RESEARCH ARTICLE

News trends and web search query of HIV/AIDS in Hong Kong

Alice P. Y. Chiu, Qianying Lin, Daihai He*

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong SAR, China

* daihai.he@polyu.edu.hk

Abstract

Background

The HIV epidemic in Hong Kong has worsened in recent years, with major contributions from high-risk subgroup of men who have sex with men (MSM). Internet use is prevalent among the majority of the local population, where they sought health information online. This study examines the impacts of HIV/AIDS and MSM news coverage on web search query in Hong Kong.

Methods

Relevant news coverage about HIV/AIDS and MSM from January 1st, 2004 to December 31st, 2014 was obtained from the WiseNews databse. News trends were created by computing the number of relevant articles by type, topic, place of origin and sub-populations. We then obtained relevant search volumes from Google and analysed causality between news trends and Google Trends using Granger Causality test and orthogonal impulse function.

Results

We found that editorial news has an impact on "HIV" Google searches on HIV, with the search term popularity peaking at an average of two weeks after the news are published. Similarly, editorial news has an impact on the frequency of "AIDS" searches two weeks after. MSM-related news trends have a more fluctuating impact on "MSM" Google searches, although the time lag varies anywhere from one week later to ten weeks later.

Conclusions

This infodemiological study shows that there is a positive impact of news trends on the online search behavior of HIV/AIDS or MSM-related issues for up to ten weeks after. Health promotional professionals could make use of this brief time window to tailor the timing of HIV awareness campaigns and public health interventions to maximise its reach and effectiveness.

Introduction

The global epidemic of HIV continues to expand. Men who have sex with men (MSM) represents a dominant route of transmission [1, 2]. In Hong Kong, ever since the first case of human immunodeficiency virus(HIV) was reported in 1984 [3], the rate of HIV infections in MSM is rapidly expanding. From 2004 to 2014, the annual number of new HIV cases had increased from 268 to 651, while Acquired Immunodeficiency Syndrome(AIDS) had also increased from 49 cases to 108 cases [3]. Homosexual transmission of HIV was growing fast from 24% to 57% during the same period [3].

Infodemiology is the study of distribution and determinants of information on the internet, and is a powerful tool to inform public health and public policy [4]. Infodemiology methods are either supply-based or demand-based. Supply-based methods analyse online information by classifying them by topic and use indicators to measure changes over time. Demand-based methods use data generated from online search behavior of people such as the number of searches on a specific topic, or the number of visits on websites with a specific topic [5].

Infodemiology had been applied in previous public health studies for various purposes, examples include its use as a syndromic surveillance tool for influenza and dengue fever [6, 7]; as a proxy indicator to measure the awareness of an autism campaign [8]; as an infodemiological tool to study online search patterns for dizziness-related information and prescription drugs [9, 10]; as an infoveillance tool to study the temporal relationships between bariatric surgery-related media events and insurance policy modification [11]; as a tool to measure sudden increases in non-cigarette tobacco product searches following cigarette tax increases [12]; as an impact measurement tool to study the online information seeking behavior following a public figure's pancreatic cancer announcement [13].

Infodemiological design was also used in HIV/AIDS studies [14, 15]. "HIV" and "AIDS" searches were shown to be correlated with HIV prevalence in different regions of USSR [14]. A recent Canadian study showed that HIV infection rates and HIV-related online campaign activities are positively correlated [15].

The impacts of HIV/AIDS news on awareness, behavior and online searches had been explored. Kalichman et al. reported that following the public announcement of HIV sero-positivity status of basketball star Earvin "Magic" Johnson, there were substantial increases in the awareness of HIV/AIDS, increase in HIV testing and changes in high risk behavior. However, such effects were transient and subsided within three weeks of the announcement [16]. Gabarron et al. demonstrated that public events such as World AIDS Day and actor Charlie Sheen's announcement of HIV positive status led to increases in Wikipedia searches for HIV-related articles on the day of the event [17].

The relationship between news coverage and online search patterns had also been studied in other diseases. During the Zika virus outbreaks in 2016, information seeking behavior about the disease substantially increased following government announcements [18]. During the weeks following the imported case of Ebola in the United States, there was significant information contagion of Ebola-related media coverage on related tweets and internet searches [19]. Niederdeppe et al's study found that there was a positive relationship between cancer news coverage and online searches [20].

Clinicians, public health professionals and the general public regularly seek health information on the internet [21]. In Hong Kong, internet uses had reached 73% of the population in 2014. Among them, 96% used the internet for information search [22], and 44% sought health information online [23]. Google had a two-third market share among search engine users in Hong Kong, followed by Yahoo [24]. Google Trends compares the relative search volumes of specified keywords entered by internet uses over time [25]. However, little is known about the

impact of routine HIV news coverage on online information searches. Our study aims to explore the temporal trends of news coverage patterns of HIV, AIDS and MSM and examine their magnitude and duration of impacts on relevant online search patterns in Hong Kong from 2004 to 2014.

Materials and methods

Google Trends

Google Trends (Google Inc, Mountain View, California, USA) is a publicly available source that provides normalised Google Trends statistics for different queries since 2004. It analyses the Google search queries to determine how many searches were done on the keywords you entered, compared to the total number of Google searches done during that time by all users. Google Trends exclude terms with very low search volumes or duplicate searches made by the same users over a very short period of time. Google Trends are normalised to a scale from 0 (i.e. relative search volume less than 1% of peak volume) to 100 (i.e. relative search volume reaches its peak). We obtained weekly data from January 1st, 2004 to December 31st, 2014, and analysed searches conducted in Hong Kong only. The keywords we considered include "HIV", "AIDS", "MSM", "gay", "homosexual", "ai zi bing" (the Chinese term for AIDS), "nan tong xing lian" (the Chinese term for MSM).

News trends

For news trends, we used WiseNews database as a news extraction tool. WiseNews is a database of newspapers, magazines, journals and newswires published in Hong Kong and other countries. The database covers 18 major local newspapers in both Chinese and English. Our search included four newspapers, Apple Daily, Oriental Daily, Metro Daily and AM730, which represented two paid newspapers and two freely distributed newspapers of top circulations in Hong Kong, from January 1st, 2004 to December 31st, 2014. We used the same set of keywords as with Google Trends. An article is retrieved if it possessed the specified keywords in the news' headlines or content. Four student helpers assisted in the data extraction process. Each student was responsible for screening half of the time period. Thus, two students worked in parallel and conducted independent screening for each study year. Any discrepancies are resolved by the first author (A.C.).

For an article to be considered relevant, it needs to meet the following inclusion criteria: the news article's primary topic is clearly related to HIV, AIDS or MSM. For example, it could be a MSM movie review, a scientific discovery related to HIV or AIDS, a sex-related crime committed by a MSM, or a HIV/AIDS awareness event. On the other hand, there is an exclusion criteria that the news article is not related to HIV, AIDS or MSM. For example, it could be news related to Middle East Respiratory syndrome's drug treatment with a brief mentioning of HIV/AIDS drug treatment as a secondary topic, or news related to MSM committing a crime, but the crime is clearly not HIV/AIDS or MSM-related, such as theft. Each article is scanned for relevance according to the specified criteria by title, and if necessary, the full news article is read and checked.

For each relevant news article, we extracted the following variables into a Microsoft Excel file: type of article (e.g. news, editorial or press release); article topic (e.g. HIV treatment, HIV vaccine, HIV prevention, HIV transmission or risk, HIV surveillance, surveys or statistics, HIV science, HIV publicity events, HIV education or awareness, HIV policy and laws, social issues, sex-related crimes, HIV testing); place of origin (e.g. Hong Kong, Macau, Mainland China, Taiwan or overseas); sub-populations covered (e.g. MSM, lesbians, people living with

HIV, general population, celebrities); and total number of news articles (i.e. the number of articles published on the same issue or topic on the same date).

Pilot study and inter-rater reliability

We conducted a pilot study where four student helpers independently extracted news articles using the methods described above, in the month of December 2004. We computed Kappa coefficient for inter-rater reliability (IRR) for each variable collected [26].

Data compilation and statistical analyses

Following the data extraction process, we created Google Trends time series and WiseNews Trends that represent aggregate time series and categories that are classified as news article type, news topic, news origin or population groups from 2004 to 2014. Data are then imported to R Statistical Software version 3.4.0 (https://cran.r-project.org/) for statistical analyses.

The steps for analysing the causality between news trends and Google Trends are as follows. We first applied moving average models to decompose Google Trends into three time series-trends, seasonal and random factors. Anderson Darling tests were then used to test for normality of the decomposed random factors. Augmented Dickey Fuller tests were used to test for stationarity of the decomposed random factors. The number of lags was determined based on four information criteria: Akaike Information Criterion (AIC), Schwarz Criterion (SC), Hannan-Quinn (HQ) and Final Prediction Error (FPE). We selected the lag which is the mode determined by these four criteria. Based on the time lag, the cause-effect relationships of news trends on Google Trends can be analysed by applying the Granger Causality tests. This statistical test does not rely on a model specification and therefore is suitable for empirical investigations of cause-effect relationships [27–29]. Granger causality test is suitable for analysing our data because of the following assumptions: (i) The effect (i.e. Google Trends) does not precede its cause temporally. (ii) The causal time series (i.e. Google Trends) contains unique information about the time series being caused (i.e. news trends that is not available otherwise) [27–29]. Statistically significant pairs of time series were further explored using impulse response function. Essentially, we investigated the reaction in Google Trends in response to changes in the news trends. Statistical significance was assessed at the 5% levels.

This study involves secondary data collection and data analyses only, and human participants are not involved. Ethical approval was not needed.

Results

Inter-rater reliability

We determined IRR during the pilot study. Four student helpers independently extracted, screened and recorded WiseNews articles from December 1st, 2004 to December 31st, 2004. 160 news records were retrieved. Kappa coefficients for different variables were computed as follows: publication date(1.00); number of articles(0.608); news origin(0.375); population group studied(0.530); article type(0.430); and article topic(0.192).

Impact of news trends on Google Trends

Using Granger Causality tests, we found that the following pairs of news trends and Google Trends are statistically significant at 5% levels: (i) editorial news trends on HIV Google Trends; (ii) editorial or other news trends on AIDS Google Trends; and (iii) MSM-related news trends on MSM Google Trends. We further explored these relationships using impulse response function and the results are summarised in Figs 1–3.
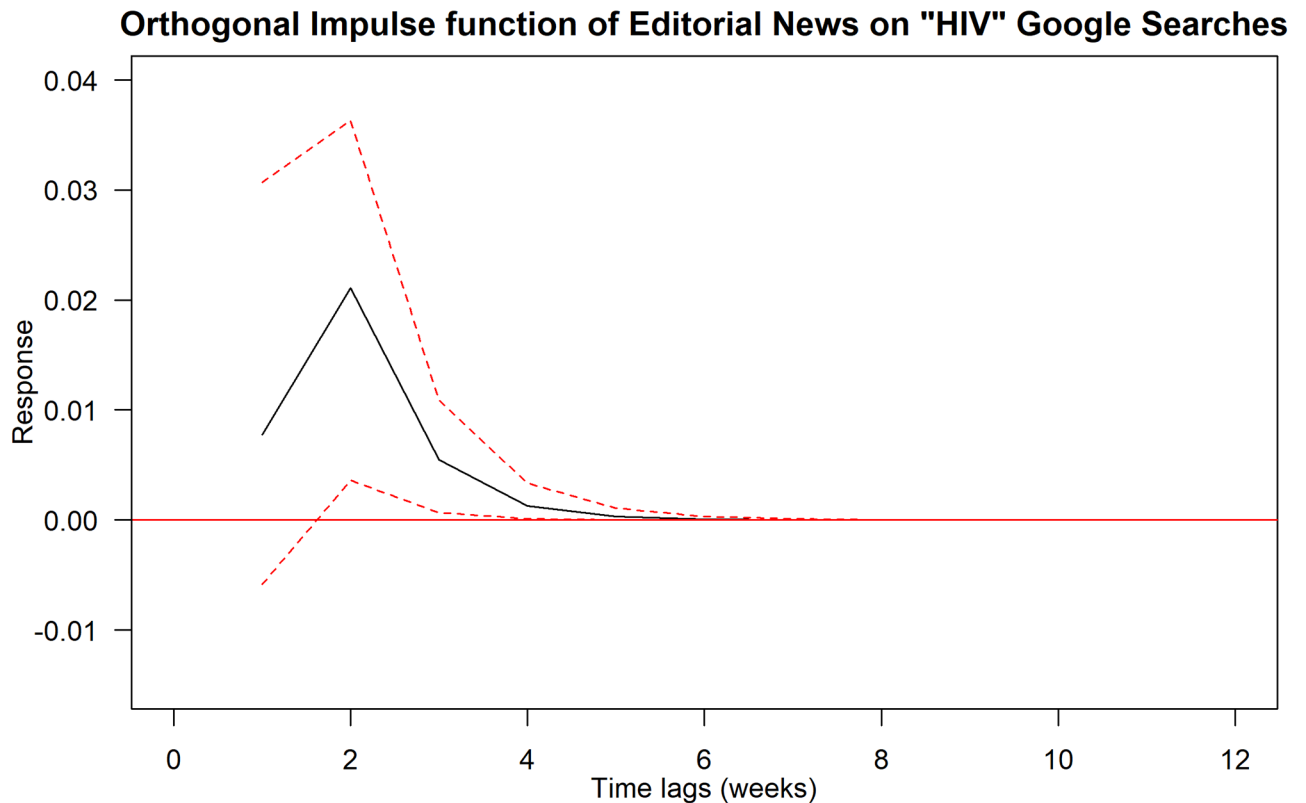
Orthogonal Impulse function of Editorial News on "HIV" Google Searches

Fig 1. Orthogonal impulse function of editorial news trends on HIV Google Trends. Black solid line shows the response curve. Red-dashed line indicates the 95% confidence interval.

Fig 1 shows the orthogonal impulse function of Editorial news trends on HIV Google Trends. Note that the impact of editorial news on "HIV" google searches increases rapidly and reaches its peak on average of two weeks after the news are published. The impact decreases rapidly three weeks after and declines gradually to zero level six weeks after.

Fig 2 displays the orthogonal impulse function of Editorial or other news on AIDS Google Trends. Again, we found that the impact of Google Trends increases after one week and reaches its peak after two weeks. It declines abruptly three weeks after. It reaches zero level five weeks after.

In Fig 3, we show the orthogonal impulse function of MSM-related news trends on MSM Google Trends. We found that there is a more fluctuating impact after one week. It drops to zero level four weeks after, but increases again and reaches its peak impact six weeks after. It continues to fluctuate around the peak level seven and eight weeks after. The impact suddenly drops to negative nine weeks after. The impact is slightly negative ten weeks after but returns to zero level 11 weeks after.

Discussion

To the best of our knowledge, this study is the first to examine the relationship between news trends and search engine queries on the topics of HIV, AIDS and MSM. We found that editorial news have a positive impact on searches on both HIV and AIDS. Also, news trends on a MSM-related topic had a predominantly positive impact on search behavior on the same topic. Such impacts lasted for up to ten weeks. Such observation is supported by the channel

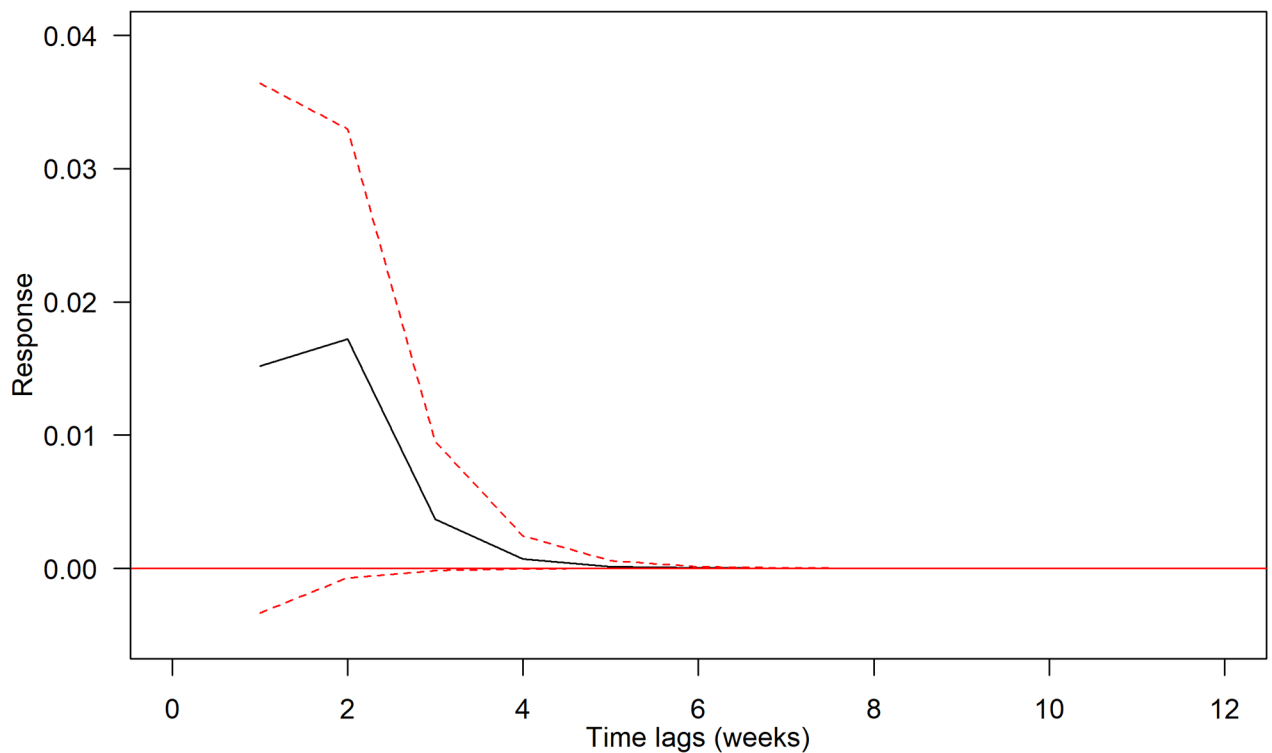## Orthogonal Impulse function of Editorial News on "AIDS" Google Searches

Fig 2. Orthogonal impulse function of Editorial news on AIDS Google Trends. Black solid line shows the response curve. Red-dashed line indicates the 95% confidence interval.

complementarity theory that was proposed by Dutta-Bergman [30, 31]. This theory implies there is a congruence between the online search behavior and news media within a specific topic [30, 31]. We could also apply this theory in our context to explain why HIV news coverage could lead to online information searches.

Gabarron et al. found that the number of HIV queries on Wikipedia increased significantly during international public events such as World AIDS Day and the announcement of the HIV-seropositivity of actor Charlie Sheen [17]. Consistent with their findings, our study further showed that Google online searches has a lagged time impact of up to ten weeks after the publication of HIV-related news.

The temporal relationships between health news and online searches had also been demonstrated in cancers [20], Ebola [19] and mammography screening [32]. Niederdeppe et al. noted a positive relationship between cancer news coverage and information seeking [20]. Towers et al. found evidence of contagion of fear during the Ebola outbreaks, where Ebola-related news videos would lead to tens of thousands of Ebola-related internet searches and tweets [19]. Weeks et al. suggests that news coverage of a mammography guideline may have driven online information seeking. In line with previous research, our study demonstrates that news coverage of HIV, AIDS or MSM related topic will stimulate online searches on these topics.

Our study has important public health implications. Health authorities and AIDS-related non-governmental organisations can make use of the time window of increased online search interest to promote awareness and provide public education on this topic [18]. Awareness
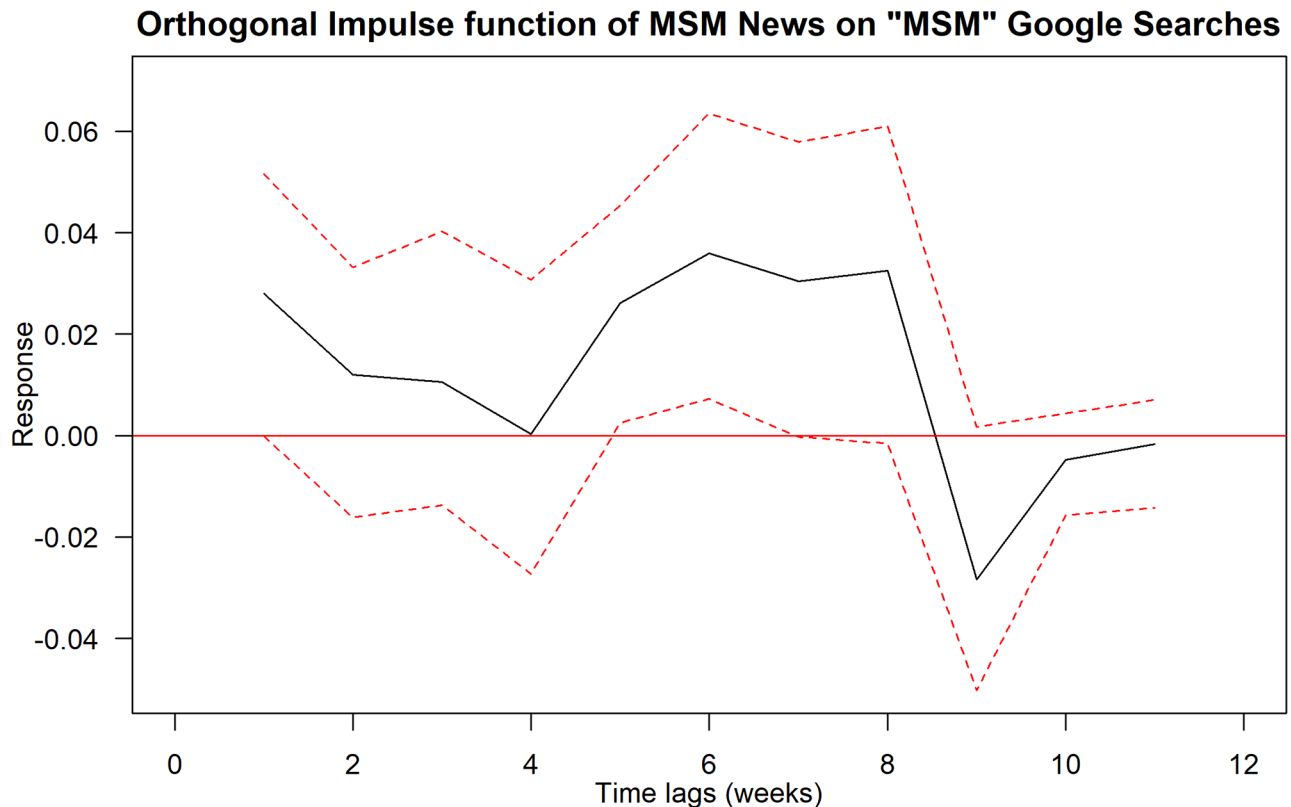
Orthogonal Impulse function of MSM News on "MSM" Google Searches

Fig 3. Orthogonal impulse function of MSM-related news trends on MSM Google Trends. Black solid line shows the response curve. Red-dashed line indicates the 95% confidence interval.

campaigns, publicity events, preventive education and other health promotion activities could be implemented after significant HIV, AIDS or MSM news to attract more public attention.

Our study has several key strengths. First, our findings are novel in identifying the duration of time-window for online search increases following HIV news coverage. Second, we used triangulation methods which employed both qualitative and quantitative techniques. Our news searches are screened for content relevance by student helpers. In contrast, previous studies mostly used keyword matching techniques without considering the content relevance [20]. Third, our study could be conducted at a low cost and cover a wide study population.

Our study is limited by several factors. First, only aggregate data were available. We could not confirm whether news and media coverage could indeed impact online searching behavior at the individual level. Second, we have included only four newspapers rather than a comprehensive coverage of all local newspapers in Hong Kong, although we believe our selection of two free and two paid newspapers of top circulation should cover a broad readership. Third, the IRR of the news coverage searches is only satisfactory.

Conclusions

This infodemiological study shows that news trends have a positive impact on online search behavior on HIV, AIDS or MSM-related issues. Health promotional professionals could make use of these insights to design timely online awareness campaigns and public health interventions that will maximise reach and effectiveness. Our research techniques could be applied to

study other sexually transmitted diseases. Our results have implications on the timing of disseminating HIV, AIDS and MSM health information online.

Supporting information

S1 Table. HIV Google Trends.
(XLSX)

S2 Table. AIDS Google Trends.
(XLSX)

S3 Table. MSM Google Trends.
(XLSX)

S4 Table. WiseNews coverage 2004–2014.
(XLSX)

Author Contributions

Conceptualization: Alice P. Y. Chiu, Daihai He.

Data curation: Alice P. Y. Chiu, Qianying Lin, Daihai He.

Formal analysis: Alice P. Y. Chiu, Qianying Lin, Daihai He.

Funding acquisition: Alice P. Y. Chiu, Daihai He.

Investigation: Alice P. Y. Chiu, Qianying Lin, Daihai He.

Methodology: Alice P. Y. Chiu, Qianying Lin, Daihai He.

Project administration: Daihai He.

Software: Qianying Lin.

Validation: Qianying Lin.

Visualization: Qianying Lin.

Writing – original draft: Alice P. Y. Chiu.

Writing – review & editing: Alice P. Y. Chiu, Daihai He.

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet. 2012; 380:367–77. https://doi.org/10.1016/S0140-6736(12)60821-6 PMID: 22819660

2. Beyrer C, Baral SD, Collins C, Richardson ET, Sullivan PS, Sanchez J, et al. The global response to HIV in men who have sex with men. Lancet. 2016; 388:198–206. https://doi.org/10.1016/S0140-6736(16)30781-4 PMID: 27411880

3. Chan K, Shu BY, Sit A, Yim S. HIV Surveillance Report—2015 Update. Centre for Health Protection, Department of Health, HKSAR. 2016.

4. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med. 2011; 40(5 Suppl 2):S154–8. https://doi.org/10.1016/j.amepre.2011.02.006 PMID: 21521589

5. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res. 2009; 11:e11. https://doi.org/10.2196/jmir.1157 PMID: 19329408

6. Thompson LH, Malik MT, Gumel A, Strome T, Mahmud SM. Emergency department and 'Google flu trends' data as syndromic surveillance indicators for seasonal influenza. Epidemiol Infect. 2014; 142:2397–405. https://doi.org/10.1017/S0950268813003464 PMID: 24480399

7. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. PLoS Negl Trop Dis. 2014; 8:e2713. https://doi.org/10.1371/journal.pntd.0002713 PMID: 24587465

8. DeVilbiss EA, Lee BK. Brief report: trends in US National autism awareness from 2004 to 2014: the impact of national autism awareness month. J Autism Dev Disord. 2014; 44:3271–3. https://doi.org/10.1007/s10803-014-2160-4 PMID: 24915931

9. Kerber KA, Skolarus LE, Callaghan BC, Zheng K, Zhang Y, An LC, et al. Consumer Demand for Online Dizziness Information: If You Build it, They may Come. Front Neurol. 2014; 5:50. https://doi.org/10.3389/fneur.2014.00050 PMID: 24795690

10. Simmering JE, Polgreen LA, Polgreen PM. Web search query volume as a measure of pharmaceutical utilization and changes in prescribing patterns. Res Social Adm Pharm. 2014; 10:896–903. https://doi.org/10.1016/j.sapharm.2014.01.003 PMID: 24603135

11. Linkov F, Bovbjerg DH, Freese KE, Ramanathan R, Eid GM, Gourash W. Bariatric surgery interest around the world: what Google Trends can teach us. Surg Obes Relat Dis. 2014; 10:533–8. https://doi.org/10.1016/j.soard.2013.10.007 PMID: 24794184

12. Jo CL, Ayers JW, Althouse BM, Emery S, Huang J, Ribisl KM. US consumer interest in non-cigarette tobacco products spikes around the 2009 federal tobacco tax increase. Tob Control. 2015; 24:395–9. https://doi.org/10.1136/tobaccocontrol-2013-051261 PMID: 24500270

13. Noar SM, Ribisl KM, Althouse BM, Willoughby JF, Ayers JW. Using digital surveillance to examine the impact of public figure pancreatic cancer announcements on media and search query outcomes. J Natl Cancer Inst Monogr. 2013; 2013:188–94. https://doi.org/10.1093/jncimonographs/lgt017 PMID: 24395990

14. Zheluk A, Quinn C, Hercz D, Gillespie JA. Internet search patterns of human immunodeficiency virus and the digital divide in the Russian Federation: infoveillance study. J Med Internet Res. 2013; 15:e256. https://doi.org/10.2196/jmir.2936 PMID: 24220250

15. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. JMIR Public Health Surveill. 2016; 2:e156. https://doi.org/10.2196/publichealth.6504 PMID: 27733330

16. Kalichman SC. Magic Johnson and public attitudes toward AIDS: a review of empirical findings. AIDS Educ Prev. 1994; 6:542–57. PMID: 7702965

17. Gabarron E, Lau AY, Wynn R. Weekly Pattern for Online Information Seeking on HIV—A Multi-Language Study. Stud Health Technol Inform. 2016; 228:778–82. PMID: 27577492

18. Southwell BG, Dolina S, Jimenez-Magdaleno K, Squiers LB, Kelly BJ. Zika Virus-Related News Coverage and Online Behavior, United States, Guatemala, and Brazil. Emerg Infect Dis. 2016; 22:1320–1. https://doi.org/10.3201/eid2207.160415 PMID: 27100826

19. Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, et al. Mass Media and the Contagion of Fear: The Case of Ebola in America. PLoS One. 2015; 10:e0129179. https://doi.org/10.1371/journal.pone.0129179 PMID: 26067433

20. Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information seeking. J Health Commun. 2008; 13:181–99. https://doi.org/10.1080/10810730701854110 PMID: 18300068

21. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. N Engl J Med. 2009; 360:2153–5, 2157. https://doi.org/10.1056/NEJMp0900702 PMID: 19423867

22. Census and Statistics Department, Hong Kong. Thematic Household Survey Report—Report No. 62-Information technology usage and penetration. 2013.

23. Yan YY. Online health information seeking behavior in Hong Kong: an exploratory study. J Med Syst. 2010; 34:147–53. https://doi.org/10.1007/s10916-008-9226-9 PMID: 20433053

24. StatCounter, Global Stats. Search Engine Market Share in Hong Kong. http://gs.statcounter.com/search-engine-market-share/all/hong-kong. Accessed on 31 May 2017.

25. Google, Inc. Google Trends. https://trends.google.com/trends/. Accessed on 31 May 2017.

26. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. 3rd ed. 2003.

27. Eichler M. Causal Inference in Time Series Analysis. In Berzuini Carlo. Causality: statistical perspectives and applications (3rd ed.). Hoboken, N.J.: Wiley. 2012; 327–352.

28. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica. 1969; 37:424–38. https://doi.org/10.2307/1912791

29. Granger CWJ. Long memory relationships and the aggregation of dynamic models. J Econom. 1980; 14: 227–38. https://doi.org/10.1016/0304-4076(80)90092-5

30. Dutta-Bergman M. Interpersonal communication after 9/11 via the telephone and the Internet: Theory of channel complementarity. New Media Soc. 2004; 6:661–75. https://doi.org/10.1177/146144804047086

31. Dutta-Bergman M. Complementarity in consumption of news types across traditional and new media. JoBEM. 2004; 48:41–60.

32. Weeks BE, Friedenberg LM, Southwell BG, Slater JS. Behavioral consequences of conflict-oriented health news coverage: the 2009 mammography guideline controversy and online information seeking. Health Commun. 2012; 27:158–66. https://doi.org/10.1080/10410236.2011.571757 PMID: 21823950