The Importance of Health Informatics in Public Health during a Pandemic
J. Mantas et al. (Eds.)

doi:10.3233/SHTI200540

Google Trends and Seasonal Effects in Infodemiology: A Use Case About Obesity

Angelo BASTERIS [a,1], Marjan MANSOURVAR [a], Uffe KOCK WIIL [a]
[a] *Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark*

Abstract. Google Trends is a free online tool that provides information on the interest on a topic measured as Google searches in the big data era. In this work, we present methods for detection of trends and seasonal effects in online interest for health topics based on Google Trends data. We present data for the term obesity worldwide and in a convenience sample of countries (Denmark, United Kingdom, USA, Japan) over the last five years. Our analysis shows that, despite obesity being one of the global health challenges, there is a decreasing trend in online interest in the topic. We also observed seasonal effects, with less interest for the topic during December and the summer months. Identifying trends and seasonal patterns in online interest can support resource allocation and planning and allows one to evaluate the effectiveness of global and local outreach efforts on public health.

Keywords. Big data, Infodemiology, Google Trends, Data analysis

1. Introduction

Infodemiology is "the science of determinants of information in the electronic mediums, like the Internet, and in population, with the aim to inform public health and public policy"[1]. Among the relevant data sources, Google Trends is an open online tool that provides information on Internet search data for queries from 2004 onwards.

Google Trends data can serve a variety of applications in healthcare, such as tracking diseases outbreaks [2], understanding public awareness about conditions (e.g. autism, [3]), and the seasonality of muscle-skeletal injuries [4]. The analysis of temporal and seasonal trends in healthcare can be used to support decisions in the planning of staff and resource allocation. However, reporting of methods in Google Trends usage needs to be improved to allow reproducibility of the findings and thus boosting Google Trends' potential for the healthcare sector [5]. In this paper, we present methods for the analysis of Google Trends data to reveal growing or decaying online interest for a (healthcare) topic among a population and seasonal patterns able to reveal whether such interest varies within a solar year. As a use case, we analyzed data for the term "obesity" worldwide and in a convenience sample of countries from the three continents in the Northern hemisphere (Denmark, United Kingdom, USA, Japan). Overall, this article presents the key points that need to be considered to achieve a robust methodological basis for using Google Trends data to analyze online queries, which is essential for ensuring the validity of the results.

[1] Corresponding Author, Angelo Basteris, Assistant Professor, SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute; E-mail: angelobasteris@gmail.com.

2. Methods

The used methodology consists of three overall phases: (1) data collection, (2) data analysis, and (3) data visualization. The first two parts are described below, while the latter is presented as part of the use case results in Section 3.

2.1. Data collection

Google Trends does not offer access to data through an official API, but it allows downloading data in csv comma separated values format. Data are available for intervals ranging from 2004 to present to the last hour, and custom intervals can be specified. Data sampling interval changes with respect to the interval length, according to Table 1. The available data is the Relative Search Volume (RSV), which is the fraction of Google queries containing the used keyword normalized concerning the peak value observed within the chosen interval. The RSV is a measure of interest for a topic. Detailed information on how to use the Trends portal and its capabilities can be found in [6].

2.2. Data analysis

STL (Seasonal Trend decomposition using LOESS) is a filtering procedure that decomposes data in a trend, a seasonal, and a residual component [7].

By using the Mann-Kendall Trend Test (M-K Test) [8] on the trend component, it is possible to reveal whether there is a significantly increasing or decreasing interest in the argument over the chosen interval. Seasonality effects can be observed by running the Kruskal-Wallis test [9] on detrended data. If a significant month effect is found, the Wilcoxon [10] test allows revealing among which months there are significant differences in interest. We implemented this process using Python 3.7 and the *statsmodel* package [11], and the software for this analysis is made publicly available.[2]

2.3. Use case: obesity

We downloaded Google Trends data for the term obesity worldwide and in a convenience sample of countries from the Northern hemisphere (Denmark, United Kingdom, USA, Japan) in the interval between 1-1-2015 and 31-12-2019. These data were analyzed with the statistical tools described above to investigate whether trends and seasonality effects exist in the online interest for obesity.

Table 1. Relation between interval duration and sampling time in Google Trends data

Interval duration	Sampling interval
Past hour	1 minute
Past 4 hour	1 minute
Past day	8 minutes
Past 7 day	1 hour
Past 30 day	1 day
Past 90 day	1 day
Past 12 months	1 week
Past 5 years	1 week
2004-present	1 month
Customized time range	Minimum 1 day

[2] Available for download at https://cutt.ly/googletrends-sdu

3. Results

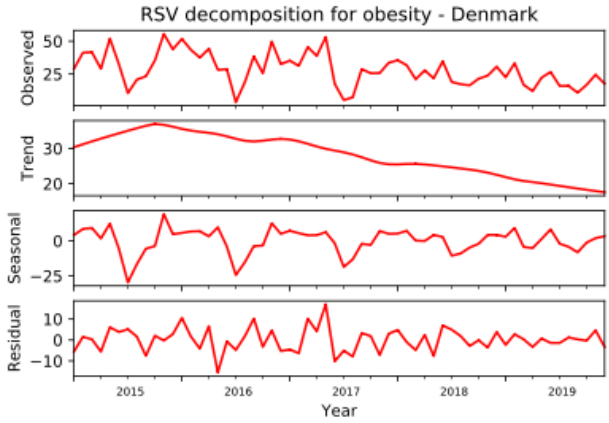

Figure 1. Seasonal trend decomposition for online relative search volume (RSV) about obesity in Denmark. The observed data are decomposed in the sum of a trend, seasonal and residual component.

Figure 1 shows the results of the seasonal-trend decomposition of online interest in obesity in Denmark. It shows that this method allows discovering behaviors (such as the decreasing trend and the periodicity, in this case) that would not be otherwise noted in the observed data. In this case, we used an additive model in which the observed data are decomposed in the sum of trend, seasonal, and residual component.

Figure 2 (left panel) displays trends worldwide and for the four selected countries. Noticeably, online interest for obesity has decreased in all countries except Japan, and changes in the slope of the trend are observed in the period 2017-2018, with a shift of several months across countries. Figure 2 (right panel) shows the seasonal component of RSV worldwide and for countries in our sample. A similar pattern is found across countries, with only Japan exhibiting a significantly different behavior during the first months of the year. Our results show that interest for obesity is increasing in the early months of the year and after the summer holidays. The significance of results for trends and season effects are shown in Table 2.

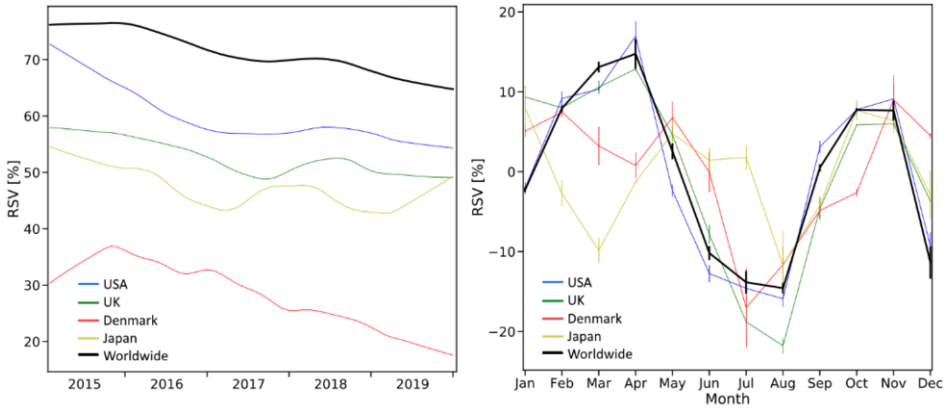

Figure 2. Trend (left) and seasonal component (right) of Relative Search Volume (RSV) about *obesity* in the last five years, worldwide and across countries in our convenience sample. We observed significantly decreasing trends and strongly significant seasonal effects – see Table 2 for a statistical analysis of these values. Errorbars in right panel show standard deviation of RSV within these months in 2015-2019.

Table 2. Significance of trends and seasonality effects in online interest for obesity

Location	Sampling Interval Trend	Seasonality
Worldwide	Decreasing (p<.001)	Yes (p<.001)
USA	Decreasing (p<.001)	Yes (p<.001)
United Kingdom	Decreasing (p<.001)	Yes (p<.001)
Denmark	Decreasing (p<.001)	Yes (p=.006)
Japan	No	Yes (p=.026)

4. Discussion and Conclusions

Despite obesity being one of the global health challenges, we observed a decrease in online interest for it in most of the countries we analyzed. This information should be used to observe whether the decreased interest corresponds to a lower burden for obesity or if, instead, newer solutions to gauge public interest need to be found. Changes in slope as those observed in 2017 show strong effectiveness of communication efforts by WHO (which issued a global warning about obesity in October 2017 [12]). The seasonality patterns of online interest can be used to allocate public initiatives about awareness on health topics during the year, such as the World Obesity Day (regularly held in October).

The main contribution of this paper is the development of a process for using seasonal-trend decomposition on Google Trends data. The process, which can be adapted to different geographical and temporal scales, is demonstrated through a use case, and the software is made publicly available for others. Understanding trends and seasonality of health related online interest can drive new research and support policymakers in a better allocation of health resources and improve the efficiency of communication efforts.

References

[1] 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. 11 (2009). doi:10.2196/jmir.1157.

[2] Pelat C, Turbelin C, Bar-Hen A et al., More diseases tracked by using google trends, Emerg. Infect. Dis. (2009). doi:10.3201/eid1508.090299.

[3] DeVilbiss EA, Lee BK, Brief Report: Trends in U.S. National Autism Awareness from 2004 to 2014: The Impact of National Autism Awareness Month, J. Autism Dev. Disord. (2014). doi:10.1007/s10803-014-2160-4.

[4] Dewan V, Sur H, Using google trends to assess for seasonal variation in knee injuries, J. Arthrosc. Jt. Surg. (2018). doi:10.1016/j.jajs.2018.02.002.

[5] Nuti SV, Wayda B, Ranasinghe I et al., The use of google trends in health care research: a systematic review., PLoS One. 9 (2014) e109583. doi:10.1371/journal.pone.0109583.

[6] Mavragani A, Ochoa G, Google Trends in Infodemiology and Infoveillance: Methodology Framework., JMIR Public Heal. Surveill. 5 (2019) e13439. doi:10.2196/13439.

[7] Cleveland RB, Cleveland WS, McRae JE et al., STL: A seasonal-trend decomposition procedure based on loess, J. Off. Stat. 6 (1990) 3–73. doi:citeulike-article-id:1435502.

[8] Hamed KH, Ramachandra Rao A, A modified Mann-Kendall trend test for autocorrelated data, J. Hydrol. (1998). doi:10.1016/S0022-1694(97)00125-X.

[9] McKight PE, Najab J, Kruskal-Wallis Test, in: Corsini Encycl. Psychol., 2010. doi:10.1002/9780470479216.corpsy0491.

[10] Ghean EA, A generalized Wilcoxon test for comparing arbitrarily singly-censored samples, Biometrika. 52 (1965) 203–23

[11] Seabold S, Perktold J, Statsmodels: Econometric and Statistical Modeling with Python, PROC. 9th PYTHON Sci. CONF. (2010).

[12] Bentham ., Di Cesare M, Bilano V et al., Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016, Lancet. (2017). doi:10.1016/S0140-6736(17)32129-3.