

LETTER

Comparison of epidemiologic surveillance and Google Trends data on asthma and allergic rhinitis in England

To the Editor,

Epidemiologic surveillance is the 'ongoing systematic collection, analysis, and implementation of health data essential to the planning, implementation and evaluation of public health practices, closely integrated with the timely dissemination of these data'.¹ Recently, the use of data from internet users' activity—measured with tools such as Google Trends (GTs)—has been proposed to complement 'classical' surveillance approaches.² In this context, infectious diseases have been widely studied,² but allergic diseases have not, as they are less often targeted by surveillance programmes. In England, however, the Oxford-Royal College of General Practitioners Research Surveillance Centre (RSC) monitors asthma, allergic rhinitis (AR) and respiratory infections, publishing weekly surveillance reports on their incidence. Such reports are based on 'the number of patients [of all ages] consulting [a set of regionally representative primary care centres in England] with new episodes of illness', differentiating incident from follow-up consultations³; methods for ensuring data quality are routinely applied.³ English general practice is a highly computerized registration-based system, with computerized medical records following the patient when they move practice, ensuring a reliable rate of disease to be reported.

In this study, we assessed the correlation of English surveillance and GTs data on asthma, rhinitis and upper respiratory tract infections (URTI) to assess whether similar patterns can be observed with these two different approaches.

We retrieved weekly surveillance data from the RSC reports on the incidence in England per 100,000 inhabitants of 'AR', 'asthma', 'URTI' and 'common cold' (a subset of URTI) for a period of 4 years (January 2016–January 2020).³

We then used GTs to retrieve relative search volume data for the topics 'AR', 'pollen', 'asthma' and 'common cold' (topics are groups of search terms concerning the same concept) for England during the same time period. Some of these topics had been previously used.⁴ Weekly surveillance and GTs data were correlated using Spearman correlation coefficients (ρ).

Using English surveillance data, 'asthma' incidence was strongly correlated with 'common cold' incidence ($\rho = 0.800$) and 'URTI' ($\rho = 0.765$), but negatively correlated with 'AR' incidence ($\rho = -0.517$) (Table 1; Figure 1).

GTs data on 'asthma' were moderately correlated with GTs data on 'common cold' ($\rho = 0.556$) and negatively correlated with GTs on 'AR' ($\rho = -0.361$). GTs data on 'AR' and 'pollen' were strongly correlated ($\rho = 0.915$).

GTs data on 'common cold' displayed a strong correlation with English surveillance data on 'asthma' ($\rho = 0.716$), 'common cold' ($\rho = 0.780$) and 'URTI' ($\rho = 0.722$). 'AR' incidence was strongly correlated with GTs data on 'AR' ($\rho = 0.902$) and 'pollen' ($\rho = 0.866$).

This study has some limitations, the most important being its ecological design and the assessment of single-country data. We analysed 2016–2019 data, as the COVID-19 pandemic could have influenced our results (in March 2020, there was a media-driven GTs peak for asthma,⁵ and increased inhalers prescriptions). However, we obtained similar results when analysing 2016–2021 data (Table S1).

Correlation does not imply causation, so we must be careful when interpreting results. Nevertheless, applying the Granger causality test for positive correlations (which assesses whether one time series is useful in forecasting another), we mostly observed biologically plausible results, with time series on URTI/common cold being useful to forecast asthma, and asthma time series not being useful to forecast URTI/common cold (Table S2).

The incidence of 'asthma' was strongly and positively correlated with that of 'URTI', but negatively correlated with that of 'AR' and with GTs data on 'pollen'/'AR'. These results are mostly explained by the different seasonality of AR and asthma, and suggest that, even though AR and pollen exposure may be linked to worsening asthma, viral infections may be a more relevant trigger for asthma exacerbations (as previously shown using GTs for 'common cold' and asthma hospitalizations in five countries⁶).

Strong correlations were also observed between surveillance and GTs data on 'AR' incidence, suggesting the potential of GTs as a complementary tool in rhinitis monitoring.

On the other hand, GTs data on 'pollen'/'AR' displayed negative correlations with GTs and surveillance data on 'common cold', suggesting that patients and primary care teams may know how to distinguish seasonal allergies from URTI (which are less common during pollen season) despite some common symptoms.

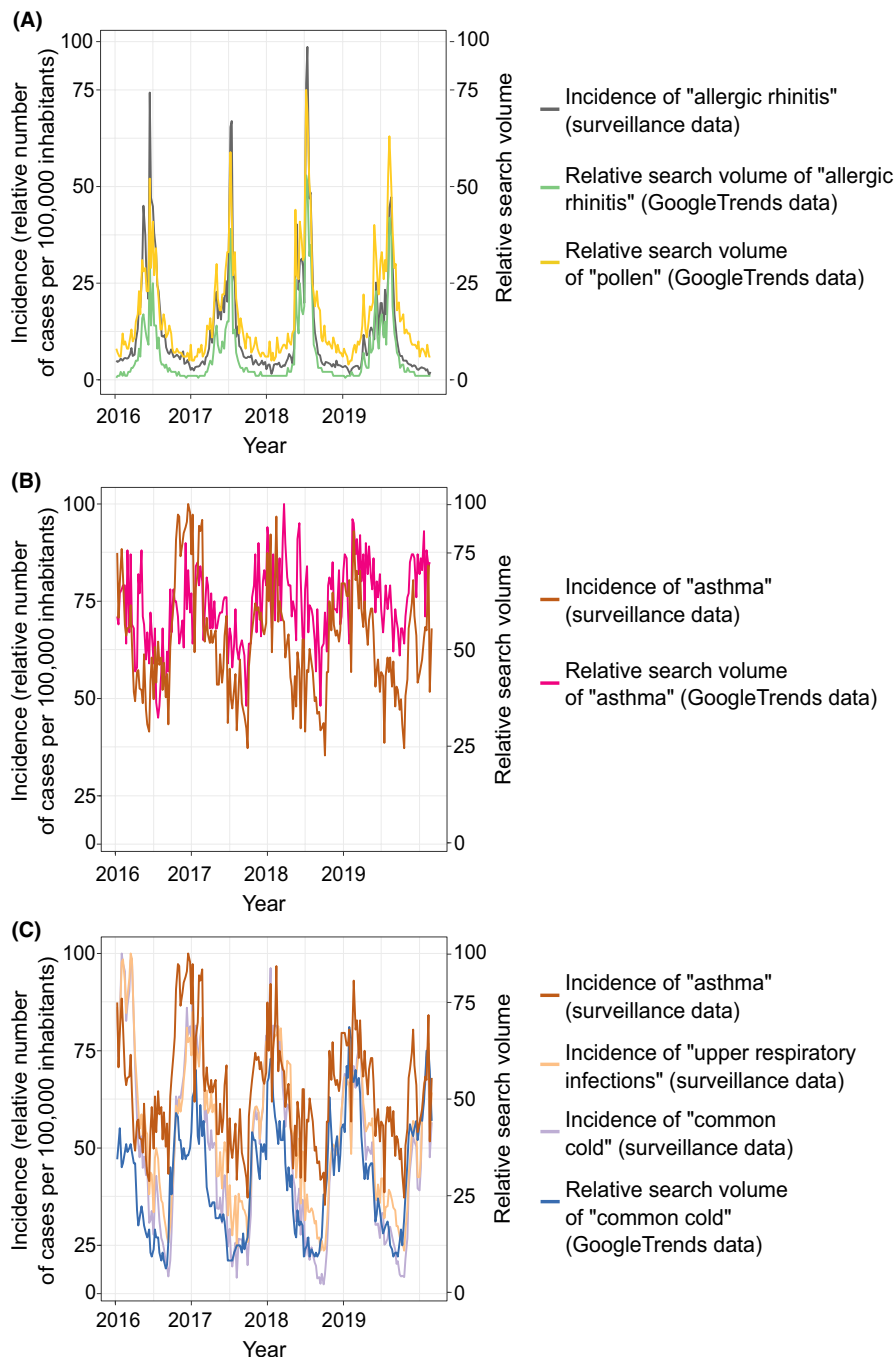
Taken together, these results point to the complementarity of classical surveillance and of GTs data (even though GTs does not

TABLE 1 Spearman correlation coefficients [95% confidence intervals] between surveillance data (cases per 100,000 inhabitants) and Google Trends data for England (3 January 2016–6 January 2020)

	Surveillance data		Google Trends data			
	Asthma	Upper respiratory tract infections	Common cold	Allergic rhinitis	Asthma	Pollen
Surveillance data						
Upper respiratory tract infections	0.765[0.699;0.820]	–	–	–	–	–
Common cold	0.800 [0.733;0.845]	0.987 [0.980;0.990]	–	–	–	–
Allergic rhinitis	–0.517 [–0.601;–0.420]	–0.517 [–0.586;–0.433]	–0.580 [–0.650;–0.505]	–	–	–
Google Trends data						
Asthma	0.490 [0.380;0.593]	0.474 [0.352;0.573]	0.480 [0.349;0.586]	–0.426 [–0.535;–0.303]	–	–
Common cold	0.716 [0.644;0.773]	0.722 [0.659;0.774]	0.780 [0.729;0.819]	–0.835 [–0.868;–0.795]	0.556 [0.458;0.647]	–
Pollen	–0.639 [–0.706;–0.563]	–0.637 [–0.690;–0.573]	–0.701 [–0.750;–0.642]	0.866 [0.813;0.902]	–0.320 [–0.444;–0.197]	–0.816 [–0.849;–0.774]
Allergic rhinitis	–0.628 [–0.695;–0.549]	–0.634 [–0.697;–0.557]	–0.693 [–0.753;–0.628]	0.902 [0.859;0.930]	–0.361 [–0.474;–0.250]	0.915 [0.875;0.943]

Colour code: Bright green – Spearman correlation coefficient ≥ 0.800 ; Dull green – Spearman correlation coefficient ≥ 0.600 and < 0.800 ; Yellow – Spearman correlation coefficient ≥ 0.400 and < 0.600 ; Bright orange – Spearman correlation coefficient ≥ -0.400 and < -0.600 ; Dull orange – Spearman correlation coefficient ≥ -0.600 and < -0.800 ; Dull pink – Spearman correlation coefficient ≥ -0.800 and < -0.600 ; Bright pink – Spearman correlation coefficient ≤ -0.800 .

FIGURE 1 Surveillance and Google Trends data on allergic rhinitis (A), asthma (B) and upper respiratory infections for England (C) (3 January 2016–6 January 2020)



solely assess incident disease and may be influenced by media coverage⁵). For example, for each topic, GTs displays a list of the top-related queries searched by internet users, possibly helping to context epidemiologic fluctuations and to raise new hypotheses (eg, on exacerbations triggers). In addition, analysis of GTs data may possibly help planning health care needs for allergic respiratory diseases (as in most countries surveillance of asthma or AR does not occur).

KEYWORDS

asthma, epidemiology, respiratory infections, rhinitis, surveillance

ACKNOWLEDGMENT

Open access funding enabled and organized by ProjektDEAL.

CONFLICT OF INTEREST

None declared.

Bernardo Sousa-Pinto^{1,2}

Josep M. Antó^{3,4,5,6}

Aziz Sheikh⁷

Simon de Lusignan^{8,9}

Tari Haahtela¹⁰

João Almeida Fonseca^{1,2}

Jean Bousquet^{11,12,13,14}

¹MEDCIDS – Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

²CINTESIS – Center for Health Technology and Services
Research, University of Porto, Porto, Portugal

³ISGlobal, Barcelona Institute for Global Health, Barcelona,
Spain

⁴Universitat Pompeu Fabra, Barcelona, Spain

⁵IMIM – Hospital del Mar Medical Research Institute, Barcelona,
Spain

⁶CIBER Epidemiología y Salud Pública – CIBERESP, Barcelona,
Spain

⁷Usher Institute, University of Edinburgh, Edinburgh, UK

⁸Nuffield Department of Primary Care Health Sciences,
University of Oxford, Oxford, UK

⁹Royal College of General Practitioners Research and
Surveillance Centre, London, UK

¹⁰Skin and Allergy Hospital, Helsinki University Hospital, and
University of Helsinki, Helsinki, Finland

¹¹Charité, Universitätsmedizin Berlin, Humboldt-Universität zu
Berlin, Berlin, Germany

¹²Department of Dermatology and Allergy, Comprehensive
Allergy Center, Berlin Institute of Health, Berlin, Germany

¹³MACVIA-France, Montpellier, France

¹⁴University Hospital, Montpellier, France

Correspondence

Jean Bousquet, CHU Arnaud de Villeneuve, 371 Avenue du
Doyen Gaston Giraud, 34295 Montpellier Cedex 5, France.

Email: jean.bousquet@orange.fr

ORCID

Bernardo Sousa-Pinto  <https://orcid.org/0000-0002-1277-3401>

Tari Haahtela  <https://orcid.org/0000-0003-4757-2156>

Jean Bousquet  <https://orcid.org/0000-0002-4061-4766>

REFERENCES

1. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev.* 1988;10:164-190.
2. 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(1):e11.
3. Smith NA, Livina VN, Byford R, Ferreira F, Yonova I, de Lusignan S. Automated differentiation of incident and prevalent cases in primary care computerised medical records (CMR). *Stud Health Technol Inform.* 2018;247:151-155.
4. Bousquet J, Agache I, Antó JM, et al. Google Trends terms reporting rhinitis and related topics differ in European countries. *Allergy.* 2017;72(8):1261-1266.
5. Sousa-Pinto B, Heffler E, Antó A, et al. Anomalous asthma and chronic obstructive pulmonary disease Google Trends patterns during the COVID-19 pandemic. *Clin Transl Allergy.* 2020;10(1):47.
6. Sousa-Pinto B, Halonen JI, Antó A, et al. Prediction of asthma hospitalizations for the common cold using Google Trends: infodemiology study. *J Med Internet Res.* 2021;23(7):e27044.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Comparison of epidemiologic surveillance and Google Trends data on asthma and allergic rhinitis in England:

Supplementary material for review

Supplementary Table 1. Spearman correlation coefficients [95% confidence intervals] between surveillance data (cases *per* 100,000 inhabitants) and Google Trends data for England (January 3, 2016 to July 3, 2021)^a

		Surveillance data				Google Trends data	
		Asthma	Upper respiratory tract infections	Common cold	Allergic rhinitis	Common cold	Pollen
Surveillance data	Upper respiratory tract infections	0.847 [0.800;0.879]	-	-	-	-	-
	Common cold	0.841 [0.800;0.875]	0.983 [0.974;0.987]	-	-	-	-
	Allergic rhinitis	-0.263 [-0.372;-0.155]	-0.207 [-0.324;-0.077]	-0.218 [-0.343;-0.091]	-	-	-
Google Trends data	Common cold	0.621 [0.544;0.687]	0.577 [0.492;0.652]	0.561 [0.475;0.643]	-0.730 [-0.777;-0.674]	-	-
	Pollen	-0.586 [-0.662;-0.506]	-0.567 [-0.648;-0.480]	-0.581 [-0.657;-0.485]	0.800 [0.749;0.839]	-0.734 [-0.783;-0.680]	-
	Allergic rhinitis	-0.563 [-0.635;-0.479]	-0.551 [-0.631;-0.466]	-0.570 [-0.650;-0.472]	0.822 [0.775;0.860]	-0.717 [-0.763;-0.661]	0.928 [0.903;0.948]

^a Google Trends data on asthma not presented on account of a large media-driven Google Trends search peak in March 2020; Colour code: Bright green – Spearman correlation coefficient ≥ 0.800 ; Dull green – Spearman correlation coefficient ≥ 0.600 and < 0.800 ; Yellow – Spearman correlation coefficient ≥ 0.400 and < 0.600 ; Bright orange – Spearman correlation coefficient ≥ -0.400 and < -0.600 ; Dull orange – Spearman correlation coefficient ≥ -0.600 and < -0.400 ; Dull pink – Spearman correlation coefficient ≥ -0.800 and < -0.600 .

Supplementary Table 2. Results (*p*-values) of the Granger causality test^a with one lag

Time series forecasting → Time series being forecasted	Pre-pandemic period (January 2016 – January 2020)	Whole period (January 2016 – July 2021)
Upper respiratory tract infections (S) → Asthma (S) ^b	0.471	0.001
Common cold (S) → Asthma (S)	0.023	<0.001
Asthma (GT) → Asthma (S)	0.350	- ^c
Common cold (GT) → Asthma (S)	<0.001	<0.001
Asthma (S) → Asthma (GT)	0.003	- ^c
Upper respiratory tract infections (S) → Asthma (GT)	0.003	- ^c
Common cold (S) → Asthma (GT)	0.001	- ^c
Common cold (GT) → Asthma (GT)	<0.001	- ^c
Asthma (S) → Upper respiratory tract infections (S)	0.509	0.858
Common cold (S) → Upper respiratory tract infections (S)	<0.001	<0.001
Asthma (GT) → Upper respiratory tract infections (S)	0.240	- ^c
Common cold (GT) → Upper respiratory tract infections (S)	<0.001	0.402
Asthma (S) → Common cold (S)	0.514	0.741
Upper respiratory tract infections (S) → Common cold (S)	<0.001	<0.001
Asthma (GT) → Common cold (S)	0.443	- ^c
Common cold (GT) → Common cold (S)	<0.001	0.981
Asthma (S) → Common cold (GT)	0.079	0.283
Upper respiratory tract infections (S) → Common cold (GT)	0.476	0.163
Common cold (S) → Common cold (GT)	0.933	0.073
Asthma (GT) → Common cold (GT)	0.698	- ^c
Allergic rhinitis (GT) → Allergic rhinitis (S)	<0.001	<0.001
Pollen (GT) → Allergic rhinitis (S)	<0.001	<0.001
Allergic rhinitis (S) → Allergic rhinitis (GT)	0.233	<0.001
Pollen (GT) → Allergic rhinitis (GT)	<0.001	<0.001
Allergic rhinitis (S) → Pollen (GT)	0.182	<0.001
Allergic rhinitis (GT) → Pollen (GT)	0.033	0.004

GT=Google Trends data; S=Surveillance data. Adjusting for the several comparisons performed by applying a Bonferroni correction, *p*-values of 0.001 or lower indicate statistically significant differences at a significance level of 0.05; ^a Despite its name, the Granger causality test does not necessarily fully assess true causality, but rather if one time series is helpful in forecasting other. This test must be carefully interpreted when both time series are driven by a common third series, or for time series displaying a non-linear association. For more information, please consult (i) Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37(3):424-438, and (ii) Grassman G. New considerations on the validity of the Wiener-Granger causality test. *Heliyon*. 2020;6:e05208. ; ^b In this case, we are assessing whether surveillance data on upper respiratory tract infections are useful to forecast surveillance data on asthma; results were not statistically significant when considering the pre-pandemic period (January 2016-January 2020), but were so when considering the whole period (January 2016-July 2021); ^c Google Trends data on asthma not presented on account of a large media-driven Google Trends search peak in March 2020.