



Published in final edited form as:

Mayo Clin Proc. 2018 September ; 93(9): 1185–1190. doi:10.1016/j.mayocp.2018.07.011.

Seasonal and Geographic Patterns in Seeking Cardiovascular Health Information: An Analysis of the Online Search Trends

Nilay Kumar, MD, Ambarish Pandey, MD, Neetika Garg, MD, Emmanuel Sampene, PhD, Carl J. Lavie, MD

Division of Hospital Medicine (N.K.), Division of Nephrology (N.G.), and Department of Biostatistics and Medical Informatics (E.S.), University of Wisconsin School of Medicine and Public Health, Madison, WI; Division of Cardiovascular Medicine, University of Texas Southwestern Medical Center, Dallas, TX (A.P.); and Division of Cardiology, John Ochsner Heart and Vascular Institute, Ochsner Clinical School, University of Queensland School of Medicine, New Orleans, LA (C.J.L.).

Abstract

Objective: To ascertain whether temporal and geographic interest in seeking cardiovascular disease (CVD) information online follows seasonal and geographic patterns similar to those observed in real-world data.

Methods: We searched Google Trends for popular search terms relating to CVD. Relative search volumes (RSVs) were obtained for the period January 4, 2004, to April 19, 2014, for the United States and Australia. We compared average RSVs by month and season and used cosinor analysis to test for seasonal variation in RSVs. We also assessed correlations between state-level RSVs and CVD burden using an ecological correlational design.

Results: RSVs were 15% higher in the United States and 45% higher in Australia for winter compared with summer ($P < .001$ for difference for both). In the United States, RSVs were 36% higher in February compared with August, while in Australia, RSVs were 75% higher in August compared with January. On cosinor analysis, we found a significant seasonal variability in RSVs, with winter peaks and summer troughs for both the United States and Australia ($P < .001$ for zero amplitude test for both). We found a significant correlation between state-level RSVs and mortality from CVD ($r = 0.62$; $P < .001$), heart disease ($r = 0.58$; $P < .001$), coronary heart disease ($r = 0.48$; $P < .001$), heart failure ($r = 0.51$; $P < .001$), and stroke ($r = 0.60$; $P < .001$).

Conclusion: Google search query volumes related to CVD follow strong seasonal patterns with winter peaks and summer troughs. There is moderate to strong positive correlation between state-level search query volumes and burden of CVD mortality.

Cardiovascular disease (CVD) is a prevalent and growing public health problem in the United States and globally.^{1,2} Seasonal variability in CVD incidence has been described

Correspondence: Address to Carl J. Lavie, MD, John Ochsner Heart and Vascular Institute, Ochsner Clinical School-The University of Queensland School of Medicine, 1514 Jefferson Highway, New Orleans, LA 70121-2483 (clavie@ochsner.org).

SUPPLEMENTAL ONLINE MATERIAL Supplemental material can be found online at <http://www.mayoclinicproceedings.org>. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Potential Competing Interests: The authors report no competing interests.

extensively.^{3,4} A winter excess in the incidence and mortality related to coronary heart disease (CHD),⁵ acute myocardial infarction,^{6,7} heart failure (HF),^{8,9} arrhythmias, and sudden cardiac death^{6,10,11} has been reported previously. In addition to temporal variability, there is significant geographic variability in CVD prevalence and mortality.^{12,13}

Current data on temporal and geographic variation in CVD are largely derived from community-based surveillance studies. Most such studies employ a retrospective observational design that is dependent on diagnostic codes for case finding that are then validated per the study protocol.¹⁴ Large retrospective hospital databases such as the National Inpatient Sample and surveys such as the National Health and Nutritional Examination Survey and the Behavioral Risk Factor Surveillance System are resource intensive and have a 2- to 3-year lag period in data availability. This lack of real-time data is a major impediment to effective CVD surveillance in the community. In the current era, passively generated digital data could potentially be leveraged to estimate disease activity in the community until validated data from traditional sources become available.

The past decade has witnessed an exponential increase in Internet penetration, with rates in many nations in the developed world now exceeding 80%. For example, nearly 80% of the US population now owns a smartphone. In 2012, there were 2.4 billion Internet users worldwide generating 2.5 quintillion (2.5×10^{18}) bytes of data every day,¹⁵ and these passively generated data provide unprecedented opportunities to study disease patterns at the community level by tapping into collective population interests and behaviors. Digital footprints left by online users could potentially serve as proxies for disease activity at a community level, thereby providing invaluable insights into temporal and spatial trends in diseases.¹⁶ In a preliminary analysis, we found that hypertension related search query volumes follow seasonal trends similar to real-world observations in the United States and Australia.¹⁷ In the current study, we analyzed data from Google Trends to ascertain whether search query volumes related to CVD show temporal and geographic variation that is parallel to real-world observations.

METHODS

Search Strategy

We searched Google Trends for search terms related to CVD in the areas of CHD, acute coronary syndrome, arrhythmias, HF, hypertension, and diabetes mellitus.¹⁸ Search query volume was filtered by the term *health* using the Google query category feature to discard non-health-related queries that could confound results. We selected a total of 28 search query terms for the 5 CVD categories (Supplemental Table 1, available online at <http://www.mayoclinicproceedings.org>). To analyze and compare patterns in the Northern and Southern Hemispheres, we used data for the United States and Australia.

Google reports relative search volumes (RSVs) on specific query terms on a normalized scale of 0 to 100. This scale represents the number of times a particular search term has been queried relative to the total number of search queries for a particular time period in a particular geographic location. These numbers do not represent absolute search volumes. Google publicly reports RSVs, which can be searched by country and for any time period

starting from January 2004. We obtained weekly RSVs for the 28 search query terms in our study from the period January 4, 2004, to April 19, 2014. Data were downloaded from Google Trends in April 2014. Google reports weekly RSVs for the United States and weekly/monthly RSVs for Australia. For 17 of the 28 search terms for Australia, only monthly RSVs were available. For search terms for Australia for which only monthly RSVs were reported, weekly RSVs were calculated using linear interpolation.

Correlation Between State-Level Search Volume and Burden of CVD Mortality

We performed an ecological correlational analysis to test whether state-level search volumes were correlated with the burden of CVD mortality in the United States. We obtained age-adjusted estimates of mortality associated with CVD, heart disease, CHD, HF, and stroke per 100,000 persons for the year 2014. These data are collected by the National Vital Statistics System and are publicly available from the Centers for Disease Control and Prevention's National Cardiovascular Disease Surveillance System.¹⁹ Google search terms for correlational analysis are listed in Supplemental Table 2 (available online at <http://www.mayoclinicproceedings.org>).

Statistical Analyses

To test for differences in mean RSVs across seasons and month, we used linear regression with season or month as a categorical predictor. Percentage change in search volume from reference was calculated by using the formula $([RSV - \beta_{\text{intercept}}]/\beta_{\text{intercept}}) \times 100$, where $\beta_{\text{intercept}}$ equals mean search volume for reference month calculated from the linear regression model. The 95% CIs for percentage change were bootstrapped using 1000 random samples.

To test for seasonal variability of RSVs, we fit a cosinor model to the detrended, demeaned time series for search volume converted to percentage difference from the linear trend using the formula $100 \times (RSV - xb)/xb$, where xb is the predicted value of search volume from a linear regression model using time as a continuous predictor. The cosinor model is given by the equation $E(Y) = A \cos(2\pi t/T + \phi)$, where Y is predicted value of RSVs from the model, A is the amplitude of variation from baseline at the peak or half the difference between the peak and trough, t is time in weeks since the start of the study, T is 52 weeks since we were testing for annual variability, and ϕ is the acrophase or timing of the peak.²⁰

The correlation between state-level RSVs and burden of CVD mortality was tested using Pearson product moment correlation coefficients. All analyses were conducted using the Stata statistical software, version MP 13.0 (StataCorp) and R statistical software, version 3.1.2 (R Project for Statistical Computing). Analyses were performed in 2017.

RESULTS

Differences in Search Volumes by Season and Month

In the United States, we found mean RSV to be significantly higher for January to June and September to November compared with the month of August ($P < .001$ for January-April and September-November compared with August, $P = .01$ for May vs August, $P = .02$ for June vs

August) In the US, mean RSV was 36% higher in February compared with August ($P<.001$). Similarly, mean RSV was 75% higher in August compared with January in Australia ($P<.001$). (Table 1). When analyzed by season, average search volumes were significantly higher in the spring, winter, and fall compared with summer for both the United States and Australia ($P<.001$ for winter, spring and fall vs summer in the US, $P<.001$ for fall and winter vs summer in Australia, $P=.02$ for spring vs summer in Australia) In the US, mean RSV was 15% higher in the winter compared with summer ($P<.001$). Similarly in Australia, mean RSV was 45% higher in winter compared with summer ($P<.001$). (Table 2).

Cosinor Analysis for Search Volume

In a cosinor model for the United States, there was a statistically significant seasonal component with amplitude of 7.43 (95% CI, 5.78–9.08; $P<.001$). The US peaks occurred in the month of February and the troughs occurred in August. Because we converted RSV to percent difference from overall mean, this means that in February, search volumes are predicted to be 7.43% higher than the overall mean. Similarly in Australia, a cosinor model revealed a significant seasonal component with amplitude of 23.99 (95% CI, 20.83–27.16; $P<.001$). In Australia, peaks occurred in June and troughs in December (Figure 1).

Association Between State-Level Search Volume and Burden of CVD Mortality

We found a moderate to strong correlation between state-level RSV and burden of CVD mortality. Pearson product moment correlation coefficient was 0.62 for state-level CVD mortality and RSV ($P<.001$), 0.58 for state-level CVD mortality from heart disease and RSV ($P<.001$), 0.48 for state-level mortality from CHD and RSV ($P<.001$), 0.51 for state-level HF mortality and RSV ($P<.001$), and 0.60 for state-level stroke mortality and RSV ($P<.001$) (Figure 2; Supplemental Figures 1–4 [available online at <http://www.mayoclinicproceedings.org>], Supplemental Table 2).

DISCUSSION

In this study using Google search query data from the United States and Australia over a period of 11 years, we made the following interesting and noteworthy observations. First, search volumes related to CVD follow strong seasonal patterns with winter peaks and summer troughs. This seasonal pattern is parallel to previously documented seasonal patterns for CVD incidence and mortality in real-world data. Second, there was a 4-month difference in the timings of the peaks in the 2 countries (February vs June), consistent with differences in seasons in the 2 regions and suggesting a possible causal link between seasons and CVD-related information-seeking behavior mediated by a higher winter incidence of CVD. Third, the magnitude of seasonal variation was higher in Australia than in the United States (amplitude 23.99 vs 7.43). Lastly, we found a statistically significant correlation between state-level burden of CVD mortality and state-level search query volumes in the United States. Taken together, we report for the first time a comprehensive analysis of seasonal and geographic variation in CVD information-seeking behavior in the United States and Australia using robust statistical methods.

Passively generated search query data from Google Trends have been used to evaluate seasonal patterns in health information-seeking behavior for a variety of communicable and noncommunicable diseases.^{21–25} Ayers et al²¹ studied seasonality in mental health-seeking behavior using Google Trends and found search volumes to peak in winter months coincident with a seasonal predisposition to several psychiatric conditions, including seasonal affective disorder. Similarly, another recent analysis of Google search query data revealed marked seasonality in information seeking for weight loss in winter months.²³ Our findings of winter peaks in seeking CVD health information in both the United States and Australia are in agreement with previously reported seasonal patterns of CVD incidence and mortality,^{4–6,10,11} suggesting a possible mechanistic link between the occurrence of CVD events and information-seeking behavior at the population level.

Several recent studies have also examined geographic distribution of search query volumes and disease activity at the population level. In a recent analysis, Nguyen et al²⁶ found a strong correlation between Google search activity at the state level and prevalence of CVD risk factors such as smoking, hypertension, and diabetes mellitus. Willard and Nguyen²⁵ found a statistically significant correlation between the state prevalence of kidney stones and search volume. Our study adds incrementally to the literature by finding a statistically significant correlation between state-specific search volumes and mortality from established CVD.

These findings may have several practical implications. First, a considerable body of research has found that a large proportion of health care information on the Internet and smartphone-based applications is either of poor quality or has poor user engagement.^{27–29} A better understanding of temporal and geographic variation in viewer interest may help with targeted dissemination of valid health care information by authoritative sources, as well as managed health care organizations with interests and incentives of having healthier populations. Second, passively generated search volume data could serve as a proxy for measuring population-level behaviors and interests, which in turn may be positively related to disease activity in the community. Because these data are generated passively and can be monitored in real time, future artificial intelligence algorithms could potentially leverage these data to augment traditional surveillance mechanisms. These data could be especially useful in low- to middle-income countries where rates of Internet penetration and smartphone ownership are growing more rapidly than traditional resources for disease surveillance at the community level.

Several limitations of our analysis should be noted. First, since longitudinal and state-specific search query data, as well as state-specific burden of CVD, are population-level averages, ie, an ecological study design, these data should not be used to infer an association between Internet search activity and CVD risk at the level of an individual. Second, search query data in our study were derived from 2 large developed nations with high rates of Internet penetration and smartphone ownership. Therefore, these findings may not be generalizable to countries where Internet penetration is not sufficiently high to measure population interest in health information. Lastly and most importantly, passively generated “big data” must never be considered to be a substitute for, but rather an adjunct to, validated data measurements by traditional surveillance mechanisms. Future studies should evaluate

the predictive utility of search query data in identifying high-risk geographic pockets of CVD and predicting the incidence of CVD where traditional surveillance data are not available.

CONCLUSION

In this study, we identified a significant seasonal pattern in CVD health-seeking behavior using Google search query data in the United States and Australia. We also found a significant correlation between CVD burden and CVD-related search volumes at the state level in the United States, which has several potential practical implications for dissemination of health care information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Grant Support: This work was supported by the Clinical and Translational Science Award program through grant UL1TR002373 from the National Institutes of Health, National Center for Advancing Translational Sciences.

Abbreviations and Acronyms:

CHD	coronary heart disease
CVD	cardiovascular disease
HF	heart failure
RSV	relative search volume

REFERENCES

1. Go AS, Mozaffarian D, Roger VL, et al.; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics2013 update: a report from the American Heart Association [published corrections appear in *Circulation*. 2013;127(1):doi:10.1161/CIR.0b013e31828124ad and *Circulation*. 2013;127(23):e841]. *Circulation*. 2013;127(1):e6-e245.
2. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010 [published correction appears in *Lancet*. 2013; 381(9874):1276]. *Lancet*. 2012;380(9859):2224–2260. [PubMed: 23582395]
3. Eng H, Mercer JB. Seasonal variations in mortality caused by cardiovascular diseases in Norway and Ireland. *J Cardiovasc Risk*. 1998;5(2):89–95. [PubMed: 9821061]
4. Spencer FA, Goldberg RJ, Becker RC, Gore JM. Seasonal distribution of acute myocardial infarction in the second National Registry of Myocardial Infarction. *J Am Coll Cardiol*. 1998; 31(6):1226–1233. [PubMed: 9581712]
5. Weerasinghe DP, MacIntyre CR, Rubin GL. Seasonality of coronary artery deaths in New South Wales, Australia. *Heart*. 2002;88(1):30–34. [PubMed: 12067938]

6. Gerber Y, Jacobsen SJ, Killian JM, Weston SA, Roger VL. Seasonality and daily weather conditions in relation to myocardial infarction and sudden cardiac death in Olmsted County, Minnesota, 1979 to 2002. *J Am Coll Cardiol*. 2006; 48(2):287–292. [PubMed: 16843177]
7. Crawford VL, McCann M, Stout RW. Changes in seasonal deaths from myocardial infarction. *QJM*. 2003;96(1):45–52. [PubMed: 12509648]
8. Barnett AG, de Looper M, Fraser JF. The seasonality in heart failure deaths and total cardiovascular deaths. *Aus N Z J Public Health*. 2008;32(5):408–413.
9. Qiu H, Yu IT, Tse LA, Tian L, Wang X, Wong TW. Is greater temperature change within a day associated with increased emergency hospital admissions for heart failure? *Circ Heart Fail*. 2013;6(5):930–935. [PubMed: 23935005]
10. Anand K, Aryana A, Cloutier D, et al. Circadian, daily, and seasonal distributions of ventricular tachyarrhythmias in patients with implantable cardioverter-defibrillators. *Am J Cardiol*. 2007; 100(7):1134–1138. [PubMed: 17884377]
11. Müller D, Lampe F, Wegscheider K, Schultheiss HP, Behrens S. Annual distribution of ventricular tachycardias and ventricular fibrillation. *Am Heart J*. 2003;146(6):1061–1065. [PubMed: 14660999]
12. Mensah GA, Mokdad AH, Ford ES, Greenlund KJ, Croft JB. State of disparities in cardiovascular health in the United States. *Circulation*. 2005;111(10):1233–1241. [PubMed: 15769763]
13. Yusuf S, Reddy S, Ôunpuu S, Anand S. Global burden of cardiovascular diseases, part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. *Circulation*. 2001;104(23):2855–2864. [PubMed: 11733407]
14. Roger VL. Cardiovascular disease surveillance in the comparative effectiveness landscape [editorial]. *Circ Cardiovasc Qual Outcomes*. 2009;2(5):404–406. [PubMed: 20031869]
15. Rainie L, Perrin A. 10 Facts about smartphones as the iPhone turns 10. Pew Research Center website, [pewresearch.org](http://www.pewresearch.org), <http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/>. Published June 28, 2017 Accessed December 18, 2017.
16. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? *JAMA*. 2014;311(14): 1399–1400. [PubMed: 24577162]
17. Kumar N, Phipps C, Garg N, Pandey A. Assessing patterns of global interest in hypertension using internet search engine data. *J Am Soc Hypertens*. 2014;8(4):e123.
18. Google Trends website. <http://www.google.com/trends/>. Accessed May 29, 2014.
19. National Cardiovascular Disease Surveillance System. Data trends & maps. Centers for Disease Control and Prevention website <http://www.cdc.gov/DHDSP/ncvdss/index.htm>. Updated July 14, 2017 Accessed May 29, 2014.
20. Barnett AG, Dobson AJ. *Analysing Seasonal Health Data*. Berlin, Germany: Springer Verlag; 2012.
21. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. *Am J Prev Med*. 2013;44(5):520–525. [PubMed: 23597817]
22. Ingram DG, Plante DT. Seasonal trends in restless legs symptomatology: evidence from Internet search query data. *Sleep Med*. 2013;14(12):1364–1368. [PubMed: 24152798]
23. Madden KM. The seasonal periodicity of healthy contemplations about exercise and weight loss: ecological correlational study. *JMIR Public Health Surveill*. 2017;3(4):e92.
24. Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis*. 2011;5(8):e1258.
25. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. *Urology*. 2013;81(1):37–42. [PubMed: 21676450]
26. Nguyen T, Tran T, Luo W, et al. Web search activity data accurately predict population chronic disease risk in the USA. *J Epidemiol Community Health*. 2015;69(7):693–699. [PubMed: 25805603]
27. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. *J Am Soc Hypertens*. 2015;9(2):130–136. [PubMed: 25660364]
28. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing Web sites a useful source of information on hypertension? *J Am Soc Hypertens*. 2014;8(7):481–490. [PubMed: 25064770]

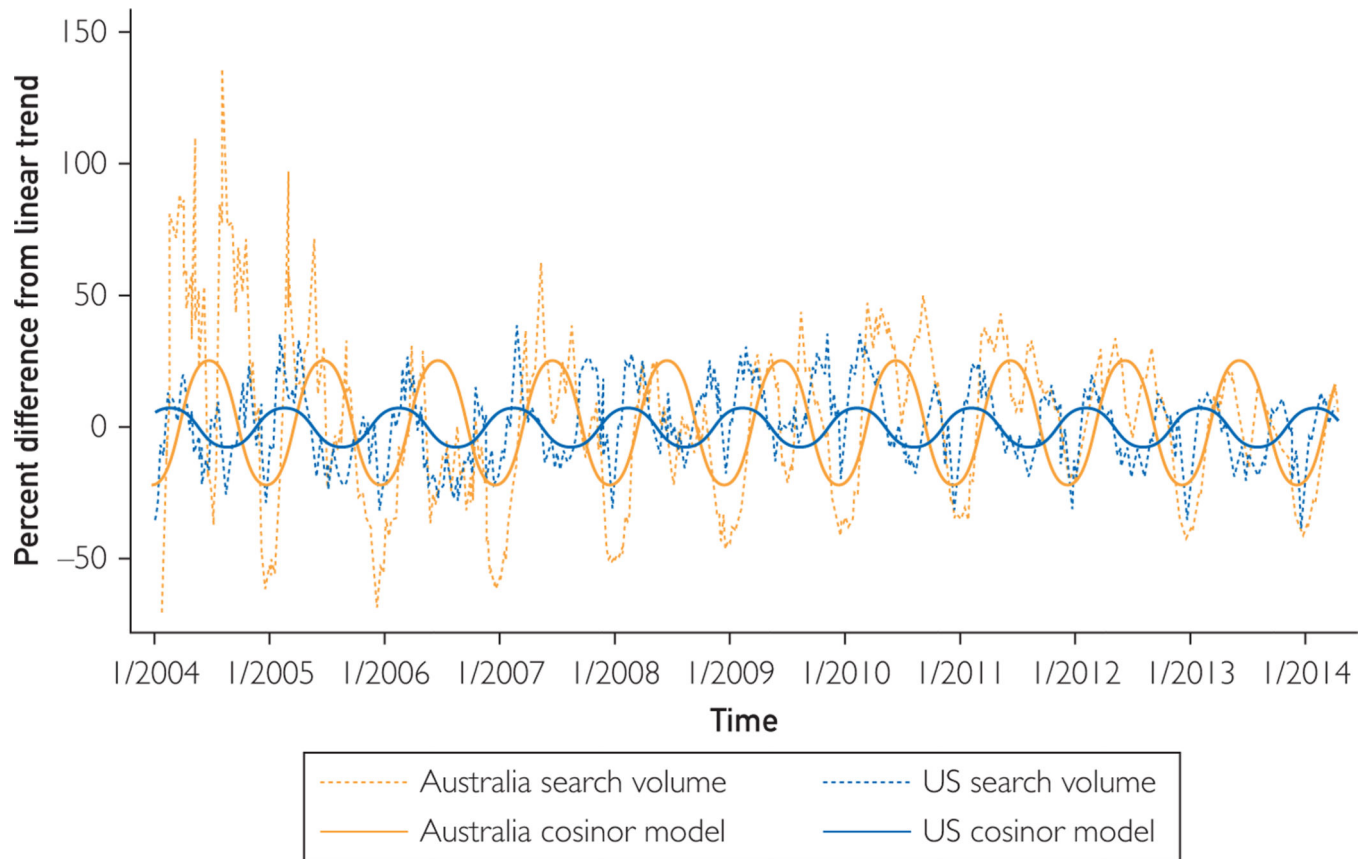
29. Garg N, Venkatraman A, Pandey A, Kumar N. YouTube as a source of information on dialysis: a content analysis. *Nephrology (Carlton)*. 2015;20(5):315–320. [PubMed: 25641264]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Peak month February for the US and June for Australia
 $P < .001$ for zero amplitude test for both US and Australia

FIGURE 1. Time series plot for detrended, demeaned search volumes in the United States (US) and Australia along with cosinor models that were fit to these data.

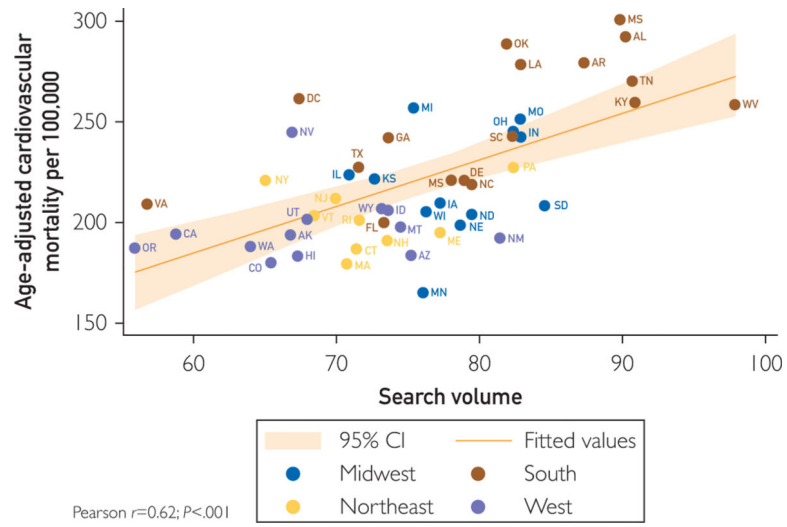


FIGURE 2. Correlation between state-level search volume and age-adjusted cardiovascular mortality per 100,000 persons.

TABLE 1.

Percent Change in Cardiovascular Disease-Related Relative Search Query Volumes Compared With Reference Month

Month	Percent change in US search volume (95% CI)	P value	Percent change in Australian search volume (95% CI)	P value
January	14.83 (7.36–22.29)	<.001	Reference	...
February	36.28 (29.02–43.54)	<.001	24.93 (4.73–45.14)	.02
March	34.56 (28.11–41.02)	<.001	81.66 (56.24–107.08)	<.001
April	27.59 (21.55–33.63)	<.001	73.15 (48.31–97.99)	<.001
May	7.24 (1.67–12.81)	.01	78.82 (52.89–104.75)	<.001
June	7.04 (0.93–13.14)	.02	50.30 (27.22–73.39)	<.001
July	4.35 (0.92 to 9.63)	.11	43.85 (21.95–65.76)	<.001
August	Reference	...	74.79 (49.41–100.17)	<.001
September	22.26 (14.53–30.00)	<.001	70.76 (45.77–95.75)	<.001
October	30.46 (23.36–37.56)	<.001	53.31 (31.28–75.34)	<.001
November	23.25 (15.75–30.77)	<.001	31.40 (12.13–50.67)	.001
December	0.76 (–5.64 to 7.16)	.82	–3.63 (–20.17 to 12.90)	.67

Percent Change in Search Volumes for Winter, Spring, and Fall Compared With Summer in the Northern and Southern Hemispheres

TABLE 2.

Season	Percent change in US search volume (95% CI)	P value	Percent change in Australian search volume (95% CI)	P value
Northern winter/Southern summer (January-March)	15.05 (9.77-20.32)	<.001	Reference	
Northern spring/Southern fall (April-June)	9.96 (6.73-13.18)	<.001	52.68 (41.54-63.82)	<.001
Northern summer/Southern winter (July-September)	Reference	...	45.05 (32.33-57.77)	<.001
Northern fall/Southern spring (Oct e Dec)	8.68 (4.55-12.82)	<.001	13.20 (2.44-23.96)	.02