




Assessing global COPD awareness with Google Trends

Anna Boehm^{1,3}, Alex Pizzini ^{1,3}, Thomas Sonnweber¹, Judith Loeffler-Ragg¹, Claudia Lamina², Guenter Weiss¹ and Ivan Tancevski¹

Affiliations: ¹Dept of Internal Medicine II, Infectious Diseases, Pneumology, Rheumatology, Medical University of Innsbruck, Innsbruck, Austria. ²Division of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. ³These two authors contributed equally to this work.

Correspondence: Alex Pizzini, Department of Internal Medicine II, Infectious Diseases, Pneumology, Rheumatology, Medical University of Innsbruck, Anichstraße 35, 6020 Innsbruck, Austria.
E-mail: alex.pizzini@i-med.ac.at

 @ERSpublications

Google Trends provides us with an important tool to evaluate public interest related to COPD and associated respiratory diseases. COPD is highly under-represented in Google search queries compared with other frequent but “less preventable” diseases. <http://bit.ly/2PAEwZW>

Cite this article as: Boehm A, Pizzini A, Sonnweber T, *et al.* Assessing global COPD awareness with Google Trends. *Eur Respir J* 2019; 53: 1900351 [<https://doi.org/10.1183/13993003.00351-2019>].

ABSTRACT Although chronic obstructive pulmonary disease (COPD) prevalence and mortality rates rise continuously, patients often remain undiagnosed, probably due to a lack of disease-related awareness. The aim of this study was to quantify public interest in COPD by analysing the frequency of web queries *via* Google.

Data from 2004 to 2018 were collected using the search engine query data analysis tool Google Trends. The relative search volume of the topic “chronic obstructive pulmonary disease” was compared with the relative search volume of nine topics representing the major causes of death in high-income countries according to the World Health Organization.

Our analysis showed highest relative search volumes for the topics “diabetes mellitus”, followed by “stroke” and “breast cancer”. The topic “chronic obstructive pulmonary disease” ranked eighth and its relative search volume clearly displayed a seasonal variation, with peaks in the first and the fourth quarter of the year.

This analysis reveals that COPD is highly under-represented in the public interest, while real-world prevalence constantly rises, indicating that there is still an urgent need to raise the levels of awareness for COPD.

This article has supplementary material available from erj.ersjournals.com

Received: Feb 19 2019 | Accepted after revision: April 23 2019

Copyright ©ERS 2019

Introduction

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of morbidity and mortality worldwide [1, 2]. Data from the Global Burden of Disease Study indicate that approximately 174.5 million people suffer from COPD worldwide, accounting for approximately 3.2 million deaths annually, but there might still be a high number of undiagnosed cases [3]. Demographic data revealed that the rapid rise of the global population and the continuous improvements in wealth will lead to a dramatic increase of chronic diseases associated with ageing, pollution and exposure to noxious fumes or vapours like cigarette smoke, such as COPD [1, 4]. The World Health Organization (WHO) ranked COPD as the fourth leading cause of death in 2000, as the third leading cause of death in 2016 and overall the Global Burden of Disease Study revealed an increase in prevalence of COPD of 44.2% from 1990 to 2015 [2, 3].

The web search engine Google is by far the most important web search engine in English-speaking countries [5]. It has previously been shown that the analysis of Google search queries may represent a powerful tool to detect the real-time global activity of diseases [6]. Using Google's search engine data analysis tool Google Trends, we investigated whether the alarming trends in COPD are also reflected in public interest in the disease.

Methods

The study was performed between July and August 2018 at the Medical University of Innsbruck (Innsbruck, Austria). Data were collected using the public web facility Google Trends (<https://trends.google.com/trends>).

Google Trends is a publicly accessible tool analysing web queries made *via* the Google search engine and displaying the results on a normalised scale. Search volume data for search terms across different geographic locations are available since 2004. Google Trends determines the proportion of searches for a user-specified term among all searches performed on Google over a specified geographic region and time period. It provides users with a graph and optional downloadable output of relative search volume. Relative search volume ranges from 1 to 100, representing search interest relative to the peak popularity for the used search term. A relative search volume value of 100 indicates peak popularity and a score of 0 indicates that the term is below 1% of its peak popularity [7, 8]. For instance, a relative search volume of 70 reflects 70% of the highest search volume monitored during the observation time. To rule out any bias for absolute search volume measurements, relative search volume indirectly corrects for internet access and population size, which have both risen over time and would bias any absolute search volume measure [8, 9]. Moreover, Google Trends automatically excludes duplicate searches, if made by the same person in a short period of time [8]. Search queries in Google Trends are defined either as a term or as a topic. The latter includes all terms that have the same idea or semantic in every language. For example, the topic "London" includes also the Spanish word "Londres" as well as the query "capital of the UK" [8, 10]. Importantly, Google Trends also allows for a direct comparison between the relative search volumes of different topics [8].

To identify the topics reflecting the top 10 causes of death according to the WHO (table 1) [2], we used the following approach. First, all search queries were defined as topics. Second, among the synonymous

TABLE 1 Relative search volume of the topics used in Google Trends reflecting the top 10 causes of death in high-income countries according to the World Health Organization [2]

Top 10 causes of death in high-income countries 2016	Mortality per 100 000 population	Chosen topic in Google Trends	Relative search volume
Ischaemic heart disease	146.7	myocardial infarction	14.81±1.85
Stroke	63.1	stroke	26.78±3.44
Alzheimer disease and other dementias	61.1	dementia	13.70±2.50
Trachea, bronchus and lung cancers	49.3	lung cancer	10.91±2.13
Chronic obstructive pulmonary disease	47.2	chronic obstructive pulmonary disease	9.15±1.21
Lower respiratory infections	37.3	pneumonia	19.94±4.43
Colon and rectum cancers	27.9	colorectal cancer	1.0±1.0
Diabetes mellitus	23.1	diabetes mellitus	76.43±8.28
Kidney diseases	18.8	chronic kidney disease	4.10±0.77
Breast cancer	15.9	breast cancer	25.22±6.93

Data are presented as n or mean±sd.

topics suggested by Google Trends, the topic with the highest relative search volume was finally included in the analysis. For instance, when searching for “ischaemic heart disease” the search topics “myocardial infarction” and “coronary artery disease” were suggested by Google Trends. In a direct comparison of these topics, “myocardial infarction” appeared to have the highest relative search volume, therefore we chose the topic “myocardial infarction” to represent “ischaemic heart disease” in our study.

According to the top 10 causes of death in high-income countries published by the WHO in 2016 (table 1), we evaluated the following topics in our study: “myocardial infarction”, “stroke”, “dementia”, “lung cancer”, “chronic obstructive pulmonary disease”, “pneumonia”, “colorectal cancer”, “diabetes mellitus”, “chronic kidney disease” and “breast cancer” [2]. To assess how often the included topics are searched in comparison with nonmedical topics, we matched “diabetes mellitus”, *i.e.* the topic with the highest relative search volume, with the topics “money” and “car”. Finally, we used a broader and more comprehensive approach to evaluate the search topic “chronic obstructive pulmonary disease” with regard to closely related respiratory diseases and symptoms; results are reported separately in the supplementary material.

On July 18, 2018, we queried Google Trends and downloaded the data. The interest by region option was set to worldwide and not limited to a specific geographic area.

For statistical analysis, we subdivided data into months (seasonal component) and quarters of the year, which were defined as: Quarter 1 (Q1)=January–March, Quarter 2 (Q2)=April–June, Quarter 3 (Q3)=July–September and Quarter 4 (Q4)=October–December.

Correlation analyses were performed using the seasonal decomposition of time series by LOESS (locally estimated scatter plot smoothing: local polynomial regression fitting), which allows for the time series to be decomposed into seasonal parts, trend and irregular components. A generalised least squares model accounting for autocorrelation between residuals was established to further evaluate whether a trend is significant over time, after adjustment for the seasonal components. The correlation structure of residuals was deduced from an automatic selection of ARIMA (AutoRegressive Integrated Moving Average) model parameters using Akaike’s information criterion.

All tests were calculated two-tailed and a p -value ≤ 0.05 indicated statistical significance. Statistical analyses were performed with SPSS version 24.0 (IBM, Armonk, NY, USA), as well as R version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria) using “nlme” [11] and “forecast” [12] libraries.

Importantly, none of the queries in the Google database for this study can be associated with a particular individual. The database retains no information about the identity, internet protocol address or specific physical location of any user. Furthermore, any original web search logs older than 9 months are being made anonymous in accordance with Google’s privacy policy (www.google.com/privacypolicy.html) [6].

Results

A comparison of all search topics revealed that “diabetes mellitus” displayed by far the highest relative search volume, with a maximum relative search volume of 100 (mean \pm SD 76.43 \pm 8.28). The second most frequent topic was “stroke”, with a maximum relative search volume of 37 (mean \pm SD 26.78 \pm 3.44), followed by “breast cancer”, with a maximum relative search volume of 52 (mean \pm SD 25.22 \pm 6.93). The topic “chronic obstructive pulmonary disease” was ranked eighth, with a maximum relative search volume of 12 (mean \pm SD 9.15 \pm 1.20), followed by the topics “chronic kidney disease” and “colorectal cancer” (figure 1). Due to the low relative search volume of “colorectal cancer” (mean \pm SD 1.0 \pm 1.0) in relation to “diabetes mellitus”, the topic “colorectal cancer” was excluded from statistical analyses. The exact ranking of the 10 search topics according to their relative search volume is shown in figure 1 and table 1. COPD-related terms and topics are described separately in the supplementary material.

Interestingly, when we compared the used search topics to common nonmedical topics such as “car” or “money”, we observed a substantial difference in the magnitude of relative search volume. For instance, when comparing “car” with “diabetes mellitus” (relative search volume mean \pm SD 84.50 \pm 6.54 *versus* 3.79 \pm 0.53), the topic “car” was searched 22.35 times more often than the topic “diabetes mellitus”.

Next, a trend over time adjusted for the seasonal component was calculated, analysing a monthly increase of relative search volume of the respective topics since January 2004. During the overall observation period (January 2004–July 2018), significant increases in relative search volume of the topics “myocardial infarction”, “stroke”, “dementia” and “pneumonia” were observed (table 2). In contrast, the topics “diabetes mellitus” and “lung cancer” showed a significant decrease of their relative search volume over time, whereas the topics “chronic obstructive pulmonary disease”, “chronic kidney disease” and “breast cancer” displayed no significant change of relative search volume during the observation period. Interestingly, decomposition of the search topics time series “chronic obstructive pulmonary disease”, “pneumonia” as well as “breast cancer” into seasonal parts showed that trends and irregular components

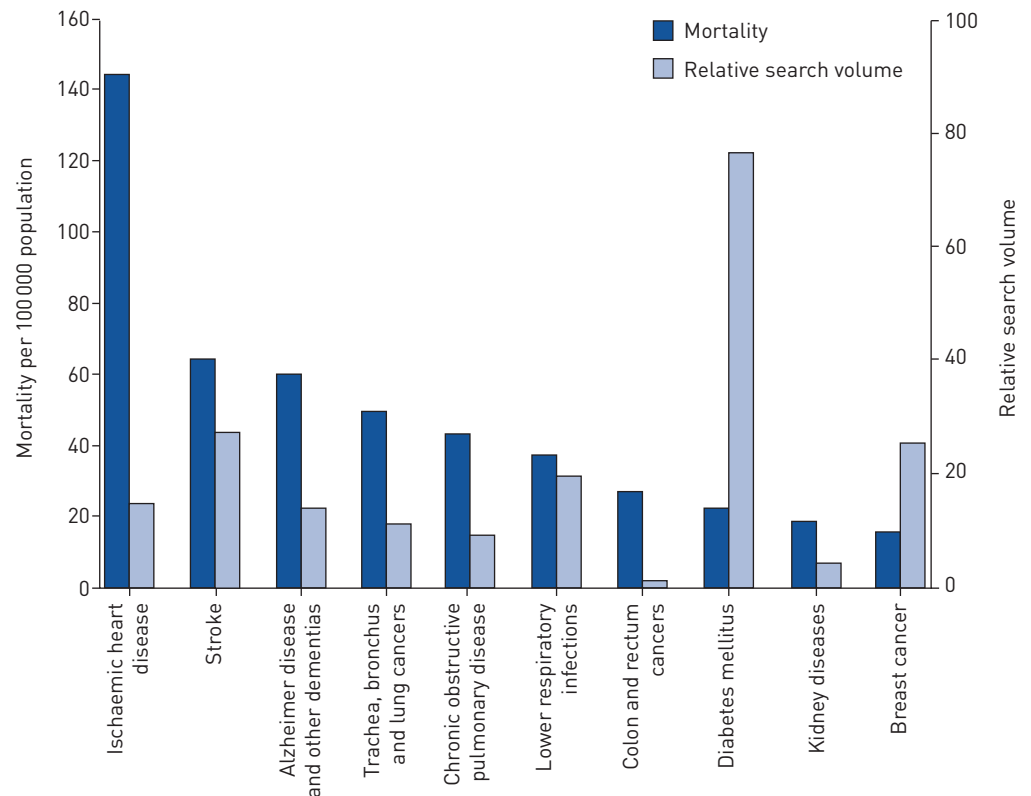


FIGURE 1 Absolute mortality rates and relative search volume by Google Trends analysis of the top 10 causes of death in high-income countries according to the World Health Organization (the corresponding Google Trends topics are given in table 1) [2].

have an annual cycle with returning peaks and lows (supplementary figures S1–S3). November represented the month with the highest “chronic obstructive pulmonary disease”-related relative search volume (mean \pm SD 10.50 \pm 0.65), January was the month with the highest peak for “pneumonia” (mean \pm SD 24.33 \pm 4.62), whereas October represented the month with the highest “breast cancer”-related relative search volume (mean \pm SD 41.93 \pm 3.85) (figure 2). Subdivision of the “chronic obstructive pulmonary disease” relative search volume into calendar quarters revealed that the median relative search volume of “chronic obstructive pulmonary disease” was highest from January to March (Q1: mean \pm SD 9.84 \pm 1.04) and lowest from July to September (Q3: mean \pm SD 8.02 \pm 0.83) (figure 3).

Discussion

COPD is one of the most alarming diseases with rising numbers in terms of incidence, prevalence, morbidity and mortality. However, the disease is largely underdiagnosed and public awareness of the

TABLE 2 Changes in relative search volume over the observation period using the generalised least squares model accounting for autocorrelation between residuals, adjusted for seasonal effects

Google Trends topic	β -estimate \pm se	p-value
Diabetes mellitus	-0.039 \pm 0.012	<0.002
Myocardial infarction	0.021 \pm 0.002	<0.001
Stroke	0.037 \pm 0.005	<0.001
Dementia	0.038 \pm 0.003	<0.001
Lung cancer	-0.031 \pm 0.005	<0.001
Chronic obstructive pulmonary disease	0.013 \pm 0.009	0.13
Pneumonia	0.051 \pm 0.005	<0.001
Chronic kidney disease	0.010 \pm 0.006	0.12
Breast cancer	-0.067 \pm 0.040	0.1

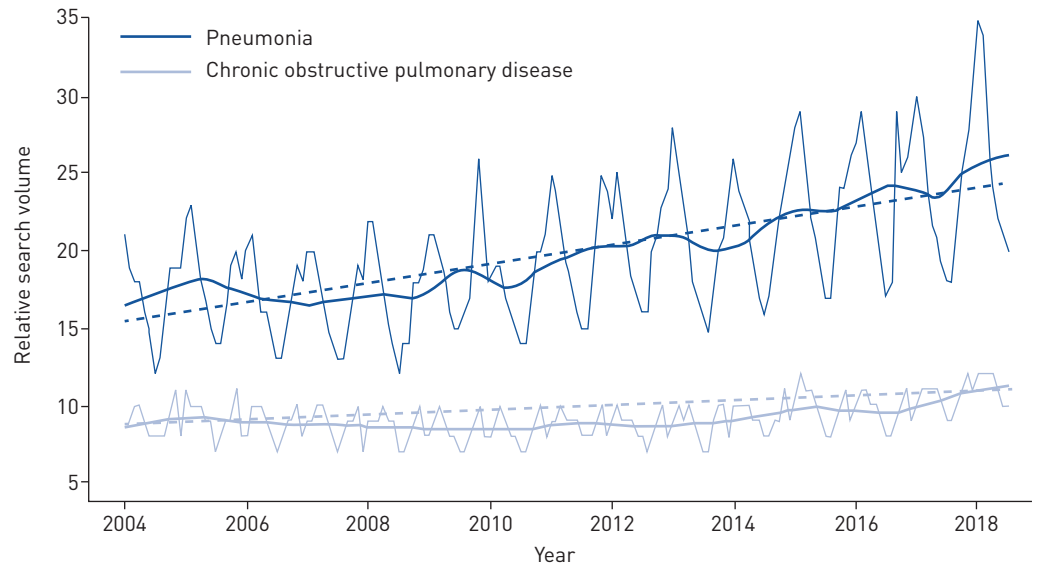


FIGURE 2 Time series for “pneumonia” and “chronic obstructive pulmonary disease” as search topics (thin lines), the trends over time using LOESS (local polynomial regression fitting) (thick lines) and the linear trends estimated from generalised least squares models (dashed lines).

diseases is low. To the best of our knowledge, this is the first report highlighting that the topic “chronic obstructive pulmonary disease” is under-represented in Google search queries compared with other frequent but less preventable diseases. Accordingly, low awareness of COPD in the general population has been reported previously [13–15]. A recent investigation by SEO *et al.* [16] assessed awareness and understanding of COPD in smokers participating in a smoking cessation programme. The authors showed that only 1% mentioned COPD as an example of a respiratory health problem, although more than two-thirds of the participants presented with COPD-related symptoms. Importantly, the authors observed a significant increase of smoking cessation willingness when awareness of COPD was raised [16].

The disparity between the relative search volume of “chronic obstructive pulmonary disease” and other topics, including “breast cancer” or “diabetes mellitus”, may have its origin in the only recently established awareness campaigns for COPD. In line with this assumption, Wikipedia contains a comprehensive article on “Breast cancer awareness” including almost 30 references, whereas the entry titled “COPD Awareness Month” fills only three sentences [17, 18].

Major efforts to raise awareness and knowledge on breast cancer have been undertaken over the past decades, making breast cancer a role model in terms of awareness promotion [19]. The pink ribbon has emerged as a brand concept for breast cancer and the National Breast Cancer Awareness Month (NBCAM) was founded in 1985 [19, 20]. The impact of the NBCAM on public interest is illustrated by cyclic relative search volume peaks during the month of October. In contrast, the COPD Foundation was established in 2004 by John W. Walsh, almost 20 years after the NBCAM [21]. In 2007, the Learn More Breathe Better programme was initiated to raise awareness for COPD and 10 years later, in 2017, the

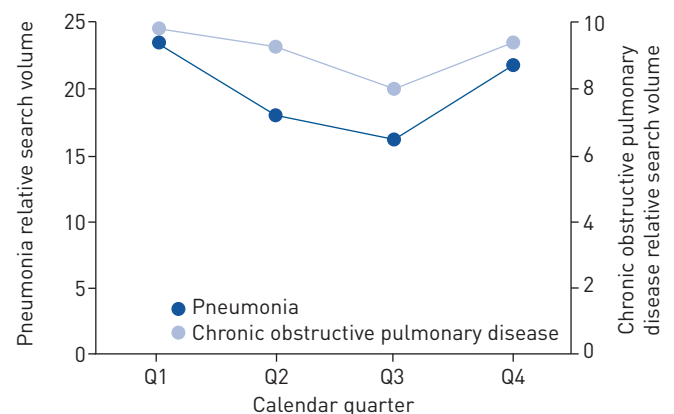


FIGURE 3 Seasonal cycle of the relative search volume for the topics “pneumonia” and “chronic obstructive pulmonary disease”. Data are presented as means (January 2004–July 2018).

National Institutes of Health and the Centers for Disease Control and Prevention together with several other federal agencies published the COPD National Action Plan to fight COPD through improvements in public education, diagnosis, treatment and prevention [13, 22]. Importantly, the COPD National Action Plan proposes to evaluate the effectiveness of information campaigns. Accordingly, the herein presented Google Trends analysis may reflect the effectiveness of the COPD Foundation to raise awareness, as our data analysis shows a significant increase of search volumes related to “chronic obstructive pulmonary disease” in November, when the COPD Awareness Month and the World COPD Day are promoted. However, it is alarming that there is no significant trend towards a general increase in the global interest for “chronic obstructive pulmonary disease”, that the awareness for other diseases is growing at a much faster pace and that the relative search volume of the topic “lung cancer”, a disease closely related to COPD and cigarette smoking, is constantly decreasing.

Interestingly, we observed a rise in the relative search volume of “chronic obstructive pulmonary disease” during the winter months, displaying an annual cycle. Similarly, a study by KUMAR *et al.* [23] reported seasonal patterns of online search trends related to cardiovascular health. We hypothesise that our findings reflect higher rates of acute COPD exacerbations during the winter months, closely related to infectious aetiologies such as *Streptococcus pneumoniae* and respiratory viruses including influenza virus [24, 25]. The fact that Google Trends is a powerful tool to analyse health-seeking behaviour was demonstrated previously by GINSBERG *et al.* [6], who tracked the geographic spread of influenza-like illness *via* Google Trends. The authors were able to show a correlation between web queries *via* Google and the number of physicians being consulted in a certain geographic region. This approach enabled calculation of the current influenza activity levels, which may be of relevance for patients affected by COPD, as similar trends could help to predict more severe exacerbations with the need of additional intensive care and mechanical ventilation therapy.

The topics “breast cancer” and “chronic obstructive pulmonary disease” revealed an awareness month-related relative search volume peak, a finding also described by SCHOOTMAN *et al.* [9], who analysed public interest in cancer screening *via* Google Trends and found a higher relative search volume during particular cancer awareness months. Accordingly, JELLYSON *et al.* [26] reported annual relative search volume peaks for “arthritis” during the National and World Arthritis Days.

Although the herein presented data show interesting new aspects related to the 10 most deadly diseases of high-income countries, we have to acknowledge potential limitations of this study as the presented data need to be carefully interpreted in the context of disease awareness. First, the data output is a relative number, therefore an increase of search volumes related to other important topics (politics, general news, technology, *etc.*) might impact on relative search volumes. Second, there is no information available about the individuals who searched for the analysed terms or topics. A bias related to high numbers of search queries by healthcare professionals, industry or marketing agencies cannot be excluded.

Finally, it is to some extent elusive which search queries are summarised in the topics defined by Google Trends algorithms, as detailed information on how Google generates this data is not provided. The selections of terms/topics might affect the results and conclusions; therefore, we decided to use the topics more accurately representing the top 10 causes of death and provide a detailed description of our data-gathering approach in order to facilitate reproducibility. The importance of accuracy in defining the search queries is exemplified when searching Google Trends for the topic “cough”. Cough is a symptom frequently associated with COPD, although not specifically representing this disease (reported in the supplementary material), thus using the query “cough” may be useful to analyse symptom-related interest but does not sufficiently represent global COPD awareness. Although the number of studies based on Google Trends is increasing, so far no standardised procedure for data collection is known and thus more guidance by Google is warranted in order to assist researchers to establish an optimal search strategy [7].

In conclusion, Google Trends provides an important tool to evaluate public interest related to COPD and associated respiratory diseases. In line with the goals of the COPD National Action Plan, Google Trends helps to collect, analyse, report and disseminate COPD-related health data. Thus, Google Trends may drive change and track progress, and may help to improve programmes to counteract the current lack of public COPD awareness.

Author contributions: A. Boehm, A. Pizzini, T. Sonnweber, C. Lamina, J. Loeffler-Ragg, G. Weiss and I. Tancevski conceived and designed the study. A. Boehm and A. Pizzini drafted the manuscript. A. Boehm collected the data. A. Boehm, A. Pizzini, C. Lamina, T. Sonnweber and I. Tancevski analysed and interpreted the data. A. Boehm, A. Pizzini, T. Sonnweber, C. Lamina, J. Loeffler-Ragg, G. Weiss and I. Tancevski revised and approved the manuscript.

Conflict of interest: None declared.

References

- 1 Vogelmeier CF, Criner GJ, Martinez FJ, *et al.* Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. *Eur Respir J* 2017; 49: 1700214.
- 2 World Health Organization. The top 10 causes of death. 2016. www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death Date last updated: May 24, 2018. Date last accessed: December 21, 2018.
- 3 GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017; 5: 691–706.
- 4 López-Campos JL, Tan W, Soriano JB. Global burden of COPD. *Respirology* 2016; 21: 14–23.
- 5 Comscore. Comscore Releases February 2016 U.S. Desktop Search Engine Rankings. 2016. www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings Date last accessed: December 1, 2018.
- 6 Ginsberg J, Mohebbi MH, Patel RS, *et al.* Detecting influenza epidemics using search engine query data. *Nature* 2009; 457: 1012–1014.
- 7 Nuti SV, Wayda B, Ranasinghe I, *et al.* The use of Google Trends in health care research: a systematic review. *PLoS One* 2014; 9: e109583
- 8 Support.google.com. Trends Help. 2019. <https://support.google.com/trends#topic=6248052> Date last accessed: January 18, 2019.
- 9 Schootman M, Toor A, Cavazos-Rehg P, *et al.* The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* 2015; 5: e006678.
- 10 Realgness.net. Understand Google Trends Search Term and Topic. 2014. <https://realguess.net/tags/google-trends> Date last accessed: November 18, 2018.
- 11 Pinheiro J, Bates D, DebRoy S, *et al.* nlme: linear and nonlinear mixed effects models. R package version 3. 2018. <https://CRAN.R-project.org/package=nlme> Date last accessed: May 7, 2019.
- 12 Hyndman R, Athanasopoulos G, Bergmeir C, *et al.* Forecast: forecasting functions for time series and linear models. R package version 8.5. 2019. <http://pkg.robjhyndman.com/forecast> Date last accessed: May 7, 2019.
- 13 National Institutes of Health, Centers for Disease Control and Prevention. COPD National Action Plan. 2017. www.nhlbi.nih.gov/sites/default/files/media/docs/COPD_National_Action_Plan_508_0.pdf Date last accessed: September 16, 2018.
- 14 Sikjær MG, Hilberg O, Fløe A, *et al.* Lack of awareness towards smoking-related health risks, symptoms related to COPD, and attitudinal factors concerning smoking: an Internet-based survey conducted in a random sample of the Danish general population. *Eur Clin Respir J* 2018; 5: 1506235.
- 15 Asai M, Tanaka T, Kozu R, *et al.* Effect of a chronic obstructive pulmonary disease (COPD) intervention on COPD awareness in a regional city in Japan. *Intern Med* 2015; 54: 163–169.
- 16 Seo JY, Hwang YI, Mun SY, *et al.* Awareness of COPD in a high risk Korean population. *Yonsei Med J* 2015; 56: 362–367.
- 17 Wikipedia. Breast cancer awareness. 2019. https://en.wikipedia.org/wiki/Breast_cancer_awareness Date last accessed: January 20, 2019.
- 18 Wikipedia. COPD Awareness Month. 2019. https://en.wikipedia.org/wiki/COPD_Awareness_Month Date last accessed: January 20, 2019.
- 19 Breastcancerconsortium.net. Critical Thinking on Breast Cancer. 2019. <http://breastcancerconsortium.net/resources/topics/breast-cancer-movement> Date last accessed: November 15, 2018.
- 20 American Psychological Association. October is National Breast Cancer Awareness Month. 2010. www.apa.org/pi/women/resources/newsletter/2010/10/breast-cancer.aspx Date last accessed: November 14, 2018.
- 21 COPD Foundation. About The COPD Foundation. 2019. www.copdfoundation.org Date last accessed: January 18, 2019.
- 22 National Heart, Lung, and Blood Institute. 2019. Join the Breathe Better Network. www.nhlbi.nih.gov/health-topics/education-and-awareness/copd-learn-more-breathe-better/join-breathe-better-network Date last accessed: April 9, 2019.
- 23 Kumar N, Pandey A, Garg N, *et al.* Seasonal and geographic patterns in seeking cardiovascular health information: an analysis of the online search trends. *Mayo Clin Proc* 2018; 93: 1185–1190.
- 24 Jenkins CR, Celli B, Anderson JA, *et al.* Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. *Eur Respir J* 2012; 39: 38–45.
- 25 Dowell SF, Ho MS. Seasonality of infectious diseases and severe acute respiratory syndrome – what we don't know can hurt us. *Lancet Infect Dis* 2004; 4: 704–708.
- 26 Jellison SS, Bibens M, Checketts J, *et al.* Using Google Trends to assess global public interest in osteoarthritis. *Rheumatol Int* 2018; 38: 2133–2136.