

RESEARCH ARTICLE

Information Seeking Regarding Tobacco and Lung Cancer: Effects of Seasonality

Zhu Zhang¹, Xiaolong Zheng¹, Daniel Dajun Zeng^{1,2*}, Scott J. Leischow³

1 The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, **2** Department of Management Information Systems, The University of Arizona, Tucson, United States of America, **3** Mayo Clinic, Scottsdale, Arizona, United States of America

* dajun.zeng@ia.ac.cn



OPEN ACCESS

Citation: Zhang Z, Zheng X, Zeng DD, Leischow SJ (2015) Information Seeking Regarding Tobacco and Lung Cancer: Effects of Seasonality. PLoS ONE 10 (3): e0117938. doi:10.1371/journal.pone.0117938

Academic Editor: Tobias Preis, University of Warwick, UNITED KINGDOM

Received: August 12, 2014

Accepted: January 4, 2015

Published: March 17, 2015

Copyright: © 2015 Zhang et al. 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 author and source are credited.

Data Availability Statement: All relevant data is within the paper and its Supporting Information files.

Funding: This research is partially supported by the National Natural Science Foundation of China (<http://www.nsf.gov.cn/>) under Grant No. 71103180 (PI: XZ), 71472175 (PI: XZ), 71025001 (PI: DDZ), 91224008 (PI: DDZ), and 61175040 (PI: WM); National Institutes of Health (NIH) of USA (<http://www.nih.gov/>) under Grant No. 1R01DA037378-01 (PI: DDZ); the Ministry of Health of China (<http://www.nhfpc.gov.cn/>) under Grant No. 2012ZX10004801 (PI: DDZ) and 2013ZX10004218 (PI: DDZ); and by the Grant No. 2013A127 (PI: DDZ). The funders had no

Abstract

This paper conducted one of the first comprehensive international Internet analyses of seasonal patterns in information seeking concerning tobacco and lung cancer. Search query data for the terms “tobacco” and “lung cancer” from January 2004 to January 2014 was collected from Google Trends. The relevant countries included the USA, Canada, the UK, Australia, and China. Two statistical approaches including periodogram and cross-correlation were applied to analyze seasonal patterns in the collected search trends and their associations. For these countries except China, four out of six cross-correlations of seasonal components of the search trends regarding tobacco were above 0.600. For these English-speaking countries, similar patterns existed in the data concerning lung cancer, and all cross-correlations between seasonal components of the search trends regarding tobacco and that regarding lung cancer were also above 0.700. Seasonal patterns widely exist in information seeking concerning tobacco and lung cancer on an international scale. The findings provide a piece of novel Internet-based evidence for the seasonality and health effects of tobacco use.

Introduction

Tobacco is the only legal drug that causes many deaths of its users when used exactly as intended by manufacturers [1]. Tobacco smoking is associated with an increased risk of ill-health, disability, and death from noncommunicable chronic diseases (e.g., lung cancer and cardiovascular disease) as well as communicable diseases such as tuberculosis [1]. Direct tobacco smoking causes about the deaths of 5 million people across the world each year, and many of these deaths occur prematurely [1,2]. Given the serious damage of tobacco smoking, tobacco control is one of the most important global public health priorities.

Exploring temporal patterns (e.g., seasonality) of smoking behavior can significantly help us to make reasonable tobacco control regulations. For instance, a previous research suggests that findings on seasonality of smoking may have major implications for the timing of tobacco control measures and interventions [3]. Several existing empirical studies [3,4,5,6,7] have found

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

that seasonality exists in the cigarette sales [3,6], the sales of nicotine replacement therapies [4], the initiation of smoking among adolescents [5], and the onset of youth smoking [7]. Most of these studies were conducted based on the actual and offline survey data. This kind of data collected by survey-based tobacco control surveillance is costly to obtain, and often lacks timeliness [8,9]. In addition, the survey data in the above previous studies is limited to the region of the USA. To the best of our knowledge, no comprehensive study on seasonality of tobacco use has been conducted on an international scale. The major reason may be that it is difficult to obtain suitable data across different countries for such study.

However, this situation presented above is changing. Recently, with the rapid growth of Internet, there is a large number of online search data about public health (e.g., tobacco and lung cancer etc.), and such search data suggests public interest in public health information. The timely, large-scale, public, and free Internet search query data provides a novel resource for public health researchers to obtain a better understanding of collective human behaviors. For instance, some previous studies used Web search queries to monitor infectious and noninfectious diseases such as influenza epidemics [10,11,12] and other diseases [13,14]. Tobacco control researchers have also used Internet search queries to explore issues such as studying who searched smoking cessation information and how they searched it using Web search engines [15,16], monitoring tax avoidance and smoking cessation after the increase of cigarette tax [17], and tracking the popularity of electronic cigarettes [18]. Additionally, the impact of World No Tobacco Day in Latin America was evaluated by health-related news stories and Internet search queries [8]. Since some previous studies show that the weather condition (especially temperature) may impact on the seasonal variation of smoking behavior [3,6], tobacco and lung cancer information seeking relevant to smoking may also have seasonal patterns. To complement the mentioned offline survey data, the aim of this paper is to exploit Internet search data to monitor continuous population interest in tobacco and lung cancer and study seasonal patterns of information seeking concerning tobacco and lung cancer.

Materials and Methods

Ethics Statement

The raw data used in this study was freely downloaded from Google Trends (<http://www.google.com/trends/>), a public service provided by Google Inc., which allows Internet users to examine trends of certain query terms by time, geographic location, and category. The search query trend data collected from Google Trends is already anonymized. The data doesn't involve human subject research, clinical trials, animal research, and observational and field studies, so we don't need to obtain ethical oversight from an Institutional Review Board (IRB) or from an appropriate data protection agency. Additionally, Google has relevant privacy policies to address ethical issues related to Google Trends. The raw datasets in this study were deposited as [S1](#) and [S2](#) Datasets in supporting information.

Data collection

Due to the above advantages of Internet search queries, Google Trends was used to monitor the incidence of depression [14], smoking cessation [17], the popularity of electronic cigarettes [18], the interest in behavior change [19], and the evaluation of World No Tobacco Day [8]. Recently, it was applied to monitor non-cigarette tobacco use [20]. Besides, another study presented an external revision to Google Flu Trends for improving the prediction accuracy [21]. On Google Trends, we used "tobacco" as the search query term to collect the search trend data about tobacco in the four main English-speaking countries (i.e., the USA, the UK, Canada, and Australia). We also used the Chinese translation of "tobacco" as the query term for the trend

data in China. In a similar way, the keyword “lung cancer” was used to collect the trend data regarding lung cancer (i.e., one of the main smoking-attributable diseases). The query terms used in this study were limited to all categories rather than a specific category like “health” or “shopping” on Google Trends, because the information related to the query term “tobacco” or “lung cancer” can cover several categories on Google Trends. The query terms “tobacco” and “lung cancer” were chosen since they were the most suitable to accurately represent information seekers’ interest in tobacco and lung cancer information, respectively. For instance, the information seeking interest indicated by the search terms “tobacco use” or “tobacco smoking” was narrower than that of the search term “tobacco”. Based on the direct observations on Google Trends, the search trends of the query term “cigarette” were similar to those of the query term “tobacco”. Besides, cigarette is one of tobacco products, and the concept of tobacco use is wider than cigarette smoking. Therefore, the search trends about cigarette were not considered in this article. The selection of the above Chinese query terms followed the same rule. The original data was downloaded on 27 October 2014, and the time interval covered by the data was from January 4, 2004 (the first date when the data was available) to October 25, 2014 (the end date). Then we extracted the ten-year data containing 522 weeks from the beginning to January 4, 2014 for convenience. The methodology to generate Google Trends data has been presented in [11]. Briefly, Google aggregates historical search query logs for chosen terms submitted within a given time frame and region. A relative value for a chosen query term can be calculated as the total search volume for the term divided by the total search volume within the chosen geographic region and time [19]. Then the relative value is normalized to the peak relative value over a given time period with respect to the same term. Finally, all relative query values are from 0 to 100.

Statistical analysis

The following statistical analysis methodology was borrowed from that in our previous work [22], but the data, related analyses, and discussion were considerably extended in this study.

Periodogram (also known as spectral plot) was applied to determine whether seasonal patterns of information seeking concerning tobacco and lung cancer existed in the corresponding search trend data across different countries, respectively. A peak in a periodogram indicated that a seasonal component existed near the frequency value corresponding to the peak. When a periodogram was generated, we applied ideal pass filter on the periodogram to extract a seasonal component from the raw trend data. The frequency interval for the filtering process just covered the peak shape in the periodogram. Ideal pass filter is a data filtering tool offered by the time series tools of MATLAB R2010b. According to the user’s guide, ideal pass filter allows only the variations in a specific frequency range. “These filters are ‘ideal’ in the sense that they are not realizable; an ideal filter is non-causal and the ends of the filter amplitude are perfectly flat in the frequency domain [23].”

Moreover, to examine linear and temporal associations of seasonal components of these trend data among countries, we calculated the pairwise cross-correlations of these seasonal components. In detail, we conducted the correlation analysis to quantify how much these seasonal components were correlated and find the time shifts in the seasonalities among different countries. If there are high cross-correlations between search trends of different countries, there are common temporal patterns in the tobacco or lung cancer information seeking behavior across countries. These common patterns of information seeking may be used in the timing of international tobacco control campaigns. For instance, if there is a high cross-correlation between the tobacco search trends of the USA and Canada, and the time lag is 0, they can collaborate to start an anti-smoking campaign at the peak of tobacco information seeking, because the

information seekers' interest to tobacco information is the highest at that time. Then the trend data concerning lung cancer was also analyzed like the tobacco data. Similarly, we measured the temporal associations between seasonal components of the trend data regarding tobacco and those regarding lung cancer by cross-correlation. Note that cross-correlation was used to measure the degree of the linear relationship between two search trends at various time lags in this study. A high correlation at a specific lag might indicate that these two trends had certain same influencing factors and a time delay [23]. The periodogram, ideal pass filter, and cross-correlation were carried out by the time series tools of MATLAB [23]. For the details of these tools, see the previous study [22] and the MATLAB manual [23].

Results

Fig. 1 showed the time series plots of the raw trend data regarding tobacco (See S1 Fig. for the trends related to lung cancer). Note that all figures in this study were produced by the time series tools of MATLAB [23]. The horizontal axis was the time line with one week as the unit time. Fig. 2 presented the periodograms of the search trend data about tobacco (See S2 Fig. for the periodograms about lung cancer). In the periodograms, the salient peak near 0.0192 on the horizontal line indicated that each search trend had a seasonal component whose period was about one year (i.e., 1 cycle / [0.0192 cycles/week] ≈ 52 weeks). Another salient peak near

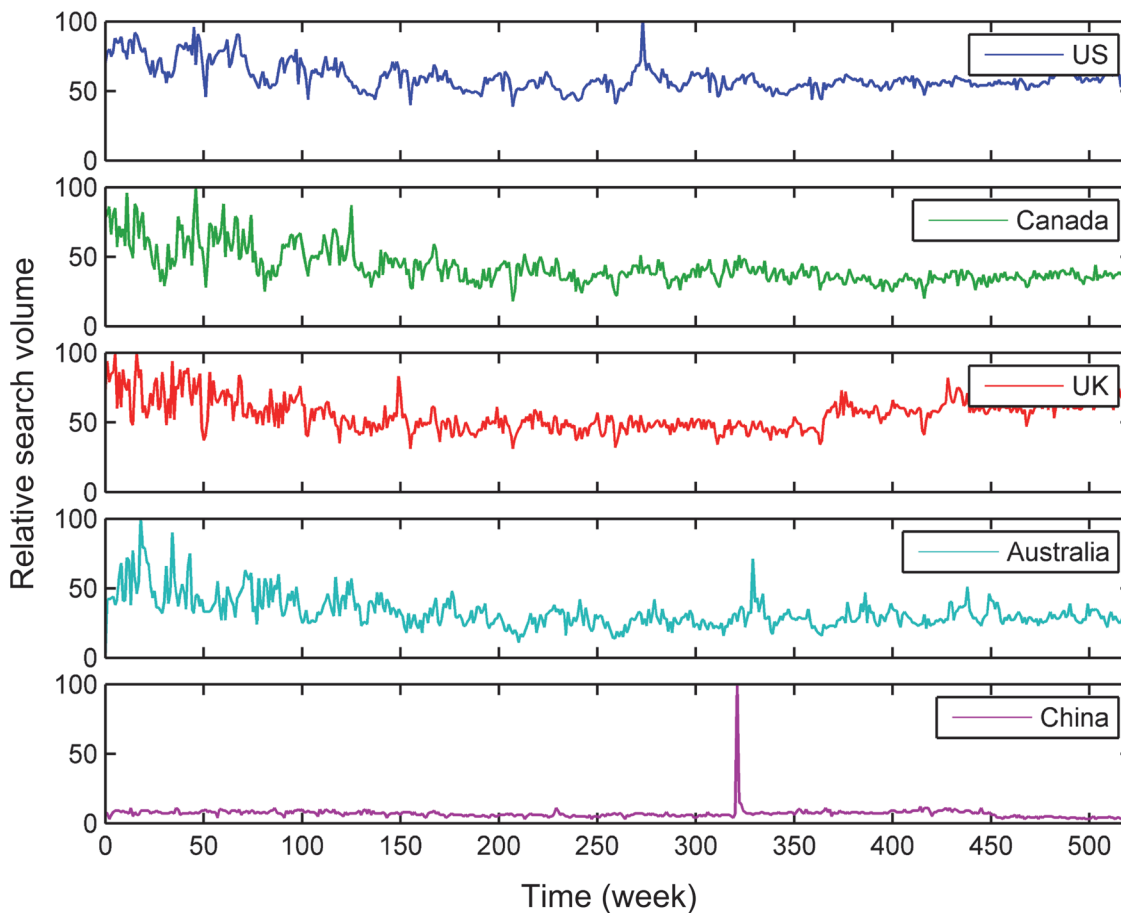


Fig 1. Raw search trends regarding tobacco from Google Trends. It contains the trends of the USA, Canada, the UK, Australia, and China. The time interval contains 522 weeks from January 4, 2004 to January 4, 2014.

doi:10.1371/journal.pone.0117938.g001

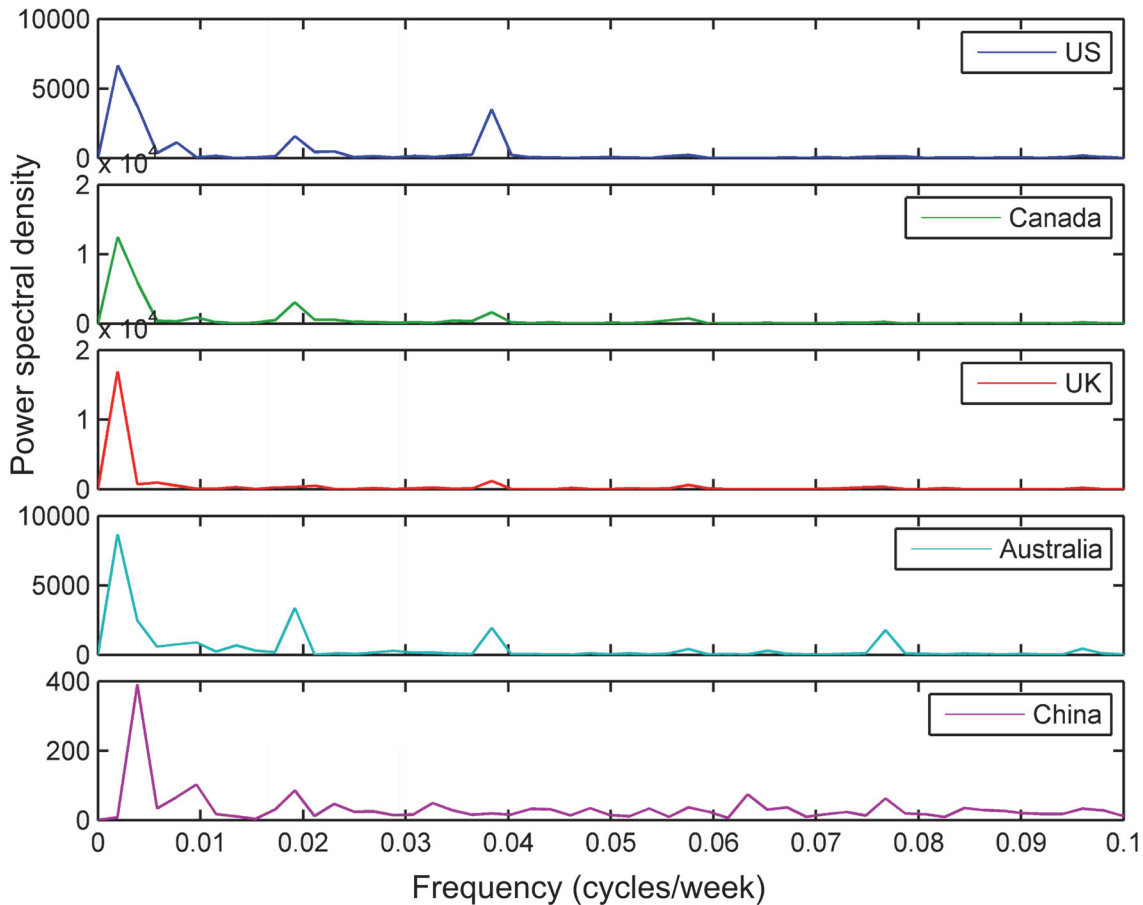


Fig 2. Periodograms of the tobacco-related search trends of the USA, Canada, the UK, Australia, and China. A peak in the periodogram indicates there is a seasonal component near the value corresponding to the peak.

doi:10.1371/journal.pone.0117938.g002

1.919×10^{-3} indicated the seasonal component whose period was the whole time interval (i.e., 1 cycle / $[1.919 \times 10^{-3} \text{ cycles/week}] \approx 521$ weeks) of the search trend. In addition, the seasonal component with half-year period was indicated by the peak near 0.04. We omitted other peaks in these periodograms, because they didn't widely exist in these periodograms.

In Fig. 2, the one-year and half-year seasonal components were more significant than other seasonal components. Since the previous study indicated the seasonality of cigarette sales in the USA had a 12-month period [3], only the one-year seasonal components were considered in detail in this paper. As illustrated in Fig. 3 and S3 Fig., the seasonal components with one-year period were extracted by using ideal pass filter on the raw trends, when we selected the frequency range from 0.015 to 0.025 in the periodograms. This frequency range only covered each peak shape near 0.0192 approximately, but didn't cover other peak shapes in Fig. 2 and S2 Fig. For Fig. 1, S1 Fig., Fig. 3 and S3 Fig., the relative search volumes of the countries on the northern (i.e., the USA, Canada, the UK, and China) and southern (i.e., Australia) hemispheres had an approximate sinusoidal pattern. The search trend of the USA almost had the same peak and trough with that of Canada, and the peak and trough of the search trend of the UK was close to that of the USA. The trend of Australia was nearly the inversion of the UK's search trend at the former half part of the search time series. The seasonal pattern of the search trend of China was not so regular as those of the other trends.

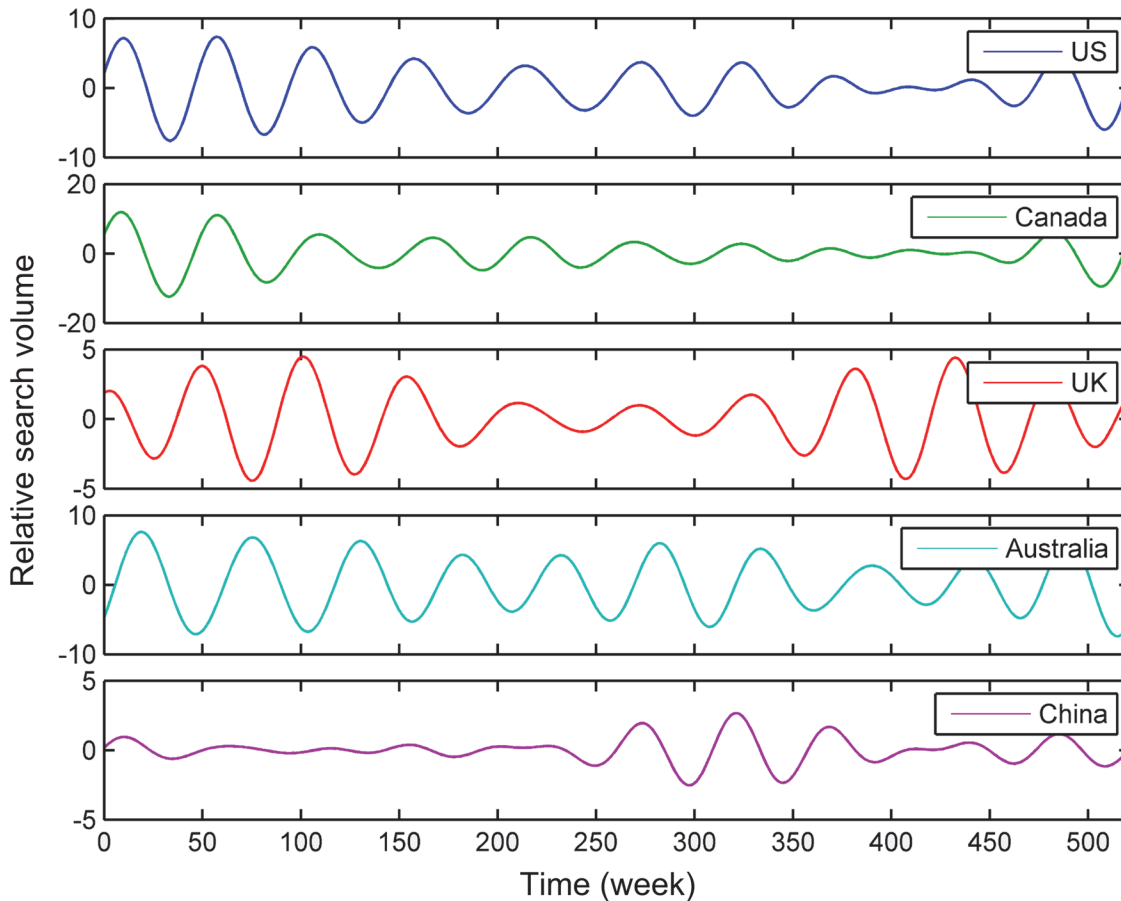


Fig 3. Seasonal components of the tobacco-related search trends of the USA, Canada, the UK, Australia, and China. These seasonal components are extracted by using ideal pass filter on the raw search trends.

doi:10.1371/journal.pone.0117938.g003

Before the cross-correlation analyses of the trends were conducted, we used Augmented Dickey-Fuller test to detect whether the extracted seasonal components of the studied trends were stationary. The results indicated all of them were stationary (See [S3](#) and [S4](#) Tables for details). In all tables of this paper, each cell contained two values: a correlation coefficient and a time lag that was in the parenthesis. For each correlation coefficient in all correlation tables, the row represented the first time series, and the column represented the second time series. Each lag value indicated that the number of time steps by which the second time series was shifted relative to the first time series. The unit of the time lag was week. In all cross-correlation plots, all lags were limited to the scale from -52 to 52 due to the one-year period of the seasonal components. Each time lag in each table was selected to provide the highest correlation among the paired cross-correlation values and time lags, and the correlation tables at lag 0 were provided in supporting information (See [S1](#) and [S2](#) Tables). For example, in [Table 1](#), the cross-correlation coefficient of the seasonal components related to tobacco search trends of the USA and Australia was 0.651, and (-15) meant Australia fell behind the USA 15 weeks in terms of these seasonal components. From [Table 1](#), we could observe that for two countries on the same hemisphere, the time lags corresponding to their correlation coefficients had smaller lengths than those of two countries on the different hemisphere except the coefficients of the UK and Canada (or China). For instance, the length of the time lag corresponding to the correlation of the USA and Canada equaled 0, smaller than that of the USA and Australia (i.e., it equaled

Table 1. The pairwise cross-correlations of seasonal components of tobacco-related search trends.

Country	Country				
	US	Canada	UK	Australia	China
US	1.000	0.906 (0)	0.717 (5)	0.651 (-15)	0.585 (0)
Canada	—	1.000	0.596 (-18)	0.788 (-14)	0.461 (-1)
UK	—	—	1.000	0.447 (33)	0.360 (-52)
Australia	—	—	—	1.000	0.581 (10)
China	—	—	—	—	1.000

The values in each cell contain two parts: a correlation coefficient and a time lag in the parenthesis.

doi:10.1371/journal.pone.0117938.t001

-15). As shown in [Table 1](#), for the correlations of these English-speaking countries, four out of six were above 0.600. Besides, two out of four were above 0.500 for the correlations of China with the other countries. Therefore, we could infer that seasonal components in the tobacco-related search trends had a significant positive correlation with each other for these four English-speaking countries.

The values in each cell contain two parts: a correlation coefficient and a time lag in the parenthesis.

As shown in [Table 2](#), the time lags of countries on the same hemisphere had smaller lengths than those on the different hemisphere except the coefficients of China and the USA (or Canada). Besides, for the correlations of these English-speaking countries, five out of six were above 0.600. Besides, one out of four was above 0.500 for the correlations of China with the other countries. The observation indicated that seasonal components in the search trends about lung cancer had a significant positive correlation with each other for these four English-speaking countries.

[Table 3](#) summarized the cross-correlations of the seasonal components of the search trends concerning tobacco and lung cancer. All correlation coefficients except that of China in the table were above 0.600. This observation indicated that the seasonal components of the search trends regarding tobacco had a significant positive correlation with that regarding lung cancer. One possible reason is that the association between tobacco and lung cancer has become very well known among those in the health communities, perhaps many information seekers will search for tobacco and lung cancer information together.

Discussion

The first important finding of this paper is that seasonality widely exists in the Internet search trend data concerning tobacco on an international scale. The second finding is that seasonal

Table 2. The pairwise cross-correlations of seasonal components of search trends regarding lung cancer.

Country	Country				
	US	Canada	UK	Australia	China
US	1.000	0.950 (1)	0.558 (5)	0.856 (12)	0.369 (51)
Canada	—	1.000	0.610 (4)	0.769 (-13)	0.348 (-51)
UK	—	—	1.000	0.612 (31)	0.719 (4)
Australia	—	—	—	1.000	0.419 (-31)
China	—	—	—	—	1.000

doi:10.1371/journal.pone.0117938.t002

Table 3. The cross-correlations between seasonal components of search trends regarding tobacco and those regarding lung cancer.

Query Term	Lung Cancer				
	US	Canada	UK	Australia	China
Tobacco	0.757 (-2)	0.845 (-1)	0.734 (0)	0.858 (0)	0.476 (51)

doi:10.1371/journal.pone.0117938.t003

components of the search trend data regarding tobacco have a significant positive correlation with each other across countries. These two findings complement the previous studies about smoking-related seasonality which used real-world survey data limited to the USA [3,4,5,6,7]. The third finding is that the Internet search trend data concerning lung cancer is similar to that of tobacco-related trend data in terms of seasonality. The fourth finding is that the seasonal components of the trend data regarding tobacco have a significant positive correlation with those regarding lung cancer. These findings about lung cancer provide a piece of novel Internet-based evidence for the health effect of tobacco use, although the significant correlation between smoking and lung cancer has been well-documented in related medical literature [1,24]. For the seasonality of the search trends and the peaks and troughs in the seasonal components of these trends, weather condition (especially temperature) may be one important factor of the reason, since some previous studies show that the weather condition (especially temperature) may be a factor contributing to the seasonal variation of smoking behavior [3,6]. In addition, the reason for the above mentioned time lag length may be that the similar weather condition (especially temperature) on the same hemisphere results in the similar smoking-related information seeking on the same hemisphere, based on the studies [3,6].

There are several implications in this study. First, the findings about seasonality in the Internet search trend data regarding tobacco and lung cancer provide the potential to help tobacco control professionals with the timing of tobacco control effort. Sheffer et al. demonstrated the increased participation in a statewide tobacco quitline service could be produced by launching a statewide media campaign timed to coincide with temporal smoking cessation behavioral patterns [25]. Similarly, it would be more effective for Internet users to access tobacco control information if some tobacco control campaigns are carried out at the peak period of tobacco information seeking. The reason is that information seekers' interest in tobacco information is the highest at that time. Second, the proposed investigation methodology for seasonality in Internet search trends can be extended to other disciplines including public health, sociology, and economics. By leveraging seasonality in Internet search trends, predictive models can be easily estimated [26]. Third, besides the Internet search query data, social media such as Facebook, Twitter, YouTube [27] etc. has become a new information source for researchers of different disciplines. Therefore, exploring collective intelligence in social media with social computing approaches [28,29] and new temporal or spatio-temporal analysis methods [30,31,32] can provide a novel insight for public health research (e.g., the analysis of smoking and smoking cessation behavior).

This study has the following limitations. First, the countries involved in the search trends only include four English-speaking developed countries and a developing country (i.e., China). The other English-speaking countries such as India, South Africa, and Pakistan are excluded in this study, because they are multilingual countries and using English query terms can't retrieve the effective and comprehensive search trends. In addition, the search trends about tobacco and lung cancer in New Zealand are not included because there are a number of missing values in the trend data. Therefore, sampling data from these countries may limit the generalizability of our findings. Second, the Internet search query data mainly reflects information seekers' interest in information regarding search query terms. We didn't compare the online search data

with real-world data, as it was very hard for us to collect such real-world tobacco and lung cancer data of these countries. Future work is needed to show to what extent the Internet search query data about tobacco and lung cancer can represent the real-world data about tobacco and lung cancer, if we can collect some appropriate real-world data. Additionally, more studies are required to investigate that what factors attribute to the reasons for the seasonality of the search trends and the peaks and troughs in the seasonal components of these trends. One possible explanation for seasonality may be that some tobacco-related conferences held at the same time of each year result in the much relevant information seeking on the Internet at the specific time of each year.

Conclusions

In summary, we studied seasonal patterns in information seeking concerning tobacco and lung cancer using Internet search information. Our findings indicate that seasonal patterns widely exist in information seeking concerning tobacco and lung cancer and have close correlations with each other across countries. Exploring the emerging online information including Internet search queries and social media information can provide novel data sources and investigation insights for public health research.

Supporting Information

S1 Fig. The raw search trends on lung cancer of the USA, Canada, the UK, Australia, and China from Google Trends, from January 4, 2004 to January 4, 2014.

(TIF)

S2 Fig. The periodograms of search trends on lung cancer in the USA, Canada, the UK, Australia, and China.

(TIF)

S3 Fig. The seasonal components of search trends about lung cancer of the USA, Canada, the UK, Australia, and China.

(TIF)

S4 Fig. The raw search trends of the query terms “cold” and “potato” of the USA.

(TIF)

S5 Fig. The periodograms of search trends of the query terms “cold” and “potato” of the USA.

(TIF)

S6 Fig. The raw search trends of the query terms “Google” and “pizza” of the USA.

(TIF)

S7 Fig. The periodograms of search trends of the query terms “Google” and “pizza” of the USA.

(TIF)

S1 Table. The pairwise cross-correlations (at lag 0) of seasonal components of tobacco-related search trends.

(DOCX)

S2 Table. The pairwise cross-correlations (at lag 0) of seasonal components of search trends regarding lung cancer.

(DOCX)

S3 Table. Stationarity test of seasonal components of tobacco-related search trends.
(DOCX)

S4 Table. Stationarity test of seasonal components of search trends regarding lung cancer.
(DOCX)

S1 Dataset. Raw search trend data regarding tobacco and lung cancer.
(ZIP)

S2 Dataset. Raw search trend data of the terms “cold”, “potato”, “Google”, and “pizza”.
(ZIP)

Acknowledgments

We thank each member of SMILES group in Institute of Automation, Chinese Academy of Sciences. Especially, we thank Kainan Cui, Chuan Luo, and Saike He for their useful discussions that improved this paper.

Author Contributions

Conceived and designed the experiments: ZZ DDZ SJL. Performed the experiments: ZZ. Analyzed the data: ZZ XZ SJL. Wrote the paper: ZZ XZ SJL.

References

1. World Health Organization (2012) WHO global report: mortality attributable to tobacco. Geneva: World Health Organization. Available: http://www.who.int/tobacco/publications/surveillance/rep_mortality_attributable/en/. Accessed 6 Aug 2014.
2. Mathers CD, Loncar D (2006) Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Med* 3: e442. PMID: [17132052](#)
3. Chandra S, Chaloupka FJ (2003) Seasonality in cigarette sales: patterns and implications for tobacco control. *Tob Control* 12: 105–107. PMID: [12612375](#)
4. Chandra S, Gitchell JG, Shiffman S (2011) Seasonality in sales of nicotine replacement therapies: patterns and implications for tobacco control. *Nicotine Tob Res* 13: 395–398. doi: [10.1093/ntr/ntq258](#) PMID: [21330268](#)
5. Colwell B, Ramirez N, Koehly L, Stevens S, Smith DW, et al. (2006) Seasonal variations in the initiation of smoking among adolescents. *Nicotine Tob Res* 8: 239–243. PMID: [16766416](#)
6. Momperousse D, Delnevo CD, Lewis MJ (2007) Exploring the seasonality of cigarette-smoking behaviour. *Tob Control* 16: 69–70. PMID: [17297079](#)
7. Wellman RJ, DiFranza JR (2003) Seasonality in onset of youth smoking parallels seasonality in cigarette sales. *Tob Control* 12: 339. PMID: [12958399](#)
8. Ayers JW, Althouse BM, Allem J-P, Ford DE, Ribisl KM, et al. (2012) A Novel Evaluation of World No Tobacco Day in Latin America. *J Med Internet Res* 14: e77. doi: [10.2196/jmir.2148](#) PMID: [22634568](#)
9. Boland M, Sweeney M, Scallan E, Harrington M, Staines A (2006) Emerging advantages and drawbacks of telephone surveying in public health research in Ireland and the U.K. *BMC Public Health* 6: 208. PMID: [16911771](#)
10. Eysenbach G (2006) Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc*: 244–248. PMID: [17238340](#)
11. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014. doi: [10.1038/nature07634](#) PMID: [19020500](#)
12. Luo Y, Zeng D, Cao Z, Zheng X, Wang Y, et al. (2010) Using multi-source web data for epidemic surveillance: A case study of the 2009 Influenza A (H1N1) pandemic in Beijing. 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI). pp. 76–81.
13. Cui K, Cao Z, Zheng X, Zeng D, Zeng K, et al. (2011) A Geospatial Analysis on the Potential Value of News Comments in Infectious Disease Surveillance. Proceedings of the 6th Pacific Asia conference on Intelligence and security informatics. Beijing, China: Springer-Verlag. pp. 85–93.

14. Yang AC, Huang NE, Peng C-K, Tsai S-J (2010) Do Seasons Have an Influence on the Incidence of Depression? The Use of an Internet Search Engine Query Data as a Proxy of Human Affect. *PLoS one* 5: e13728. doi: [10.1371/journal.pone.0013728](https://doi.org/10.1371/journal.pone.0013728) PMID: [21060851](https://pubmed.ncbi.nlm.nih.gov/21060851/)
15. Cobb NK (2010) Online consumer search strategies for smoking-cessation information. *Am J Prev Med* 38: S429–432. doi: [10.1016/j.amepre.2009.12.001](https://doi.org/10.1016/j.amepre.2009.12.001) PMID: [20176318](https://pubmed.ncbi.nlm.nih.gov/20176318/)
16. Cobb NK, Graham AL, Byron MJ, Niaura RS, Abrams DB, et al. (2011) Online Social Networks and Smoking Cessation: A Scientific Research Agenda. *J Med Internet Res* 13: e119. doi: [10.2196/jmir.1911](https://doi.org/10.2196/jmir.1911) PMID: [22182518](https://pubmed.ncbi.nlm.nih.gov/22182518/)
17. Ayers JW, Ribisl K, Brownstein JS (2011) Using Search Query Surveillance to Monitor Tax Avoidance and Smoking Cessation following the United States' 2009 "SCHIP" Cigarette Tax Increase. *PLoS one* 6: e16777. doi: [10.1371/journal.pone.0016777](https://doi.org/10.1371/journal.pone.0016777) PMID: [21436883](https://pubmed.ncbi.nlm.nih.gov/21436883/)
18. Ayers JW, Ribisl KM, Brownstein JS (2011) Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 40: 448–453. doi: [10.1016/j.amepre.2010.12.007](https://doi.org/10.1016/j.amepre.2010.12.007) PMID: [21406279](https://pubmed.ncbi.nlm.nih.gov/21406279/)
19. Carr LJ, Dunsiger SI (2012) Search Query Data to Monitor Interest in Behavior Change: Application for Public Health. *PLoS one* 7: e48158. doi: [10.1371/journal.pone.0048158](https://doi.org/10.1371/journal.pone.0048158) PMID: [23110198](https://pubmed.ncbi.nlm.nih.gov/23110198/)
20. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, et al. (2014) Monitoring of non-cigarette tobacco use using Google Trends. *Tob Control*. doi: [10.1136/tobaccocontrol-2013-051276](https://doi.org/10.1136/tobaccocontrol-2013-051276).
21. Santillana M, Zhang DW, Althouse BM, Ayers JW (2014) What Can Digital Disease Detection Learn from (an External Revision to) Google Flu Trends? *Am J Prev Med*. doi: [10.1016/j.amepre.2014.05.020](https://doi.org/10.1016/j.amepre.2014.05.020).
22. Zhang Z, Zheng X, Zeng DD, Cui K, Luo C, et al. (2013) Discovering seasonal patterns of smoking behavior using online search information. 2013 IEEE International Conference on Intelligence and Security Informatics (ISI). pp. 371–373.
23. MathWorks MATLAB 7.11.0 (R2010b).
24. U.S. Department of Health and Human Services (2004) *The Health Consequences of Smoking: A Report of the Surgeon General*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. PMID: [20669512](https://pubmed.ncbi.nlm.nih.gov/20669512/)
25. Sheffer MA, Redmond LA, Kobinsky KH, Keller PA, McAfee T, et al. (2010) Creating a perfect storm to increase consumer demand for Wisconsin's Tobacco Quitline. *Am J Prev Med* 38: 014.
26. Choi H, Varian H (2012) Predicting the present with google trends. *Economic Record* 88: 2–9.
27. Luo C, Zheng X, Zeng DD, Leischow S (2014) Portrayal of electronic cigarettes on YouTube. *BMC Public Health* 14: 1028. doi: [10.1186/1471-2458-14-1028](https://doi.org/10.1186/1471-2458-14-1028) PMID: [25277872](https://pubmed.ncbi.nlm.nih.gov/25277872/)
28. Li X, Mao W, Zeng D, Wang F-Y (2008) Agent-based social simulation and modeling in social computing. *Intelligence and Security Informatics*: Springer Berlin Heidelberg. pp. 401–412.
29. Zeng D, Chen H, Lusch R, Li S-H (2010) Social Media Analytics and Intelligence. *IEEE Intelligent Systems* 25: 13–16.
30. Chang W, Zeng D, Chen H (2005) Prospective spatio-temporal data analysis for security informatics. 8th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 1120–1124.
31. Lu H-M, Zeng D, Chen H (2010) Prospective infectious disease outbreak detection using Markov switching models. *IEEE Transactions on Knowledge and Data Engineering* 22: 565–577.
32. Cui K, Zheng X, Zhang Z, Zeng D (2014) Analyzing Spatio-temporal Patterns of Online Public Attention in Emergency Events: A Case Study of 2009 H1N1 Influenza Outbreak in China. In: Zheng X, Zeng D, Chen H, Zhang Y, Xing C et al., editors. *Smart Health*: Springer International Publishing. pp. 44–50.