https://doi.org/10.5272/jimab.2021271.3493

Journal of IMAB - *Annual Proceeding (Scientific Papers).* 2021 Jan-Mar;27(1)

Journal of IMAB
ISSN: 1312-773X
https://www.journal-imab-bg.org

Original article

INTERNET BASED DATA AS A POWERFUL TOOL FOR TRACKING INFECTIOUS DISEASE DYNAMICS

Tatina T. Todorova, Gabriela S. Tsankova, Neli M. Ermenlieva
Department of Microbiology and Virology, Faculty of Medicine, Medical University Varna, Bulgaria.

ABSTRACT

Purpose: Digital epidemiology (using Internet data as an epidemiological tool) is one of the most concurrent and novel approaches in the field of infectious disease tracking. We aimed to see if the disease-related information seeking via one of the most popular search engine (Google) correlates with the real varicella epidemiology in Bulgaria.

Material/Methods: We compared the relative search index obtained by Google Trends with a keyword 'varicella' and the real incidence data for 2016-2018 in Bulgaria. Correlation analysis and t-test were used to see for association and difference in the means. Additionally, we summarized what type of information people find when searching for 'varicella'.

Results: Dynamics of Google Trends with search term 'varicella' followed almost exactly the real dynamics of reported varicella cases in Bulgaria (Pearson's correlation coefficient of 0.85). The seasonality and the regional spread of varicella infection were also confirmed by the digital epidemiology (Pearson's correlation coefficient between the search index and the monthly cases was 0.93 and between the search index and the individual district incidence was 0.35). The majority of the information found by people during the seasonal varicella outbreaks was scientifically relevant and correct.

Conclusions: Digital data received by non-epidemiological methods could be successfully used to study varicella dynamics, as high correlation exists between the data collected via the traditional methods and via the new, digital methodology.

Keywords: digital epidemiology, chicken pox, notifiable diseases, Google Trends,

INTRODUCTION

Traditional (classical) epidemiology works with data collected by obligatory or voluntary reporting of disease cases to the respective regional, national and/or international public health authorities. Although well working for years and most of the notifiable diseases, this system could be questionable in times of emergency when sudden epidemic outbreaks happen, or unknown non-notifiable disease appears. Recent examples include the Ebola virus disease outbreak in West Africa (2013-2016) and the on-going COVID-19 pandemic. In such a situation, a significant delay and inaccuracies in infectious disease reporting may exist because of the not prepared system. Therefore, there is a critical need for new, alternative and real-time methods for gathering relevant epidemiological data.

A decade ago, the interest of researchers had focused on the possible use of continuously growing Internet data generated by the search of information or sharing of experience, especially when people feel threatened of some communicable disease. Such data could be search queries, social media posts, webpage access logs and mobile phone network data[1]. The information retrieved via these tools represents the basis of the digital epidemiology, which "uses data that was generated outside the public health system, i.e. with data that was not generated with the primary purpose of doing epidemiology".[2]

The digital epidemiology seems simple, widely available and almost free approach. However, the major challenge is the validation of the data. We are still in the beginning, and the successful implementation requires the gathering of as much as possible data and comparing against valid epidemiological data generated with the classical methods. In this context, the varicella is one of the most "convenient" diseases – first, it is an obligatory/voluntary notifiable infection in a high number of countries and notification data exist for long periods; second, it is present as endemic in all regions but has a well-defined seasonality and locality, therefore an evident searching pattern should be expected.

Recently, the information repository Google Trends has been used to examine the digital epidemiology of varicella in 36 countries for over 11 years. [3] Varicella dynamics retrieved from the digital data showed high correlation with the real data and the seasonal and geographic distribution of the disease was found to be related to the information-seeking behavior of people. [3]However, Bulgaria was not included in the aforementioned analysis. Thus, we aimed to investigate this promising new epidemiological tool in the case of Bulgaria by comparing the existingvaricella incidence data for the last three years and the disease-related searching behavior of the population. Additionally, we also try to analyze what type of information people find when searching for varicella information on the Internet.

MATERIALS AND METHODS

Google Trends is a free tool provided by Google Inc. It summarizes the overall/regional Internet searchingof keywords on specific topics. Google Trends shows the frequency of each search term (or combination of terms) entered into the search engine in relation to the total search volume over a defined period. It represents the results for the search of a given term as search volume index ranging from 0 to 100.

The number of varicella cases in Bulgaria is freely available on a monthly base from the National Center of Public Health and Analyses (NCPHA), Ministry of Health. [4]Varicella is a compulsorily notifiable disease in Bulgaria – all medical practitioners should report possible, probable and confirmed cases on a daily base to the Regional Health Inspectorate of the corresponding district. Cases are then summarized and reported to the national bodies. The NCPHA aggregates the data and publishes the monthly and annual incidence rates at the district and country level.

We collect the search volume indexes for the term 'varicella' (in Bulgarian) for each of the last three years (2016-2018).The Google Trends data are given on a weekly base, whereas varicella cases in Bulgaria are summarized on a monthly base. To deal with this issue, we used the strategy described. [3]In brief, we converted the weekly trends to monthly data by repeating the weekly values at daily intervals. We then assigned the daily values to their appropriate month of the year. For each month, we calculated the mean of the daily values, which were used for further analyses.Google Trends allows this analysis to be performed for the whole territory of Bulgaria, as well as shows data for the separated districts – the district with the highest searching scores 100 and the other regions are ranged accordingly to the frequency of 'varicella' searches. The Google Trends results were obtained and analyzed in April 2019.

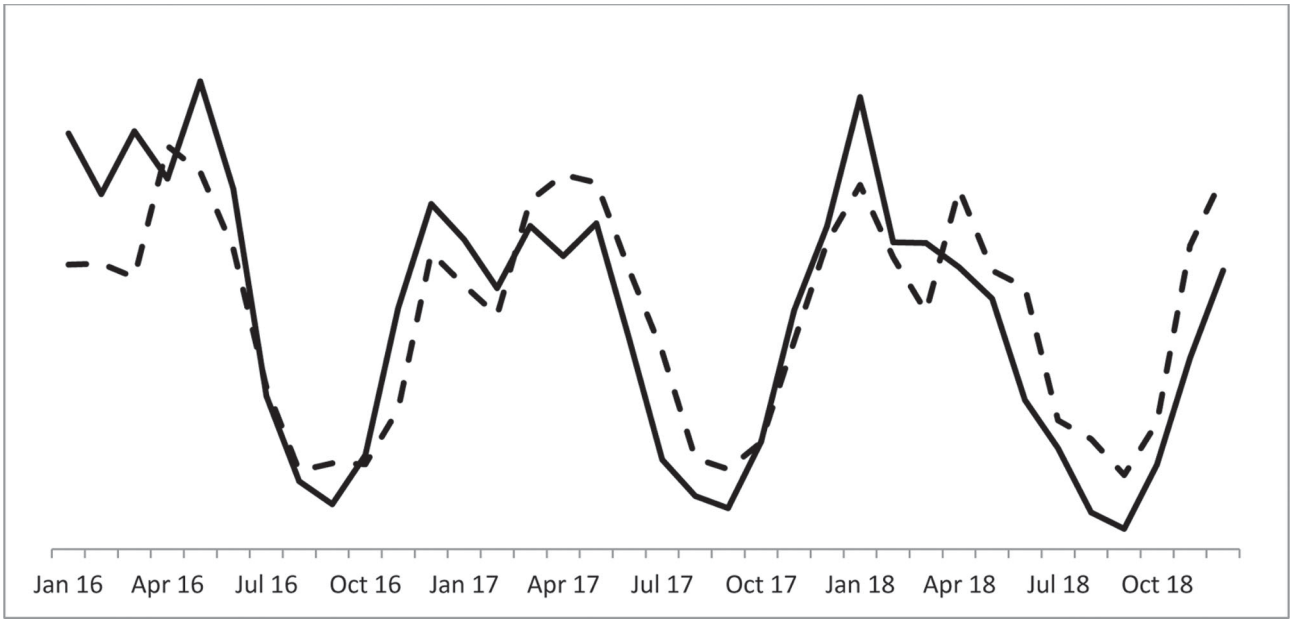
To test for correlation with the real incidence data and to test for difference in the means, Pearson's correlation coefficient and t-test were used as statistical tools in Excel. P-value of 0.05 was considered as statistically significant.

To test what people find when searching on the web, we performed a Google search with the term 'varicella' on 07.06.2019. The term was entered in Bulgarian, and the first 3 pages (30 results) were analyzed and carefully scrutinized for the quality of the information present.

RESULTS

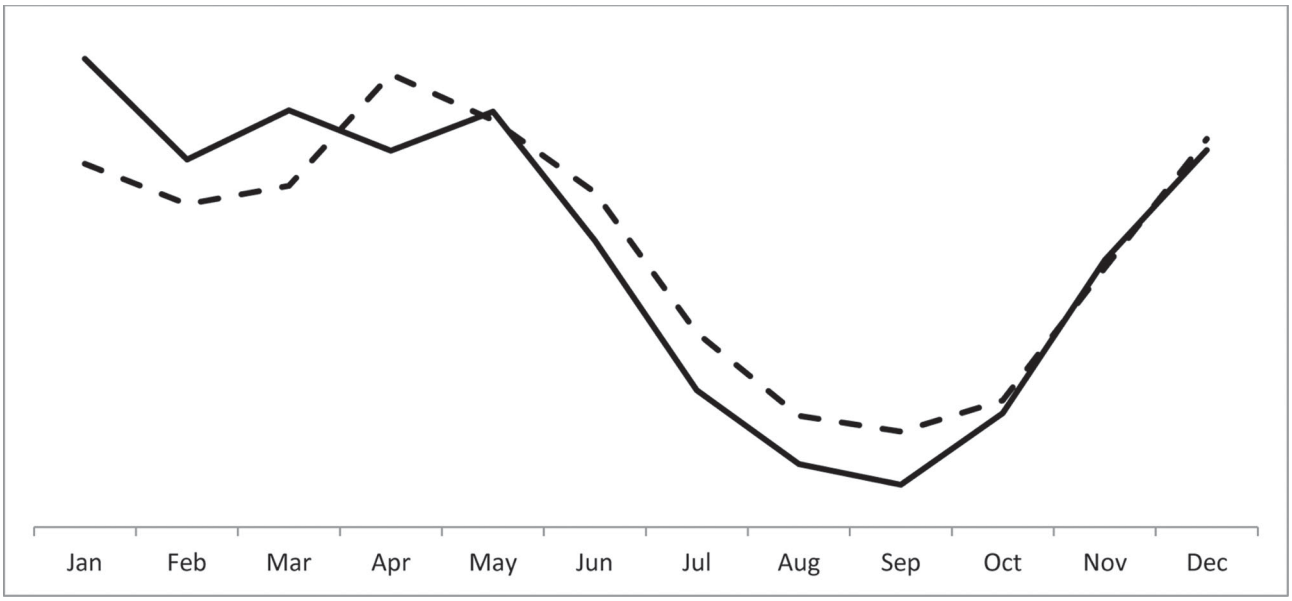
Dynamics of Google Trends with search term 'varicella' (in Bulgarian) followed almost exactly the real dynamics of reported varicella cases for the period 2016-2018. The calculated Pearson's correlation coefficient between the two sets of data was 0.85 ($p <<< 0.001$) (Figure 1).

Fig. 1. Three-year dynamics of reported varicella distribution (solid line, asnumber of cases per month) and 'varicella' Google searches (dashed line, as relative search index) in Bulgaria (2016-2018).



The well-documented seasonality of varicella infection was also confirmed by the digital epidemiology – most of the searches are during the Winter-Spring season when the majority of the disease cases have been reported, and the least searches for varicella were during the summer months when virus transmission was almost negligible (Figure 2).Pearson's correlation coefficient between the search index and the seasonality was 0.93 ($p <<< 0.001$).

Fig. 2. Annual seasonality of varicella distribution (solid line, asnumber of cumulative cases per month for the three years) and 'varicella' Google searches (dashed line, as average of the relative search indexes for the three years of interest) in Bulgaria (2016-2018).



Bulgarian districts have different varicella incidences (even differences of more than 10-fold in the mean annual incidence rates exist among the districts). [5] The search queries in the different districts did not follow exactly the true varicella distribution, however, a moderate positive correlation was found (Pearson's correlation coefficient of 0.35 (p = 0.07). Most of the Google searches were concentrated in the 14 largest regions in which the mean of the Google Trends values was 75.57±17.49 while in the other 14 regions of Bulgaria it was 60.54±19.25, p = 0.04.

These results clearly show that people are looking for disease-related information on the Internet, but it is also important to know what they find when searching. We performed a Google search with the term 'varicella' during a month when the seasonal outbreak was already calming down. The first 30 results (the first three pages) were analyzed. Nineteen out of 30 results were articles containing general information about the symptoms, epidemiology and treatment of the disease; three were press releases with current epidemiological data; two were advertisements of symptomatic remedies; two were leaflets of approved vaccines against varicella; one was an article containing information for the planned implementation of the vaccine in Bulgaria; one was information about varicella infection of a leading anti-vaccer; one was a forum with opinions how to relieve symptoms in children, and one was a petition against the planned implementation of anti-varicella vaccine in Bulgaria. We should mention that the leading result was the official information about the disease on the Ministry of Health site and most of the other Google results were also relevant – 26 of them were with clearly stated authors or were issued by reputed medical laboratories/associations/sites. Even the petition against the obligatory varicella vaccination, although anonymous contained scientific-based information and links to scientific (PubMed) publications reporting negative side effects and complications.

DISCUSSION

In the modern digital world, people prefer to search for information (including medical information) on their phones/computers rather call a doctor or other specialist for advice. Currently, in case of an epidemic outbreak (seasonal or unpredictable), the most innate and simplest behavior is to search and to post via the web and/or social media. The frequency of searching and posting could be used as a good approximation for the actual spreading of an infectious disease. [6]However, together with the ethical considerations[7], the validation of the retrieved data is an important challenge.

The current study validates digital data for the 3-year frequency of varicella infection (2016-2018) against the real epidemiological data for the disease incidence in Bulgaria. We showed that the generated Google Trends search indexes reflected at a high degree the real incidence of varicella in Bulgaria (positive correlation of 85%). A similarpositive association was received by Bakker et al. [3]where Internet-related searching behavior was found to correlate well with the original epidemiology of varicella, especially in countries without massive anti-varicella immunization. We also found an association between the individual Google searching and the temporal and spatial pattern of varicella distribution in Bulgaria – the seasonality of the varicella was almost identicalvia the traditional and the digital approaches, whilethe correlation between the district incidence and the regional Google Trends data was less visible but also present.

Finally, in this work, we review the information obtained when searching in Google with a keyword 'vari-

cella' (in Bulgarian). Based on our perception, the majority of the information found by people was scientifically relevant and correct. We should point out that our study was performed in a period when the possible recommended implementation of the anti-varicella vaccine has been discussed in Bulgaria. However, we obtained only one 'anti-vaccer' result in the 30 leading Google results, and it was scientifically based – false statements and popular anti-vaccer beliefs were not present – it was a petition against the planned vaccine usage with a list of scientific publications for cases with adverse effects and reactions after the vaccination.

CONCLUSIONS

Digital data received by non-epidemiological methods could be successfully used to study varicella dynamics, as high correlation exists between the data collected via the traditional methods and via the new, digital methodology.

REFERENCES:

1. Park H-A, Jung H, On J, Park SK, Kang H. Digital Epidemiology: Use of Digital Data Collected for Non-epidemiological Purposes in Epidemiological Studies. *Healthc Inform Res.* 2018 Oct;24(4):253–62. [PubMed]
2. Salathe M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc policy.* 2018 Jan 4; 14(1):1. [PubMed]
3. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proc Natl Acad Sci U S A.* 2016 Jun 14;113(24): 6689-94. [PubMed]
4. National Center of Public Health and Analyses. Public Health Statistics Annual Bulgaria. 2019. [Internet].
5. Todorova TT. Varicella infection in a non-universally vaccinated population: Actual epidemiology in Bulgaria (2013–2015). *J Infect Public Health.* 2018 May-Jun;11(3):326-30. [PubMed]
6. Velasco E. Disease detection, epidemiology and outbreak response: the digital future of public health practice. *Life Sci Soc Policy.* 2018 Apr 1; 14(1):7. [PubMed]
7. Samerski S. Individuals on alert: digital epidemiology and the individualization of surveillance. *Life Sci Soc Policy.* 2018 Jun 14;14(1):13. [PubMed]

Please cite this article as: Todorova TT, Tsankova GS, Ermenlieva NM. Internet based data as a powerful tool for tracking infectious disease dynamics. *J of IMAB.* 2021 Jan-Mar;27(1):3493-3496.
DOI: https://doi.org/10.5272/jimab.2021271.3493

Received: 26/03/2020; Published online: 05/01/2021

Address for correspondence:
Tatina Todorova
Department of Microbiology and Virology, Faculty of Medicine, Medical University Varna,
3, Bregalniza Str., 9002 Varna, Bulgaria
E-mail: Tatina.Todorova@mu-varna.bg