Infodemiology of systemic lupus erythematous using Google Trends

(Article begins on next page)

Infodemiology of Systemic Lupus Erythematous using Google Trends

Massimo Radin and Savino Sciascia

Center of Research of Immunopathology and Rare Diseases- Coordinating Center of Piemonte and Valle d'Aosta Network for Rare Diseases, Department of Rare, Immunologic, Hematologic and Immunohematologic Diseases, and SCDU Nephrology and Dialysis, S. Giovanni Bosco Hospital

Running Title

Infodemiology of SLE

Corresponding Author:

Savino Sciascia, MD, PhD;

Center of Research of Immunopathology and Rare Diseases- Coordinating Center of Piemonte and Valle d'Aosta Network for Rare Diseases, and SCDU Nephrology and Dialysis, S. Giovanni Bosco Hospital

Piazza del Donatore di Sangue 3, 10154, Turin, Italy.

Email savino.sciascia@unito.it Tel +390112402056 Fax +390112402052

Abstract

Objective

People affected by chronic rheumatic conditions, such as systemic lupus erythematosus (SLE), frequently rely on Internet and search engines to look for terms related to their disease, and its possible causes, symptoms and treatments. "Infodemiology" and "infoveillance" are two recent terms created to describe a new developing approach for public health, based on Big Data monitoring and data mining. In this study, we aim to investigate trends of Internet research linked to SLE and symptoms associated to the disease applying a Big Data monitoring approach.

Methods

We analyzed the large amount of data generated by Google Trends, considering 'lupus', 'relapse' and 'fatigue' in a 10-year web-based research. Google Trends automatically normalized data for the overall number of searches, and presented them as relative search volumes, in order to compare variations of different search terms across regions and periods. The Menn-Kendall test was used to evaluate the overall seasonal trend of each search term and possible correlation between search terms.

Results

We observed a seasonality for Google search volumes for lupus-related terms. In Northern hemisphere, relative search volumes for 'lupus' were correlated with 'relapse' ($\tau=0.85$; $p=0.019$) and with fatigue ($\tau=0.82$; $p=0.003$), whereas in Southern hemisphere we observed a significant correlation between 'fatigue' and 'relapse' ($\tau=0.85$; $p=0.018$). Similarly, a significant correlation between 'fatigue' and 'relapse' ($\tau=0.70$; $p<0.001$) was seen also in the Northern hemisphere.

Conclusion

Despite the intrinsic limitations of this approach, Internet-acquired data might represent a real-time surveillance tool and an alert for healthcare systems in order to plan the most appropriate resources in specific moments with higher disease burden.

Introduction

People with chronic autoimmune conditions (such as systemic lupus erythematous – SLE-, multiple sclerosis, rheumatoid arthritis) frequently use search engines, such as Google, to look for terms related to their disease, and its possible causes, symptoms and treatments[1]. The large amount of information coming from searches conducted through Google generates trend data, and can be analyzed with Google Trends, a publicly available tool that compares the different geospatial and temporal patterns in search volumes for user-specified terms[2]. Therefore, Google Trends can provide indirect estimates of the burden of diseases and symptoms, and, accordingly, has been used for surveillance studies[2]. Despite the enormous potentiality and applicability of Google Trends in different fields of medicine have been noticed only recently, this tool has been already been successfully applied for monitoring infectious diseases (among the others, for tuberculosis and influenza)[3], suicides patterns[4] and epidemiology of drug addiction[5].

Similarly, Google Trends for some diseases have positively been associated with disease prevalence in some studies[6,7]. Different clinical and epidemiological studies have been conducted on seasonality of SLE, and focused on flare, fatigue and periodicity of onset of clinical manifestations. Overall, SLE disease activity has been associated with specific seasonal patterns in both, Northern and Southern hemispheres, possibly in relation to sun exposure, meteorological factors and vitamin D levels[8–11]. Therefore, the present 10-year retrospective web-based research investigated the seasonality of Google Trends search volumes for 'lupus' and for SLE-related search terms, in the Northern and Southern hemispheres.

Methods

Search terms, considered on Google Trends separately, were: 'lupus' (requiring the exact phrase to be present in the search field), 'relapse' and 'fatigue' (http://www.google.com/trends with the following filters: 'Worldwide', in 'All Categories' and for 'Web Searches'). Google Trends automatically normalized data for the overall number of searches, and presented them as relative search volumes (on a scale from 0 to 100), in order to compare variations of different search terms across regions and periods[12]. This approach has been applied and validated elsewhere[6].

In order to reduce the background noise from populations with limited Internet searches and/or with lower relative interest in Lupus, countries were selected according to the following criteria: (1) presence of a relative search volume >50 for 'lupus'; (2) presence of at least 20 million inhabitants. Canada, France, Italy, the UK and the US, were included from the Northern hemisphere, and Australia, Argentina, Brazil and Perù, from the Southern hemisphere. For those countries where English is not the first language, additional analyses were conducted by including search terms in the local language (language-by-language search is provided as additional material). For statistical purposes, countries were grouped in relation to the hemisphere (Northern or Southern). Relative search volumes for 'lupus', 'relapse' and 'fatigue', were extracted from January 2006 to January 2016, and grouped for each month of observation. The terms "relapse" and "fatigue" were used in a *combined* search with the term "lupus" in order to increase the accuracy of our analysis for lupus-related search volumes".

The Menn-Kendall test was used to evaluate the overall seasonal trend of each search term; results are presented as τ coefficients, that express the measured ordinal association between research terms. Generalised linear models were employed to assess possible differences in relative search volumes for 'lupus', 'relapse' and 'fatigue',

between months of the year in each hemisphere; results are presented as rate ratio (adjusted on the median) and 95% CIs. Finally, possible correlations between different search terms were evaluated with the Kendall test; results are presented as τ coefficients. SPSS (IBM SPSS Statistics for Windows, Version 19.0 Armonk, NY: IBM Corp) has been used for data processing and analysis. Results have been considered statistically significant if $p < 0.05$.

Results

We observed a seasonality for Google relative search volumes for lupus-related terms (Figure 1). In the Northern hemisphere, 'Lupus'-relative search volumes presented the highest levels in May and June (Late Spring Time) and in January and February (Winter Time); the lowest levels are observed in November and December (Winter Time). Specular to this patter, in the Southern hemisphere, highest levels of 'Lupus'-relative search volumes were seen in May and June (Winter Time) and in January and February (Summer Time); the lowest levels were observed in July and August (Winter Time).

'Relapse'-relative search volumes presented higher levels in April and May (Spring) and January and February (Winter) in the Northern hemisphere; lower levels in June and July (Summer). In the Southern hemisphere, highest levels of 'relapse'-relative search volumes were seen in April and June (Autumn/Winter); the lowest levels are observed in February and December (Summer Time).

'Fatigue'-relative search volumes presented higher levels in May and June (Late Spring) and lower levels in November and December (Late Autumn) in the Northern hemisphere. In the Southern hemisphere, highest levels of 'fatigue'-relative search volumes were seen in March and October; the lowest levels are observed in June and July (Winter Time).

In Northern hemispheres, relative search volumes for 'lupus' were correlated with 'relapse' ($\tau=0.85$; $p=0.019$) and with fatigue ($\tau=0.82$; $p=0.003$), whereas in Southern

hemisphere we observed a significant correlation between 'fatigue' and 'relapse' (τ=0.85; p=0.018). Similarly , a significant correlation between 'fatigue' and 'relapse' (τ=0.70; p<0.001) was seen also in the Northern hemisphere.

Discussion

The terms "infodemiology" and "infoveillance" were coined by Gunther Eysenbach when describing a new developing approach for public health[13].[14], based on Big Data monitoring and data mining, within the conceptual framework of e-health and health Web 2.0[15,16]. Albeit limitations and concerns still exist, one could not ignore that the Internet and the new medical informatics tools are defining new directions in the field of the epidemiological research, indicating new trends and strategies.

Google Trends search volumes are commonly found to be increased for conditions with higher social impact or, at least, for periods with higher disease burden and, thus, are gaining importance in surveillance studies on the most epidemiologically relevant diseases[2].

Google Trends for 'lupus' presented a specific seasonality, with peaks in Spring and Winter in both hemispheres, possibly as a result of seasonal variations in the disease burden. Interestingly, in the Northern hemisphere, relative search volumes for 'lupus' and 'relapse' were positively associated, and presented a similar annual cyclical sinusoidal pattern. Similar results were observed for 'lupus' and 'fatigue'. Those trends might reflect what previously shown in clinical studies evaluating the occurrence of relapses across different seasons. Among others, Chiche et al. reported a seasonal pattern for lupus flares among patients living in Southern France, with most flares in spring, in correlation with an increase in temperature and duration of sunshine[10]. On the other hand, Dall'Ara reported the occurrence of winter lupus flares in associated with lower vitamin D levels[8]. Patients with SLE are prone to hypo-vitaminosis D because of their photosensitivity. Vitamin D has beneficial effects not only on bone

metabolism but also on the function of the immune system. Thus, once could speculate that vitamin D insufficiency detected in the wintertime can be either a predisposing factor for flare or the consequence of the flare itself in SLE patients.

The possible heterogeneity observed comparing the results between Northern and Southern hemispheres could be justified by the difference among Internet users in the different countries, being USA and Europe the geographical areas with more search volume[17].

The main limitation of the present study is that search trends might be produced by people other than patients with SLE, who are nevertheless interested in this topic. Furthermore, available data are clearly limited to Google users, and are related to the possibility to use a computer with Internet access. From that perspective, however, it should be noted that sixty-five percent of patients with SLE have disease onset between the ages of 16 and 55[18], distribution that mirrors the age-frame of the majority of Google users[19]. Therefore, a non-representative sampling bias might have occurred due to different factors, such as age, disability, income or preferred search engine.

In conclusion, the current study provided additional evidence for seasonality of lupus by using Google Trends. Indeed, Internet-acquired data might represent a real-time surveillance tool and an alert for healthcare systems in order to plan the most appropriate resources in specific moments with higher disease burden. As a future perspective, lupus-related Google Trends should be validated with external clinical data sets.

Legend

Figure 1:

Profile plots for relative search volumes. Profile plots show the relative search volumes for 'lupus' (A and D), 'relapse' (B and E) and 'fatigue' (C and F), in the Northern and Southern hemispheres, respectively, for different months of the year, with estimates sinusoidal curves.

References

1   Brigo F, Lochner P, Tezzon F, *et al.* Web search behavior for multiple sclerosis: An infodemiological study. *Mult Scler Relat Disord* 2014;**3**:440–3. doi:10.1016/j.msard.2014.02.005

2   Nuti S V, Wayda B, Ranasinghe I, *et al.* The use of google trends in health care research: a systematic review. *PLoS One* 2014;**9**:e109583. doi:10.1371/journal.pone.0109583

3   Dukic VM, David MZ, Lauderdale DS. Internet queries and methicillin-resistant Staphylococcus aureus surveillance. *Emerg Infect Dis* 2011;**17**:1068–70. doi:10.3201/eid/1706.101451

4   Gunn JF, Lester D. Using google searches on the internet to monitor suicidal behavior. *J Affect Disord* 2013;**148**:411–2. doi:10.1016/j.jad.2012.11.004

5   Deluca P, Davey Z, Corazza O, *et al.* Identifying emerging trends in recreational drug use; outcomes from the Psychonaut Web Mapping Project. *Prog Neuro-Psychopharmacology Biol Psychiatry* 2012;**39**:221–6. doi:10.1016/j.pnpbp.2012.07.011

6   Moccia M, Palladino R, Falco A, *et al.* Google Trends: new evidence for seasonality of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2016;**87**:1028–9. doi:10.1136/jnnp-2016-313260

7   Holmberg M, Murtonen A, Elovaara I, *et al.* Increased Female MS Incidence and Differences in Gender-Specific Risk in Medium- and High-Risk Regions in Finland from 1981-2010. *Mult Scler Int* 2013;**2013**:182516. doi:10.1155/2013/182516

8   Dall'Ara F, Andreoli L, Piva N, *et al.* Winter lupus flares are associated with low vitamin D levels in a retrospective longitudinal study of Italian adult patients. *Clin Exp Rheumatol*;**33**:153–8.

9   Duarte-García A, Fang H, To CH, *et al.* Seasonal variation in the activity of

systemic lupus erythematosus. *J Rheumatol* 2012;**39**:1392–8. doi:10.3899/jrheum.111196

10   Chiche L, Jourde N, Ulmann C, *et al.* Seasonal variations of systemic lupus erythematosus flares in southern France. *Eur J Intern Med* 2012;**23**:250–4. doi:10.1016/j.ejim.2011.12.006

11   Zhang H, Xu S, Tang D, *et al.* Seasonal distribution of active systemic lupus erythematosus and its correlation with meteorological factors. 2011;**66**:1009–13. doi:10.1590/S1807-59322011000600015

12   Tyack PL, Calambokidis J, Friedlaender A, *et al.* Formal Comment on Schorr GS, Falcone EA, Moretti DJ, Andrews RD (2014) First Long-Term Behavioral Records from Cuvier's Beaked Whales (Ziphius cavirostris) Reveal Record-Breaking Dives. PLoS ONE 9(3): e92633. doi:10.1371/journal.pone.0092633. *PLoS One* 2015;**10**:e0142287. doi:10.1371/journal.pone.0142287

13   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. doi:10.2196/jmir.1157

14   Eysenbach G, Eysenbach G, Eysenbach G, *et al.* Infodemiology and Infoveillance. *Am J Prev Med* 2011;**40**:S154–8. doi:10.1016/j.amepre.2011.02.006

15   Burkow TM, Vognild LK, Krogstad T, *et al.* An easy to use and affordable home-based personal eHealth system for chronic disease management based on free open source software. *Stud Health Technol Inform* 2008;**136**:83–8.

16   Giustini D. How Web 2.0 is changing medicine. *BMJ* 2006;**333**:1283–4. doi:10.1136/bmj.39062.555405.80

17   No Title.

18   N R. Clinical features of systemic lupus erythematosus. In: *Textbook of Rheumatology, Kelley WN, Harris ED, Ruddy S, Sledge CB (Eds), WB Saunders,*

Philadelphia. 1981.

19   (http://online-behavior.com/analytics/demographics).