

BRIEF COMMUNICATION OPEN

Google searches do not correlate with melanoma incidence in majority English speaking countries

Laura McDonald¹, Alex Simpson², Sophie Graham², Anna Schultze², Beth Nordstrom³, Piyush Durani⁴, Faisal Mehmud⁴ and Sreeram V. Ramagopalan¹

Recent reports have suggested that internet search behaviour may be a valuable tool to estimate melanoma incidence and mortality. Previous studies have used incorrect statistical methods, were focussed on the United States and/or did not use non-cancer control search terms to provide a context for interpreting the effects seen with the cancer-related terms. Using more robust statistical methods we found that no cancer search terms were significantly, or strongly correlated with melanoma incidence in 6 countries.

npj Digital Medicine (2018)1:44; doi:10.1038/s41746-018-0050-4

INTRODUCTION

The internet is an important source of health information for patients. In 2013 surveys showed that >70% of adults looked up online health information in the preceding year.¹ Although many of these online health searches may be more general or related to an already-diagnosed condition or planned treatment, 35% of Americans reported looking online specifically to determine what medical condition they may have.¹ Google is the world's most popular search engine² and ~5% of all Google searches are for health-related information.³ This vast volume of search data is therefore a potential option to provide a robust and real-time surveillance system of epidemics and diseases.⁴

To this end, recent reports have suggested that Google internet search behaviour may be a valuable tool to estimate cancer incidence and/or mortality, particularly when national registry data are unavailable.^{5,6} Most of these studies, however, have been focussed largely on the United States (US).⁵⁻⁷ Further, some studies have used potentially inappropriate statistical methods,^{5,6} neglecting to consider the auto-correlated nature of the time series search data (i.e., the dependency of observed search activity on previous search activity).⁸ One such study found that the search term "skin cancer", but not "melanoma", was correlated with melanoma mortality, but not incidence in the US.⁵ Another study, also limited to the US and not correcting for autocorrelation, found the "melanoma" search term to be weakly associated with both melanoma incidence and mortality.⁶ Without appropriate control terms independent from cancer terms, the correlations observed in both cases could be spurious, as any number of other non-cancer related terms might have also demonstrated a similar correlation (for example, Google searches in general show seasonal patterns that would correlate generically with multiple phenomena with the same seasonal pattern). In order to address the concerns raised with previous studies, we sought to examine correlations of internet search behaviour and melanoma incidence in the US and other majority native English-

speaking countries, accounting for auto-correlation in our analysis as well as assessing the relative specificity of cancer search terms.

RESULTS

Correlations of SVI and melanoma incidence in the US states between 2011 and 2014 are shown in Table 1. Overall, no search terms were strongly correlated with melanoma incidence. The strongest correlation observed was 0.31 (95% confidence interval (CI), -0.19 to 0.43) with "Melanoma". None of the correlations achieved statistical significance.

Correlations of SVI over time with melanoma incidence in English speaking countries is shown in Table 2. The strongest correlation was observed with the "Melanoma" search term (0.52 (95% CI, -0.43 to 0.89)) and again, none of the correlations achieved statistical significance.

DISCUSSION

In this study we found that no Google search terms we tested were significantly, or strongly correlated with melanoma incidence in 6 countries. Using a mixed model, we have accounted for the auto-correlation in the time series search data and by including control search terms, we were able to assess inferential relationships between search terms and melanoma incidence. Indeed, given that all of the results were non-significant, we were unable to confirm previously published findings presented by Wehner et al.,⁶ however, these results do corroborate those of Bloom et al.⁵ in that at the US state level, skin cancer/melanoma searches do not correlate with melanoma incidence. Such differences might have occurred in the former case due to a differing time period used; Google SVI is standardised to the maximum value of search volume in the time period considered, meaning that for different but overlapping time periods, the relative overlapping values for SVI can be drastically different.⁹ It has been suggested that internet search data may be particularly useful where national

¹Centre for Observational Research and Data Sciences, Bristol-Myers Squibb, Uxbridge UB8 1DH, UK; ²Real-World Evidence, Evidera, London W6 8DL, UK; ³Real-World Evidence, Evidera, Waltham, MA 02541, USA and ⁴Bristol-Myers Squibb, Uxbridge UB8 1DH, UK
Correspondence: Sreeram V. Ramagopalan (sreeram.ramagopalan@bms.com)

Received: 27 June 2018 Revised: 7 August 2018 Accepted: 8 August 2018

Published online: 05 September 2018

Table 1. Correlation of cancer and control search terms against melanoma incidence in US states between 2011–2014

	Cancer term		Control term	
	Skin cancer	Melanoma	European Union	Google
Incidence	−0.049	0.313	−0.004	0.117
95% CI	(−0.298, 0.336)	(−0.190, 0.434)	(−0.424, 0.592)	(−0.267, 0.507)
p-value	0.485	0.596	0.499	0.535

Table 2. Correlation of cancer and control search terms against melanoma incidence for English speaking countries (England, Republic of Ireland, Australia, New Zealand, Canada) between 2004 and 2016

	Cancer term		Control term	
	Skin cancer	Melanoma	European Union	Google
Incidence	−0.032	0.523	−0.028	0.207
95% CI	(−0.657, 0.709)	(−0.434, 0.887)	(−0.389, 0.205)	(−0.0130, 0.480)
p-value	0.981	0.662	0.983	0.874

registry data on cancer are unavailable. As we have found here that there was no strong or significant association between search terms and cancer incidence, using search data as a proxy for estimating cancer incidence therefore appears unwarranted.

As cancer incidence trends are relatively slow to change, Google search data may be more useful for tracking epidemic conditions where incidence can dramatically change over a few weeks or months. The Google flu trend prediction tool was developed on this premise,⁴ but was discontinued as correlation of search terms with flu incidence were not causally related and as such when internet search patterns changed temporally, the flu trend prediction tool was not able to accurately predict disease occurrence.⁴ Improvements to infectious disease prediction have come when search data is combined with traditional surveillance data.¹⁰

In summary, whilst the public health benefits of being able to track disease incidence using internet search volumes would be significant, such as for epidemic prediction and surveillance, more thought is likely needed to do this with rigour to enable robust insights to be obtained.

METHODS

Age-adjusted melanoma incidence averaged for males and females was obtained for the US for 2011–2014,¹¹ England¹² for 2004–16, New Zealand¹³ for 2004–15, Australia¹⁴ for 2004–13, Canada¹⁵ for 2004–16 and the Republic of Ireland for 2004–14.¹⁶ Google Trends¹⁷ quantifies interest in topics at the population level by analysing all search queries for a specific term. Search volume indexes (SVIs) are normalized values based on total searches during a specified period per selected region. We obtained SVIs for each US state or non-US country from January 1, 2004, to January 1, 2017, for the search terms “melanoma” and “skin cancer” (representing a lay term for melanoma). Control search terms were “Google” and “European Union” and SVIs for these were obtained for the same time periods. Control search terms were selected by the authors to determine if unrelated terms might also show similar correlations to the cancer related search terms, to this end the control terms were utilised in exactly the same way as the cancer related search terms. For the US state model, annualised SVIs were available directly from Google with each state contributing a data point for each year. For the English-speaking countries model, we retrieved annual SVI data for each term that was stratified by

month for each country and calculated the average of these 12 monthly values to create an annual estimate in order to match the annual cancer incidence data. Due to the autocorrelation within these values, we derived a mixed model to calculate a correlation coefficient between cancer incidence and SVIs over time. An unstructured covariance matrix was used initially in the models as per Hamlett et al.¹⁸ However, in cases where convergence was not achieved and would not produce viable estimates, we used an autoregressive structure as the next best choice. Bootstrapped 95% confidence limits for the correlation coefficients were estimated and p-values obtained from an asymptotic Fisher test. All analysis was performed using SAS version 9.4 (SAS Institute Inc., Cary, North Carolina).

Code availability statement

SAS version 9.4 was used for all analyses.

DATA AVAILABILITY

The datasets analysed during the current study are publicly available with websites cited for all data.

AUTHOR CONTRIBUTIONS

S.R. is the guarantor of and conceived and designed the study. A.S. led the analysis. All contributed to the analysis and interpretation of the data. L.M. wrote the first draft and all contributed to subsequent drafts and the final paper. Funding This study was supported in part by Bristol Myers-Squibb.

ADDITIONAL INFORMATION

Competing interests: A.S., S.G., A.S., and B.N. are employed by Evidera. L.M., P.D., F. M., and S.R. are employees of Bristol Myers-Squibb.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- McCarthy, D. M. et al. What did you google? Describing online health information search patterns of ED patients and their relationship with final diagnoses. *West. J. Emerg. Med.* **18**, 928–936 (2017).
- Web search engine. Wikipedia (2018).
- A remedy for your health-related questions: health info in the Knowledge Graph. Official Google Blog.
- Lazer, D., Kennedy, R., King, G. & Vespignani, A. The parable of google flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014).
- Bloom, R., Amber, K. T., Hu, S. & Kirsner, R. Google search trends and skin cancer: evaluating the US population's interest in skin cancer and its association with melanoma outcomes. *JAMA Dermatol.* **151**, 903–905 (2015).
- Wehner, M. R., Nead, K. T. & Linos, E. Correlation among cancer incidence and mortality rates and internet searches in the United States. *JAMA Dermatol.* **153**, 911–914 (2017).
- Phillips, C. A. et al. Relationship between state-level google online search volume and cancer incidence in the United States: retrospective study. *J. Med. Internet Res.* **20**, e6 (2018).
- Lospinoso, D. J., Lospinoso, J. A. & Miletta, N. R. The impact of ultraviolet radiation on sunburn-related search activity. *Dermatol. Online J.* **23**, (2017).
- Rogers, S. What is Google Trends data — and what does it mean? *Medium* (2016).
- Yang, S. et al. Advances in using Internet searches to track dengue. *PLoS Comput. Biol.* **13**, e1005607 (2017).
- CDC—United States Cancer Statistics—NPCR. (2018). <https://www.cdc.gov/cancer/npcr/uscs/index.htm>. Accessed 7th July 2018
- ONS. Release Edition Reference Tables. (2010). <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcmm%3A77-223048>. Accessed 1st February 2012
- Cancer data and stats. Ministry of Health NZ. <https://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/cancer-data-and-stats>. Accessed 22nd February 2018
- Australia, C. Cancer statistics. (2013). <https://cancer australia.gov.au/affected-cancer/cancer-statistics>. Accessed 22nd February 2018
- Canadian Cancer Statistics publication—Canadian Cancer Society. <http://www.cancer.ca/en/cancer-information/cancer-101/canadian-cancer-statistics-publication/?region=on>. Accessed 22nd February 2018

16. National Cancer Registry Ireland | Essential information on cancer in Ireland. <https://www.ncri.ie/>. Accessed 22nd February 2018
17. Google Trends. Google Trends. <https://trends.google.com/trends/?geo=US>. Accessed 7th July 2018
18. Hamlett, A., Ryan, L. & Wolfinger, R. On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *SAS Users Group Int. Proc. Stat. Data Anal. Sect.* 1–7 (2004).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018