jung H, Chung K. Sequential pattern profiling based bio-detection for smart health service. Cluster Comput 2014 Apr 3;18(1):209-219. [doi: 10.1007/s10586-014-0370-3]
10. Pramanik I, Lau R, Demirkan H, Azad A. Smart health: big data enabled health paradigm within smart cities. Expert Syst Appl 2017 Nov;87:370-383 [FREE Full text] [doi: 10.1016/j.eswa.2017.06.027]
11. Preis T, Moat HS, Stanley HE, Bishop SR. Quantifying the advantage of looking forward. Sci Rep 2012;2:350 [FREE Full text] [doi: 10.1038/srep00350] [Medline: 22482034]
12. Preis T, Moat HS, Stanley HE. Quantifying trading behavior in financial markets using Google Trends. Sci Rep 2013;3:1684 [FREE Full text] [doi: 10.1038/srep01684] [Medline: 23619126]
13. Google Trends. URL: https://trends.google.com/trends/explore [accessed 2017-08-11] [WebCite Cache ID 6sdKCrWKt]
14. Brigo F, Lochner P, Tezzon F, Nardone R. Web search behavior for multiple sclerosis: an infodemiological study. Mult Scler Relat Disord 2014 Jul;3(4):440-443. [doi: 10.1016/j.msard.2014.02.005] [Medline: 25877054]
15. Bragazzi NL, Bacigaluppi S, Robba C, Nardone R, Trinka E, Brigo F. Infodemiology of status epilepticus: a systematic validation of the Google Trends-based search queries. Epilepsy Behav 2016 Feb;55:120-123. [doi: 10.1016/j.yebeh.2015.12.017] [Medline: 26773681]
16. Brigo F, Igwe SC, Ausserer H, Nardone R, Tezzon F, Bongiovanni LG, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. Epilepsy Behav 2014 Feb;31:67-70. [doi: 10.1016/j.yebeh.2013.11.020] [Medline: 24361764]
17. Bragazzi NL, Dini G, Toletone A, Brigo F, Durando P. Leveraging big data for exploring occupational diseases-related interest at the level of scientific community, media coverage and novel data streams: the example of silicosis as a pilot study. PLoS One 2016;11(11):e0166051 [FREE Full text] [doi: 10.1371/journal.pone.0166051] [Medline: 27806115]
18. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: evidence From Taiwan. J Med Internet Res 2015;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
19. Rossignol L, Pelat C, Lambert B, Flahault A, Chartier-Kastler E, Hanslik T. A method to assess seasonality of urinary tract infections based on medication sales and Google Trends. PLoS One 2013;8(10):e76020 [FREE Full text] [doi: 10.1371/journal.pone.0076020] [Medline: 24204587]
20. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty 2015 Dec 10;4:54 [FREE Full text] [doi: 10.1186/s40249-015-0090-9] [Medline: 26654247]
21. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
22. Davidson MW, Haim DA, Radin JM. Using networks to combine big data and traditional surveillance to improve influenza predictions. Sci Rep 2015 Jan 29;5:8154 [FREE Full text] [doi: 10.1038/srep08154] [Medline: 25634021]
23. Dukic V, Lopes H, Polson N. Tracking epidemics with Google Flu Trends data and a state-space SEIR model. J Am Stat Assoc 2012 Aug 14;107(500):1410-1426. [doi: 10.1080/01621459.2012.713876]
24. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Information seeking regarding tobacco and lung cancer: effects of seasonality. PLoS One 2015;10(3):e0117938 [FREE Full text] [doi: 10.1371/journal.pone.0117938] [Medline: 25781020]
25. Sentana-Lledo D, Barbu CM, Ngo MN, Wu Y, Sethuraman K, Levy MZ. Seasons, searches, and intentions: what the Internet can tell us about the bed bug (Hemiptera: Cimicidae) epidemic. J Med Entomol 2016 Jan;53(1):116-121. [doi: 10.1093/jme/tjv158] [Medline: 26474879]
26. Fenichel EP, Kuminoff NV, Chowell G. Skip the trip: air travelers' behavioral responses to pandemic influenza. PLoS One 2013;8(3):e58249 [FREE Full text] [doi: 10.1371/journal.pone.0058249] [Medline: 23526970]
27. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Tracking dabbing using search query surveillance: a case study in the United States. J Med Internet Res 2016 Sep 16;18(9):e252 [FREE Full text] [doi: 10.2196/jmir.5802] [Medline: 27637361]
28. Zheluk A, Quinn C, Meylakhs P. Internet search and krokodil in the Russian Federation: an infoveillance study. J Med Internet Res 2014 Sep 18;16(9):e212 [FREE Full text] [doi: 10.2196/jmir.3203] [Medline: 25236385]
29. Gamma A, Schleifer R, Weinmann W, Buadze A, Liebrenz M. Could Google Trends be used to predict methamphetamine-related crime? An analysis of search volume data in Switzerland, Germany, and Austria. PLoS One 2016;11(11):e0166566 [FREE Full text] [doi: 10.1371/journal.pone.0166566] [Medline: 27902717]
30. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
31. Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. NCHS Data Brief 2012 May(94):1-8 [FREE Full text] [Medline: 22617340]
32. FastStats: Asthma. National Center for Health Statistics. URL: http://www.cdc.gov/nchs/fastats/asthma.htm [accessed 2018-02-15] [WebCite Cache ID 6sdJKRPmf]

XSL•FO
RenderX

33. Centers for Disease Control and Prevention. Most recent asthma data URL: http://www.cdc.gov/asthma/most_recent_data.htm [WebCite Cache ID 6sdJSHspd]
34. Asthma and African Americans.: US Department of Health and Human Services URL: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=15 [accessed 2018-02-21] [WebCite Cache ID 6xP2qadGK]
35. Global Initiative for Asthma. 2017 Pocket guide for asthma management and prevention URL: http://ginasthma.org/download/520/ [WebCite Cache ID 6xFIlR1Ry]
36. Johnston NW, Sears MR. Asthma exacerbations: epidemiology. Thorax 2006 Aug;61(8):722-728 [FREE Full text] [doi: 10.1136/thx.2005.045161] [Medline: 16877691]
37. Larsen K, Zhu J, Feldman LY, Simatovic J, Dell S, Gershon AS, et al. The annual September peak in asthma exacerbation rates. Still a reality? Ann Am Thorac Soc 2016 Feb;13(2):231-239. [doi: 10.1513/AnnalsATS.201508-545OC] [Medline: 26636481]
38. Gerhardsson DVM, Gustafson P, McCrae C, Edsbäcker S, Johnston N. Seasonal and geographic variations in the incidence of asthma exacerbations in the United States. J Asthma 2017 Oct;54(8):818-824. [doi: 10.1080/02770903.2016.1277538] [Medline: 28102717]
39. Gillisen A. Patient's adherence in asthma. J Physiol Pharmacol 2007 Nov;58 Suppl 5(Pt 1):205-222 [FREE Full text] [Medline: 18204131]
40. Axtell S, Haines S, Fairclough J. Effectiveness of various methods of teaching proper inhaler technique. J Pharm Pract 2017 Apr;30(2):195-201. [doi: 10.1177/0897190016628961] [Medline: 26912531]
41. Pinnock H. Supported self-management for asthma. Breathe (Sheff) 2015 Jun;11(2):98-109 [FREE Full text] [doi: 10.1183/20734735.015614] [Medline: 26306110]
42. Winer RA, Qin X, Harrington T, Moorman J, Zahran H. Asthma incidence among children and adults: findings from the Behavioral Risk Factor Surveillance system asthma call-back survey—United States, 2006-2008. J Asthma 2012 Feb;49(1):16-22. [doi: 10.3109/02770903.2011.637594] [Medline: 22236442]
43. Centers for Disease Control and Prevention. Vital Signs: Asthma in the US URL: http://www.cdc.gov/vitalsigns/asthma/ [WebCite Cache ID 6sdJfBsTX]
44. McCallum M, Bury G. Public interest in the environment is falling: a response to Ficetola (2013). Biodivers Conserv 2014 Feb 14;23(4):1057-1062. [doi: 10.1007/s10531-014-0640-7]
45. Jun S, Park D. Consumer information search behavior and purchasing decisions: empirical evidence from Korea. Technol Forecast Soc Change 2016 Jun;107:97-111. [doi: 10.1016/j.techfore.2016.03.021]
46. Han S, Chung H, Kang B. It is time to prepare for the future: forecasting social trends. In: Computer Applications for Database, Education, and Ubiquitous Computing. Berlin: Springer; 2012:325-331.
47. Scharkow M, Vogelgesang J. Measuring the public agenda using search engine queries. Int J Public Opin Res 2011 Mar 01;23(1):104-113. [doi: 10.1093/ijpor/edq048]
48. How Trends data is adjusted. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2017-08-11] [WebCite Cache ID 6sdJp7avA]
49. Hyndman R, Athanasopoulos G. Forecasting Principles and Practice. 2014. URL: https://www.otexts.org/fpp [accessed 2018-02-15] [WebCite Cache ID 6xFJlXCQI]
50. Centers for Disease Control and Prevention. 2015. National Health Interview Survey (NHIS) Data: 2015 lifetime asthma, current asthma, asthma attacks among those with current asthma URL: https://www.cdc.gov/asthma/nhis/2015/data.htm [WebCite Cache ID 6sdJhJx5r]
51. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System (BRFSS) prevalence data: asthma URL: https://www.cdc.gov/asthma/brfss/default.htm [WebCite Cache ID 6sdJl5fpw]
52. Centers for Disease Control and Prevention. Surveillance Case Definitions URL: https://wwwn.cdc.gov/nndss/conditions/ [WebCite Cache ID 6wPl4S6NQ]
53. Centers for Disease Control and Prevention. Breathing easier URL: https://www.cdc.gov/asthma/pdfs/breathing_easier_brochure.pdf [WebCite Cache ID 6u57QQ7tx]
54. Centers for Disease Control and Prevention. Data and Surveillance: asthma URL: https://www.cdc.gov/asthma/tables_graphs.htm [WebCite Cache ID 6wPlB6D40]
55. Centers for Disease Control and Prevention. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010 URL: https://www.cdc.gov/nchs/data/databriefs/db94.pdf [WebCite Cache ID 6xFKYZQio]
56. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using Web data: the case of influenza in Greece and Italy using Google Trends. JMIR Public Health Surveill 2017 Nov 20;3(4):e90 [FREE Full text] [doi: 10.2196/publichealth.8015] [Medline: 29158208]
57. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using Twitter data. J Med Internet Res 2017 Sep 12;19(9):e315 [FREE Full text] [doi: 10.2196/jmir.7393] [Medline: 28899847]
58. Pesälä S, Virtanen MJ, Sane J, Mustonen P, Kaila M, Helve O. Health information-seeking patterns of the general public and indications for disease surveillance: register-based study using Lyme disease. JMIR Public Health Surveill 2017 Nov 06;3(4):e86 [FREE Full text] [doi: 10.2196/publichealth.8306] [Medline: 29109071]

XSL•FO
RenderX

59. Liu K, Huang S, Miao Z, Chen B, Jiang T, Cai G, et al. Identifying potential norovirus epidemics in China via Internet surveillance. J Med Internet Res 2017 Aug 08;19(8):e282 [FREE Full text] [doi: 10.2196/jmir.7855] [Medline: 28790023]
60. Mavragani A, Tsagarakis K. YES or NO: predicting the 2015 GReferendum results using Google Trends. Technol Forecast Soc Change 2016 Aug;109:1-5. [doi: 10.1016/j.techfore.2016.04.028]
61. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 9;311(14):1399-1400. [doi: 10.1001/jama.2014.1505] [Medline: 24577162]
62. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
NHIS: National Health Interview Survey

Edited by G Eysenbach; submitted 11.08.17; peer-reviewed by N Bragazzi, Z Zhang, A Zheluk; comments to author 21.09.17; revised version received 15.10.17; accepted 13.01.18; published 12.03.18

Please cite as:
Mavragani A, Sampri A, Sypsa K, Tsagarakis KP
Integrating Smart Health in the US Health Care System: Infodemiology Study of Asthma Monitoring in the Google Era
JMIR Public Health Surveill 2018;4(1):e24
URL: http://publichealth.jmir.org/2018/1/e24/
doi:10.2196/publichealth.8726
PMID:29530839

©Amaryllis Mavragani, Alexia Sampri, Karla Sypsa, Konstantinos P Tsagarakis. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 12.03.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX