



# Using Google Search Data for Sunburn as a Surrogate for Sunburn Prevalence in the United States

## ABSTRACT

**Objective:** The investigators sought to evaluate the feasibility of using state-based Google Trends® search volume data for sunburn as a surrogate marker for state sunburn prevalence. **Design:** State-based search volumes for sunburn were assessed for associations with environmental risk factors for ultraviolet (UV) exposure. **Setting:** Search volume data for sunburn were queried from google.trends.com for all United States (US) searches from January 2004 to December 2017. UV exposure data came from publicly available databases. **Participants:** This analysis included searches occurring in the US. **Main Outcomes and Measures:** Risk factors for UV exposure included degrees latitude, annual number of clear days, average annual temperature, mean state elevation, number of low/moderate/high/very high/extreme UV index days, state outdoor recreation tax revenue, and state consumer spending on outdoor recreation. Regressions and correlations between state searches for sunburn and risk factors for UV exposure were assessed using linear regression and Pearson correlations. **Results:** Searches for sunburn were significantly associated with state degree latitude (coef= -0.59,  $r=-0.47$ ,  $p=0.001$ ); number of low UV index days (coef= -0.37,  $r=-0.46$ ,  $p=0.001$ ); moderate UV index days (coef=1.46,  $r=0.36$ ,  $p=0.01$ ); high UV index days (coef=0.30,  $r=0.43$ ,  $p=0.002$ ); and average annual temperature (coef=0.37,  $r=0.45$ ,  $p=0.001$ ). **Conclusion:** Searches for sunburn in the United States are directly correlated with certain UV exposure measures. These data suggest that search volume for sunburn may be used as a surrogate marker for state sunburn prevalence. **KEYWORDS:** General dermatology, epidemiology, public health, sunburn, melanoma, prevalence, Google Trends

by ZACHARY H. HOPKINS, MD; RYAN CARLISLE, BS; and AARON M. SECREST, MD, PhD

*Dr. Hopkins is with the Department of Internal Medicine, Mr. Carlisle is with the School of Medicine, and Dr. Secrest is with the Departments of Dermatology and Population Health Sciences—all with the University of Utah in Salt Lake City, Utah.*

*J Clin Aesthet Dermatol.* 2019;12(2):32–36

Sunburn signifies an intermittent pattern of exposure to ultraviolet (UV) radiation and is an important risk factor for melanoma.<sup>1</sup> Quantifying the prevalence of sunburn, especially state-based, in the United States (US) is difficult. Currently, the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS) are the largest US surveys for assessing sunburn occurrence, but neither offer consistent, granular state-based data to determine sunburn prevalence by state. For example, the BRFSS assessed sunburn prevalence for the years 1999, 2003, 2004, 2010, and 2012.<sup>2</sup> However, in 2010 and 2012, data regarding sunburn was gathered from only four and six states, respectively. By quantifying each state's sunburn prevalence, hotspots of increased prevalence can be identified and educational campaigns can be tailored to local needs.

Google Trends® (GT) has the ability to track public interest in a topic at the state level and has emerged as a useful technology in epidemiology. GT generates data called “search volume indices” (SVIs), which are standardized measures of searches for each search term in a certain geographical location over a specified time period. Understanding trends

in public search interests can provide valuable epidemiological insights. However, attempts to extrapolate search data to behavior patterns requires validation with outside measures. This principle was applied to the H1N1 influenza outbreaks and has been explored for tanning bed use.<sup>3</sup>

Recently, the association between sunburn searches on GT and UV index (UVI) within each state was explored.<sup>4</sup> While this study demonstrated a strong association between UVI and sunburn searches, the results were not directly comparable between states, as data for each state were pulled separately. GT data is normalized to a 0- to 100-points scale, with 100 points being the highest search for that search query and the rest of the searches being scaled to it. Thus, when each search is pulled independently, the scale is different for each state. As a result, the relationships reported are accurate within each state but cannot be compared between states. To remedy this, sunburn searches must be queried at the national level, with all search data normalized to the state with the highest search volumes (100 points) for a given time period and the other state search volumes scaled relative to that state.

**FUNDING:** Dr. Secrest is supported by research grants from the American Skin Association, National Eczema Association, and National Psoriasis Foundation; however, no specific funding was provided for this project.

**DISCLOSURES:** Dr. Secrest is on the advisory board of VisualDx (Rochester, New York), developer of a diagnostic clinical decision support database system. The other authors report no conflicts of interest relevant to the content of this article.

**CORRESPONDENCE:** Aaron M. Secrest, MD, PhD; Email: aaron.secrest@hsc.utah.edu

Also, while the relationship to UVI and sunburn search interest is important in validating this surrogate prevalence measure, UVI is not the only factor contributing to sunburn. For example, temperature can affect sunburn because temperatures that are too high might encourage people to seek shelter, whereas high or moderate UVIs with more moderate temperatures might encourage longer exposure times.<sup>5,6</sup> Also, outdoor recreation interests of the population might affect sunburn. States with less outdoor recreation might, in theory, relay less risk due to less exposure.

In this analysis, GT data were used to explore whether state-based SVIs for sunburn are associated with state-based measures of UV radiation exposure, both overall and temporally. In addition to UVI, a variety of environmental and behavioral variables were used.

## METHODS

Search volume data for the term *sunburn* were downloaded from trends.google.com by state from January 1, 2004, to December 31, 2017. GT generates SVIs by sampling the number of searches performed in a specified geographic location and time frame—adjusting for total searches. For state-based data, similar outputs are generated, but the SVI for each state represents the averaged search volume within each state. As such, all state SVIs are scaled to the state with the highest number of searches (SVI=100 points). To assess temporality patterns in state data, we extracted state-based data for one winter (10/01/2016–3/31/2017) and one summer (4/1/2017–9/31/2018).<sup>4</sup> With state-based data, it is currently not possible to select for multiple discrete time periods (e.g., data for each winter from 2004–2017), so one winter and one summer were retrieved for comparative purposes.

Proxy measures of sun exposure were divided into environmental and behavioral risk factors, although some overlap existed. Environmental factors included state median degrees latitude, mean state elevation, mean annual number of clear days, average annual temperature, and annual number of days at various UVI levels.<sup>7–10</sup> The UVI scale, as defined by the National Oceanic and Atmospheric Administration is as follows: low (0–2 points); moderate (3–5 points); high (6–7 points); very high (8–10 points); and extreme (>11 points).<sup>7</sup> Behavioral factors included per-capita state tax revenue

**TABLE 1.** Univariable linear regressions for associations between searches for sunburn (sunburn search volume) and predictors of sunburn

VARIABLE	COEFF†	95% CI	P-VALUE
Latitude	-0.59	(-0.91, -0.26)	0.001
State region			
Northeast	Reference	Reference	-
Southeast	8.13	(2.94, 13.3)	0.003
Midwest	2.61	(-2.68, 7.90)	0.33
Southwest	0.36	(-7.04, 7.76)	0.92
West	-0.14	(-5.67, 5.40)	0.96
Mean state elevation	-0.0005	(-0.002, 0.001)	0.41
Number of days annually in specific UVI ranges			
Low	-0.24	(-0.37, 0.10)	0.001
Moderate	1.46	(0.37, 2.55)	0.01
High	0.30	(0.02, 0.48)	0.002
Very high	0.16	(-0.03, 0.36)	0.10
Extremely high	0.34	(-0.31, 0.99)	0.30
Number of clear days	0.03	(-0.04, 0.10)	0.44
Average temperature	0.37	(0.16, 0.59)	0.001
State tax revenue from outdoor activities (per capita)	0.003	(-0.01, 0.02)	0.64
State consumer spending on outdoor activities (per capita)	-0.0005	(-0.002, 0.001)	0.64
BRFSS-reported prevalence of sunburn (2004)	-0.09	(-0.38, 0.20)	0.53

BRFSS: Behavioral Risk Factor Surveillance System; CI: confidence interval; Coef: correlation coefficient; UVI: ultraviolet index  
†Regression coefficient

**TABLE 2.** Pearson correlations between sunburn search volume and factors predicting ultraviolet (UV) exposure. Since Hawaii was an outlier, it was removed for these analyses. Since categorical variables are not allowed within Pearson correlation, state region was not included.

VARIABLE	COEF†	P-VALUE
Latitude	-0.47	0.001
Mean state elevation	-0.12	0.41
Number of days annually in specific UVI ranges		
Low	-0.46	0.001
Moderate	0.36	0.01
High	0.43	0.002
Very high	0.23	0.10
Extremely high	0.15	0.30
Number of clear days	0.11	0.44
Average temperature	0.45	0.001
State tax revenue from outdoor activities (per capita)	0.07	0.64
State consumer spending on outdoor activities (per capita)	-0.14	0.35
BRFSS-reported prevalence of sunburn (2004)	-0.09	0.53

BRFSS: Behavioral Risk Factor Surveillance System; r: Pearson correlation coefficient; UVI: ultraviolet index  
†Pearson correlation coefficient

generated from outdoor activities, per-capita consumer spending on outdoor activities, and mean annual temperature.<sup>11</sup> Temperature was included as an environmental variable but could arguably be behavioral as well, owing to studies demonstrating its effect on a population's likelihood to be outdoors and exposed to UV light.<sup>5,6</sup> Raw state-based monetary measures were divided by each state's 2010 population to estimate per-capita consumer spending and tax revenue.

For each set of SVIs (overall, summer, winter), the mean, standard deviation (SD), and range were calculated. Means of these three datasets were compared using analysis of variance. To quantify how much variability the season added to the change in SVI between the summer and winter datasets as well if these changes were uniform across the country, multivariable linear regression of SVI was performed with season and geographic region as predictors.

Regions were defined according to the census bureau, with certain regions being split into smaller areas to better match weather patterns. Regions and their respective states include the following: Northeast (Maine, Massachusetts, Rhode Island, Connecticut, New Hampshire, Vermont, New York, Pennsylvania, New Jersey, Delaware, and Maryland); Southeast (Washington DC, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Alabama, Mississippi, Arkansas, Louisiana, and Florida); Midwest (Ohio, Indiana, Michigan, Illinois, Missouri, Wisconsin, Minnesota, Iowa, Kansas, Nebraska, South Dakota, and North Dakota); Southwest (Texas, Oklahoma, New Mexico, and Arizona); and West (Colorado, Wyoming, Montana, Idaho, Washington, Oregon, Utah, Nevada, California, Alaska, and Hawaii).<sup>12</sup>

Univariable linear regression was performed to evaluate relationships between sunburn searches and environmental and behavior risk factors. Analysis of initial residuals versus plotted values revealed that the model did not meet the assumption of linearity. The SVI for sunburn searches for Hawaii was a large outlier and was hypothesized to be a contributor to this problem. Since Hawaii is a mathematical outlier as well as a geographic, environmental, and demographic outlier (plausibly, most searches are being performed by tourists), Hawaii was removed from the dataset and all regressions then met the assumption of linearity. A

multivariable linear regression model was not pursued due to limitations in sample size ( $N=51$ ).<sup>13</sup>

To assess relationships between sunburn searches and predictors of UV exposure, Pearson correlation was performed. As Pearson correlations are highly sensitive to outliers and the data were otherwise rather normal in their distribution, Hawaii was removed for these analyses as well.

Unfortunately, since state-based BRFSS data on sunburn was only available for 1999, 2002, and 2004 and GT data started from 2004, comparisons using BRFSS data were limited. Sunburn prevalence data from the year 2004 was compared with the corresponding GT data using univariable linear regression and Pearson correlations. Notably, the time spans of the two datasets differ (GT data are an average of sunburn search data from 2004–2017). Thus, conclusions should be considered with caution.

Statistical significance was established as  $p<0.05$ . All statistics were two-sided and were conducted using STATA version 14.2 (StataCorp. LLC, College Station, Texas).

## RESULTS

State-based sunburn searches were split into three groups: sunburn overall, sunburn summer, and sunburn winter. SVIs for sunburn overall ranged from 19 points (Alaska) to 100 points (Hawaii), with a mean SVI of 38.6 points (SD: 11.1 points). SVIs for sunburn summer ranged from 16 points (Alaska) to 100 points (Hawaii), with a mean SVI of 40.8 points (SD: 12.1 points). SVIs for sunburn winter ranged from three points (Rhode Island) to 100 points (Hawaii), with a mean SVI of 11.7 points (SD: 13.3 points). Because the highest SVIs for all three query outputs were for the same state (Hawaii), the scale was the same and comparisons between the three datasets were possible. There was no difference between the sunburn overall mean and the sunburn summer mean ( $p=0.36$ ), but the means of both the sunburn overall and sunburn summer were greater than those for winter ( $p<0.001$ ). Figures 1A to 1C demonstrate the average SVIs for each state of overall, during summer, and during winter, respectively. Using SVI data from the winter and summer sunburn datasets, linear regression suggested that season (summer vs. winter) explained 57 percent of the variability in SVI ( $r^2=0.57$ ,  $p<0.001$ ). While a regression model could not

be adjusted at the state level, adjusting for US region did not significantly alter the overall fit of the model ( $r^2=0.59$  vs.  $0.57$ ) and did not significantly contribute to the model ( $p=0.54$ ). Pearson pairwise correlation between the three datasets revealed that overall sunburn searches were more correlated with the summer ( $r=0.94$ ,  $p<0.001$ ) rather than winter searches ( $r=0.83$ ,  $p<0.001$ ). Summer and winter sunburn searches were the least related ( $r=0.73$ ,  $p<0.001$ ).

GT also generates top related search terms associated with the search terms queried. For sunburn searches, the top five related searches were as follows: 1) "sunburn peeling," 2) "get rid of sunburn," 3) "sunburn remedies," 4) "how to get rid of sunburn," and 5) "sunburn relief."

Relationships between overall sunburn SVI and sun-exposure risks using univariable linear regression are shown in Table 1. Factors associated with sunburn included latitude (regression coefficient [coef]=  $-0.59$ ,  $p=0.001$ ), Southeast states (coef= $8.13$ ,  $p=0.003$ ), mean annual temperature (coef= $0.37$ ,  $p=0.001$ ), mean annual number of low UVI days (coef= $-0.24$ ,  $p=0.001$ ), moderate UVI days (coef= $1.46$ ,  $p=0.01$ ), and high UVI days (coef= $0.30$ ,  $p=0.002$ ). Mean state elevation, outdoor activity consumer spending, and annual number of very high and extremely high UVI days were not associated with sunburn searches.

Likewise, the strength of associations between sunburn searches and predictors of UV exposure were evaluated using Pearson correlation (Table 2). Factors correlated with sunburn searches included: latitude ( $r=-0.47$ ,  $p=0.001$ ); mean annual temperature ( $r=0.45$ ,  $p=0.001$ ); and low, moderate, and high numbers of annual UVI days ( $r=-0.46$ ,  $p=0.001$ ;  $r=0.36$ ,  $p=0.01$ ; and  $r=0.43$ ,  $p=0.002$ , respectively). All other factors were not significant.

Both linear regression and Pearson correlations demonstrated no significant relationship between the number of people reporting sunburn in 2004 and averaged search data for sunburn.

## DISCUSSION

During the H1N1 influenza outbreak, GT helped estimate influenza incidence and identify areas with emerging cases, allowing for pointed public health intervention in these areas.<sup>3</sup> Although this model ultimately developed errors that caused it to fail, the concept has since

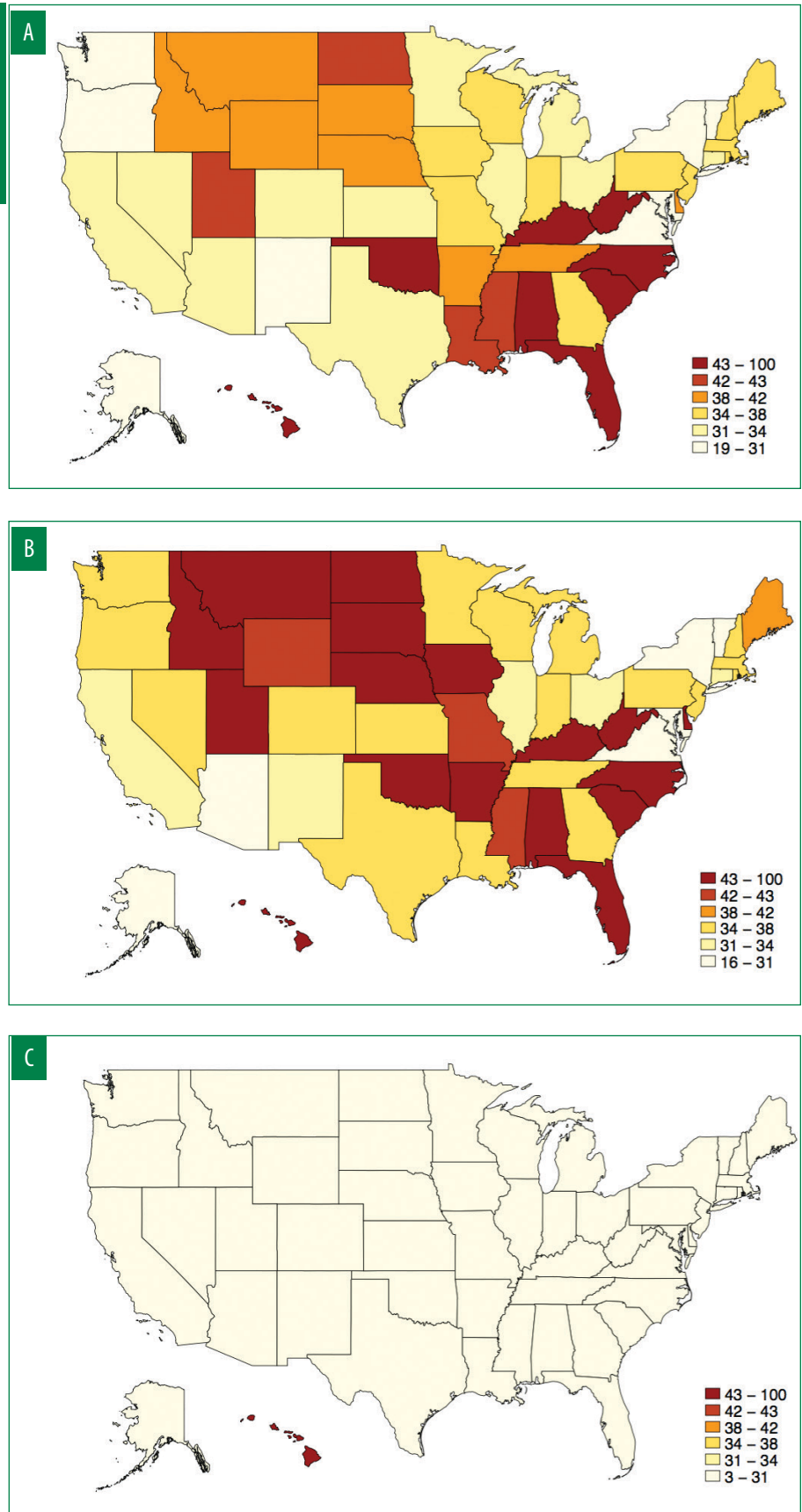
**FIGURE 1.** Heat maps of the United States, with darker states having higher levels of searches for sunburn—A) overall searches from 2004 to 2017; B) searches from April 1, 2017, to September 30, 2017, to represent search volumes for sunburn during the summer months; C) searches from October 1, 2016, to March 31, 2017, to represent search volumes for sunburn during the winter months

triggered numerous studies in different fields.<sup>14</sup> Within dermatology, GT has been used to track US interest in tanning beds, and these interests aligned well with national survey data.<sup>15,16</sup>

Unfortunately, tracking sunburn prevalence is particularly challenging. Nationally, BRFSS and NHIS surveys estimated sunburn prevalence at 33.4 percent (2004), 34.4 percent (2005), 29.1 percent (2010), 34.3 percent (2012), and 31.6 percent (2015), but granularity is a major limitation. State-based sunburn survey data from BRFSS are currently only available for 2004, but, ideally, state-based estimates of sunburn prevalence for multiple years are needed to evaluate trends, identify struggling areas, and guide focused educational campaigns. Thus, we explored associations between sunburn searches and state-level environmental and behavioral factors that act as surrogates for sunburn prevalence in an attempt to validate using GT search data as surrogates for sunburn prevalence.

The seasonality of sunburn searches has been described previously.<sup>4</sup> However, state-based GT data are averaged over the entire time-course designated. To delineate how state-based sunburn searches changed by season, we designated two six-month intervals for data to reflect one summer and one winter. No difference existed in mean SVI overall versus summer, but winter searches were significantly fewer in number (Figures 1A–1C). As expected, linear regression suggested that season accounted for most of the variability between the two seasonal datasets (57%). Adding geographic region to the model did not significantly alter the model. We therefore felt the overall dataset spanning 2007 to 2017 would offer the best picture of sunburn searches by state.

Qualitatively, GT also reports the top searches associated with the keyword queried. The top five searches related to sunburn centered around sunburn-related behaviors. Searching for “severe sunburn relief” suggests the searcher or someone they know has experienced a severe





sunburn. While not directly testable, these associated searches lend support to the idea that sunburn searches are performed for reasons related to being sunburned (prevalence).

Additionally, sunburn searches by state match well with risk factors for sunburn. States at higher latitudes and with greater annual percentages of low UVI days had fewer sunburn searches. Conversely, states with higher annual temperatures and more moderate/high UVI days had more sunburn searches. Interestingly, no correlation was seen between states with more very high or extreme UVI days, perhaps suggesting exposure to these levels is less tolerable or people are more aware of the need for protection. Indeed, moderate UVI days were association in regression analysis with the greatest increase in sunburn searches, similar to other reports.<sup>4</sup>

The lack of significance with behavioral risk factors is expected, as these measures are rather removed from actual behaviors and these data lack granularity on type of outdoor activity. UV exposure can vary greatly by outdoor activity. Higher average temperatures, which are associated with more sunburn searches, typically correlate with more people in these locations spending more time outdoors, often with less protective clothing.<sup>5</sup> Additionally, each state's behavioral and/or cultural features might be better analyzed in a state-by-state fashion. For example, California has a relatively high UVI, yet searches for sunburn were lower in California than in other states, for unclear reasons. However, identifying states at a higher risk as well as identifying exceptions such as these could guide future studies looking at specific states in isolation.

In this analysis, we did not find significant associations between averaged GT search data and BRFSS-reported sunburn prevalences. These data should be evaluated with caution, as BRFSS was only available for one year (2004) and might not adequately account for recent trends in sun exposure or more recent sunburn prevalence calculations.

**Limitations.** GT research has several limitations. GT data are limited to those who can access the Internet, but regular Internet use is exhibited by more than 85 percent of the US population.<sup>17</sup> Additionally, searches are likely

skewed to represent a younger, computer-literate population. Third, GT software is limited in the way in which data are standardized to the highest search volume. With outliers (e.g., Hawaii), other values contract to distort the spread of searches around the median. Lastly, with state-based data, since the data are resampled each time a new database is queried, the direct comparison or compilation of multiple databases is questionably comparable. We advocate that GT researchers should continue contacting Google to request better access to current data or increased functionality of existing software. In this case, since state-based data must be averaged over time periods, it would be useful to be able to select such for certain months of each year and also for the output data to include measures of variance. Furthermore, our claim that searches for sunburn are related to sunburn prevalence would be strengthened if comparisons with state-based survey data or state-based medical data could be done. However, these are data that are currently unavailable in some cases. Despite limitations, these data suggest that online search interest in the term "sunburn" can serve as a surrogate for sunburn prevalence.

## CONCLUSION

Searches for sunburn in the US are correlated with a variety of markers for UV exposure, including latitude, average annual temperature, and number of moderate and high UVI days. These data suggest that sunburn searches can serve as a surrogate marker for sunburn prevalence with further validation. In the meantime, these data can be used for local/regional targeting of specific, online risk factor education campaigns.

## REFERENCES

- Dennis LK, Vanbeek MJ, Beane Freeman LE, et al. Sunburns and risk of cutaneous melanoma: does age matter? A comprehensive meta-analysis. *Ann Epidemiol*. 2008;18(8):614–627.
- Centers for Disease Control and Prevention (CDC). Sunburn prevalence among adults—United States, 1999, 2003, and 2004. *MMWR Morb Mortal Wkly Rep*. 2007;56(21):524–528.
- Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012–1014.
- Lospinoso DJ, Lospinoso JA, Miletta NR. The impact of ultraviolet radiation on sunburn-related search activity. *Dermatol Online J*. 2017;23(8). pii: 13030/qt6cs1n9nd.
- Marion JW, Lee J, Rosenblum JS, Buckley TJ. Assessment of temperature and ultraviolet radiation effects on sunburn incidence at an inland U.S. Beach: a cohort study. *Environ Res*. 2018;161:479–484.
- Diffey BL. Time and place as modifiers of personal UV exposure. *Int J Environ Res Public Health*. 2018;15(6). pii: E1112.
- National Oceanic and Atmospheric Administration site. National Centers for Environmental Information. <https://www.ncdc.noaa.gov>. Accessed 5 Feb 2019.
- Reese R. List of latitudes and longitudes for every state. Ink Plant Code site. <https://inkplant.com/code/state-latitudes-longitudes>. Accessed 4 Feb 2019.
- New World Encyclopedia site. List of U.S. states by elevation. [http://www.newworldencyclopedia.org/entry/List\\_of\\_U.S.\\_states\\_by\\_elevation](http://www.newworldencyclopedia.org/entry/List_of_U.S._states_by_elevation). Accessed 5 Feb 2019.
- Current Results Publishing Ltd site. Average annual temperature for each US state. <https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php>. Accessed 5 Feb 2019.
- Outdoor Industry Association site. Outdoor recreation economy. 2017. <https://outdoorindustry.org/advocacy/>. Accessed 5 Feb 2019.
- United States Census Bureau site. Census regions and divisions of the United States. [https://www2.census.gov/geo/docs/maps-data/maps/reg\\_div.txt](https://www2.census.gov/geo/docs/maps-data/maps/reg_div.txt). Accessed 5 Feb 2019.
- Wilson Van Voorhis CR, Morgan BL. Understanding power and rules of thumb for determining sample sizes. *Tutor Quant Methods Psychol*. 2007;3(2):43–50.
- Lazer D, Kennedy R, King G, Vespignani A. Big data: the parable of Google Flu—traps in big data analysis. *Science*. 2014;343(6176):1203–1205.
- Guy GP Jr, Berkowitz Z, Holman DM, Hartman AM. Recent changes in the prevalence of and factors associated with frequency of indoor tanning among US adults. *JAMA Dermatol*. 2015;151(11):1256–1259.
- Kirchberger MC, Heppt MV, Eigentler TK, et al. The tanning habits and interest in sunscreen of Google users: what happened in 12 years?. *Photodermatol Photoimmunol Photomed*. 2017;33(2):68–74.
- Internet Live Stats site. Number of people with internet in the United States. <http://www.internetlivestats.com/>. Accessed on 14 Feb 2018.

JCAD