Searching for Cedar: Geographic Variation in Single Aeroallergen Shows Dose Response in Internet Search Activity

Thomas J. Willson, MD¹,², Alexandra Shams², Joshua Lospinoso, PhD³, Erik Weitzel, MD¹,², and Kevin McMains, MD²,⁴

Abstract

Objective. To assess the effect of a single aeroallergen on Internet search activity in 3 distinct geographic regions over time.

Study Design. A retrospective time series analysis.

Setting. Tertiary academic medical center.

Methods. Mountain Cedar pollen counts spanning the periods from 2011 to 2014 were collected for San Antonio, Dallas–Ft Worth, and Houston metro regions (Texas). Google search data for the corresponding period and regions were also obtained for the search terms pollen and mountain cedar. Data were analyzed through time series plots and autoregressive integrated moving average.

Results. Seasonal maximum pollen counts were greatest in San Antonio, ranging from 5413 to 9982 grains per cubic centimeter. In San Antonio, there was a strong, positive, and statistically significant effect of mountain cedar pollen count on search activity for both pollen (P = .001) and mountain cedar (P < .001) when evaluated by autoregressive integrated moving average.

Conclusions. Environmental levels of a single allergen—mountain cedar pollen—show a geographic and temporal relationship between Internet searches and mountain cedar pollen, based on Google Trends. These data are useful to understand patients’ perceptions of the health risk presented by mountain cedar pollen. In turn, this information affords practitioners the opportunity to offer targeted patient education in a geographically unique region, as well as adjust surveillance of related pathologies.

Keywords
aeroallergen, allergen, mountain cedar, juniper, Google, Internet

Received May 26, 2015; revised July 13, 2015; accepted July 29, 2015.

Mountain cedar (Juniperus ashei) is an evergreen tree that populates the southern United States, occupying 8.6 million acres in Texas.¹ Although the area is vast, its region of influence is relatively localized to central and southern Texas (Figure 1). This dioecious species is unique, with the male plant pollinating during the winter months instead of the spring. Each male plant is capable of producing 1 billion grains of pollen. Pollen produced by this species is highly allergenic and has been documented as the number one cause of allergic rhinoconjunctivitis in South Texas.¹ Since mountain cedar is so allergenic and pollinates in a season when other plants are dormant, it is a useful model for studying the effect of individual allergens on specific geographic populations. It is well documented that patients often utilize the Internet to access information about their health prior to seeking medical care. Correlation of health-related Internet search activity with environmental factors according to Google Trends has been demonstrated and has been previously used to track historical and real-time influenza outbreaks.²–⁴ The focus of this study was to investigate Internet search activity related to the presence of a single allergen over several regions.

¹Department of Otolaryngology–Head and Neck Surgery, San Antonio Military Medical Center, San Antonio, Texas, USA
²Uniformed Services University of Health Sciences, Bethesda, Maryland, USA
³Portia Statistical Consulting, LLC, San Antonio, Texas, USA
⁴ENT Section, South Texas Veterans Health Care System, San Antonio, Texas, USA

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting true views of the Department of the Army, Department of the Air Force, or the Department of Defense.

Corresponding Author:
Thomas J. Willson, MD, Department of Otolaryngology–Head and Neck Surgery, San Antonio Military Medical Center, 3551 Roger Brooke Dr, San Antonio, TX 78234, USA.
Email: S10twillson@gmail.com
Approval was obtained from the San Antonio Military Medical Center Institutional Review Board. Mountain cedar pollen counts spanning the periods from 2011 to 2014 were collected for San Antonio (SAT), Dallas–Ft Worth (DFW), and Houston (HOU) metro regions in Texas. Archived pollen data were collected from the following National Allergy Bureau collection sites: for DFW, http://entdocs.com/pollen-count/; for HOU, the city website at http://www.houstontx.gov/health/pollen-and-mold-spore-count-archives; and for SAT, from Wilford Hall Ambulatory Surgical Center records. Since mountain cedar pollen counts are generally close to 0 outside its season, counts were collected only from October 10 to April 15 for each year analyzed. Pollen counts were transformed to a 100-point relative index by setting the maximum observed pollen count for the data set to 100. This provided ease of visualization graphically. Google search data for the corresponding period and regions were also obtained for the search terms pollen and mountain cedar. These data are accessible at http://google.com/trends. Search terms were input and results returned. Regions of interest were then selected and data sets downloaded. (While the Google Trends site displays results in a time series graph, a comma-separated values [or csv] file is made available for raw data download.) Search data are compiled and stored by Google and are specific to time and region. Its proprietary algorithm computes a baseline level of activity in each region and then detects search activity for terms that exceeds the geographic and temporal baseline. A more detailed explanation of this process is found online at https://support.google.com/trends/answer/4355213?hl=en.

Time series plots were constructed by region to compare data visually and to assess for temporal correlation. Statistical analysis of the data was performed through an autoregressive integrated moving average (ARIMA) using R statistical software v. 3.1 (Vienna, Austria). Time series modeling with ARIMA statistical models is common practice in many fields and can help detect the influence of an independent time series on a dependent time series. ARIMA models allow analysis controlled for seasonal variation, lagged effects, and other so-called nuisance variables that affect the variables of interest. In this study, ARIMA models were used to determine what effect observed mountain cedar pollen count had on Google search activity. Due to seasonal availability of mountain cedar pollen measurements, the data were smoothed by computing an 8-week moving average. The Box-Jenkins approach was used to model data for comparison, to show “goodness of fit” and to assess for significance. The ARIMA modeling technique was applied only to the data...
acquired from SAT because of low search activity levels in the DFW region and undetectable levels in HOU.

Results

Visual analysis showed a strong temporal relationship between the cedar pollen counts in SAT and annual search activity for the terms mountain cedar and pollen in January and February (Figure 2). Seasonal maximum counts in SAT ranged from 5413 grains per cubic centimeter to 9982 during the periods studied. A similar temporal relationship was seen weakly in DFW (Figure 3), where counts are much lower than observed counts in SAT, ranging from 635 to 3844 grains per cubic centimeter. Time series plots based on data recorded from HOU (Figure 4) did not show a temporal relationship between searches and pollen count. Counts in HOU were the lowest of any region studied and ranged between 416 and 1265 grains per cubic centimeter.

There was a strong, positive, and statistically significant estimate for the effect of mountain cedar pollen count on search activity for both pollen ($P = .001$) and mountain cedar ($P < .001$). The effect estimates generated from the analysis are 0.012 for pollen and 0.016 for mountain cedar. This finding means that a 1-unit increase in the mountain cedar pollen count increases Google search activity for pollen by 0.012 units and for mountain cedar by 0.016 units. During some of the spikes, pollen counts reach into the thousands, which translates to a large change in search activity. It is also worth noting that there seems to be some lagged effect of pollen count on the pollen search activity. That is, after 2 weeks of elevated pollen levels, there is an additional contribution to the effect on search activity.

Data for the DFW and HOU local areas were not analyzed by ARIMA, due to the scarcity of search activity. In DFW, searches for pollen occurred more frequently than those for mountain cedar; however, activity levels for mountain cedar...
could be reported only monthly. While it is possible to complete analysis with these data, the ability to hone in on any meaningful response in search activity would be severely limited by the excessive effect of data smoothing. HOU search activity for mountain cedar remained 0 throughout the period studied. Here modeling does not add value to the analysis and was not performed.

**Discussion**

Mountain cedar is an established cause of rhinoconjunctivitis in South Texas and has proven to be useful for studying Aeroallergen effects and treatments. Since mountain cedar pollinates in January when most other flora are dormant, seasonal allergic symptoms in the areas studied can be isolated to this individual allergen. This model was particularly useful in this study since 3 of the state’s major cities observe very different mountain cedar pollen levels. This effect was borne out in the search data collected in the cities studied.

Of the 3 areas for which data are evaluated, SAT experiences by far the highest levels of mountain cedar. Mountain cedar is a robust cause of rhinoconjunctivitis in the winter months in South Texas, including the major city of SAT. In a study conducted in 2007, 43% of patients who underwent skin testing for known atopy in SAT exhibited a reaction to the mountain cedar allergen. This fact demonstrates the large number of sensitized individuals living in the region who may be affected by exposure to mountain cedar pollen levels. This effect was borne out in the search data collected in the cities studied.

Of the 3 areas for which data are evaluated, SAT experiences by far the highest levels of mountain cedar. Mountain cedar is a robust cause of rhinoconjunctivitis in the winter months in South Texas, including the major city of SAT. In a study conducted in 2007, 43% of patients who underwent skin testing for known atopy in SAT exhibited a reaction to the mountain cedar allergen. This fact demonstrates the large number of sensitized individuals living in the region who may be affected by exposure to mountain cedar pollen levels. This effect was borne out in the search data collected in the cities studied.

**Figure 4.** Time series plot for Dallas–Ft Worth. Actual pollen counts are denoted for the seasonal maximum observed count in pollen per cubic centimeter.

In HOU, it seems to be relatively unaffected by mountain cedar. While there are visible peaks in the mountain cedar pollen counts in the local area, the absolute counts are much smaller relative to SAT. These low-level counts do not appear to exert an effect on the Google search trends for the observed terms *mountain cedar* and the *pollen*. In fact, the searches for *pollen* in HOU are at annual low during the January/February period. The search activity for *mountain cedar* was 0 throughout the duration of the period observed.

As depicted in **Figure 1**, mountain cedar is not native to the HOU area. Thus, pollen reaches the area only when weather conditions blow pollen grains from central Texas. The analysis of the data collected for the local HOU region elucidates 2 interesting findings. First, these data serve as a negative control, helping to validate mountain cedar pollen as a robust aeroallergen model for an effect on Internet search activity. Low observed mountain cedar pollen counts correlate with reduced or absent levels of Internet search activity during the cedar season. Second, while there are very low levels of mountain cedar observed, there was no increase in search activity for *mountain cedar* or *pollen*. This suggests that there is a threshold for clinical significance that mountain cedar pollen counts do not reach in HOU. Although it would be nice to demonstrate the threshold, quantitative pollen thresholds have been difficult to establish for mountain cedar and seem to have been somewhat arbitrarily assigned in therapeutic trials based on range of symptom response.

This study demonstrates the impact of a single aeroallergen on a geographically distinct population’s Internet search behavior. Previous studies assessing perceived risk in association with Internet search activity have reported correlation. An increased perceived risk surrounding a medical problem, whether infectious or noninfectious, may lead to increased seeking and health care resource consumption. Internet search activity represents patients’ perceptions of health risk, perceived or real need for health care, and in some instances actual burden of disease.
demonstrates that presence of pollen in a geographic region does not necessarily drive related Internet searches. However, when a threshold is surpassed, a population’s perception of health risk drives search activity increases, as seen in DFW and SAT data sets. These data might be used to target periods when web-based patient education materials may be more useful to the public. Furthermore, it gives insight into how the general public goes about searching for such information. It should also lead the practitioner to have a heightened awareness of the potential for exacerbation of related comorbid conditions, such as asthma and obstructive sleep apnea.

When large data sets are utilized as in this study, the results should be taken with caution. Assigning true epidemiologic value to these data is beyond the scope of the study and should be considered the next step. DFW data were not modeled; they were available only as a monthly figure and thus would not have provided meaningful comparison. In spite of these limitations, the ability to trend an effect produced by actual pollen counts is not completely lost.

**Conclusions**

Mountain cedar pollen is a highly allergenic aeroallergen. Environmental levels of mountain cedar pollen show a temporal relationship between Internet searches for pollen and mountain cedar using Google Trends. Furthermore, the intensity of search activity differs among 3 unique geographic regions in which the level of airborne mountain cedar pollen differs. This suggests that there may be a threshold level of mountain cedar pollen required to initiate the observed effects. Google Trends data are useful to understand a patient population’s understanding of health risk presented by mountain cedar pollen. This understanding affords practitioners the opportunity to offer targeted patient education in a geographically unique region, as well as adjust surveillance of related pathologies.

**Acknowledgments**

Special thanks to Julie Wilson for her artistic work, Dr Jeffrey Adelglass, Robert Gomez, and the City of Houston for providing open access to its archived pollen records for this study.

**Author Contributions**

Thomas J. Willson, data collection, analysis and interpretation, manuscript preparation/review and revision, final approval, accountability for all aspects of work; Alexandra Shams, data collection, data analysis, manuscript review/preparation, final approval, accountability for all aspects of work; Joshua Lospinoso, data analysis, manuscript review/preparation, final approval, accountability for all aspects of work; Erik Weitzel, data interpretation, manuscript preparation and review, final approval, accountability for all aspects of work; Kevin McMains, data interpretation, manuscript preparation and review, final approval, accountability for all aspects of work.

**Disclosures**

**Competing interests:** None.

**Sponsorships:** None.

**Funding source:** None.

**References**