



## Early View

Original article

### **Assessing global COPD awareness with *Google Trends***

Anna Boehm, Alex Pizzini, Thomas Sonnweber, Judith Loeffler-Ragg, Claudia Lamina, Guenter Weiss, Ivan Tancevski

Please cite this article as: Boehm A, Pizzini A, Sonnweber T, *et al.* Assessing global COPD awareness with *Google Trends*. *Eur Respir J* 2019; in press (<https://doi.org/10.1183/13993003.00351-2019>).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

Copyright ©ERS 2019

## Assessing global COPD awareness with *Google Trends*

Anna Boehm<sup>1\*</sup>, Alex Pizzini<sup>1\*</sup>, Thomas Sonnweber<sup>1</sup>, Judith Loeffler-Ragg<sup>1</sup>, Claudia Lamina<sup>2</sup>,  
Guenter Weiss<sup>1</sup>, Ivan Tancevski<sup>1</sup>

### **Affiliations:**

<sup>1</sup> Department of Internal Medicine II, Infectious Diseases, Pneumology, Rheumatology,  
Innsbruck Medical University, Innsbruck, Austria

<sup>2</sup> Division of Genetic Epidemiology, Innsbruck Medical University, Innsbruck, Austria

\* AB and AP contributed equally to this manuscript.

**Correspondence:** Alex Pizzini, Department of Internal Medicine II, Infectious Diseases,  
Pneumology, Rheumatology, Innsbruck Medical University, Anichstraße 35, 6020 Innsbruck,  
Austria. E-Mail: [alex.pizzini@i-med.ac.at](mailto:alex.pizzini@i-med.ac.at)

Take home message: *Google Trends* provides us with an important tool to evaluate public interest related to COPD and associated respiratory diseases. COPD is highly underrepresented in *Google* search queries compared to other frequent but “less preventable” diseases.

## **Abstract**

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality constituting the third leading cause of death worldwide. Although 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 2004-2018 were collected using the search engine query data analysis tool *Google Trends*. The relative search volume (RSV) of the topic COPD was compared with the RSV of nine topics representing the major causes of death in high-income countries according to the WHO.

Our analysis showed highest RSVs for diabetes, followed by stroke and breast cancer. COPD ranked 8<sup>th</sup>, and its RSV clearly displayed a seasonal variation, with peaks in the first and the fourth quarter of the year. In summary, 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.

Key words: COPD, Google Trends, awareness, major causes of death, WHO, Prevention

## **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 aging, pollution, and exposure to noxious fumes or vapors like cigarette smoking, such as COPD [1,4]. The WHO ranked COPD as the fourth leading cause of death in 2000, as third leading cause of death in 2016 and overall the Global Burden of Disease Study revealed an increase in prevalence of COPD by 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* (GT; <https://trends.google.com/trends/>) we investigated whether the alarming trends in COPD are also reflected in the public interest.

## Methods

The study was performed between July and August 2018 at Innsbruck Medical University.

Data were collected using the public web facility GT, available from <https://trends.google.com/trends/>.

GT is a publicly accessible tool analysing web queries made via the *Google* search engine and displaying the results on a normalized scale. Search volume data for search terms across different geographical locations is available since 2004. GT 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 (RSV). RSV ranges from 1–100, representing search interest relative to the peak popularity for the used search term. An RSV value of 100 indicates peak popularity and a score of 0 indicates that the term is below 1 percent of its peak popularity [7,8]. For instance, an RSV of 70, reflects 70 percent of the highest search volume monitored during the observation time. To rule out any bias for absolute search volume measurements, RSV indirectly corrects for internet access and population size, which both rose over time and would bias any absolute search volume measure [8,9]. Moreover, GT automatically excludes duplicate searches, if made by the same person in a short period of time [8]. Search queries in GT 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, GT also allows for a direct comparison between the RSVs of different topics [8].

To identify the topics reflecting the top ten causes of death according to the WHO (**table 1**), we used the following approach: first, all search queries were defined as topics. Second, among the synonymous topics suggested by GT, the topic with the highest RSV 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 GT. In a

direct comparison of these topics, “myocardial infarction” appeared to have the highest RSV, therefore we chose the topic “myocardial infarction” to represent the term "ischaemic heart disease" in our study.

According to the top ten causes of death in high-income countries published by the WHO in 2016 (listed in **table 1**, [2]), 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". To assess how often the included topics are searched in comparison to non-medical topics, we matched "diabetes mellitus" - the topic with the highest RSV- with the topics "money" and "car". Finally, we used a broader and more comprehensive approach to evaluate the search topic "COPD" in regard to closely related respiratory diseases and symptoms; results are separately reported in the online supplement.

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 certain geographical area.

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

Correlation analyses were performed using the Seasonal Decomposition of Time Series by Loess (Local Polynomial Regression Fitting), which allows to decompose the time series into seasonal part, trend and irregular components. A generalized 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-model parameters using AIC criterion.

All tests were calculated two-tailed, and a p-value of 0.05 indicated statistical significance. Statistical analyses were performed with SPSS 24.0 statistical package (IBM Corp., Armonk,

NY, USA), as well as R version 3.5.0 (R Core Team 2018 – 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 (IP) 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 (<http://www.google.com/privacypolicy.html>) [6].

## Results

A comparison of all search topics revealed that "diabetes mellitus" displayed by far the highest RSV, with a maximum of 100 and a mean RSV of  $76.43 \pm 8.28$  SD. The second most frequent term was "stroke" with a maximum RSV of 37 and a mean of  $26.78 \pm 3.44$  SD, followed by "breast cancer" presenting a maximum RSV of 52 and a mean of  $25.22 \pm 6.93$  SD. The term "COPD" was ranked 8<sup>th</sup>, with a maximum RSV of 12 and a mean of  $9.15 \pm 1.20$  SD, followed by the topics "chronic kidney disease" and "colorectal cancer" (**Figure, a**). Due to the low RSV of "colorectal cancer" (mean = 1.0, SD = 1.0) in relation to "diabetes mellitus", the term "colorectal cancer" was excluded from statistical analyses. The exact ranking of the 10 search topics according to their RSV is shown in **Figure, a** and **table 1**. COPD-related terms and topics are separately described in the online supplement.

Interestingly, when we compared the used search topics to common non-medical topics such as "car" or "money", we observed a substantial difference in the magnitude of RSV. For instance, when comparing "car" with "diabetes mellitus" (RSV mean =  $84.50 \pm 6.54$  SD vs. RSV mean =  $3.79 \pm 0.53$  SD), the term "car" was searched 22.35 times more often than the term "diabetes mellitus".

Next, a trend over time adjusted for the seasonal component was calculated, analysing a monthly increase of RSV of the respective topics since January 2004. During the overall observation period (January 2004 – July 2018), significant increases in RSV 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 RSV over time, whereas the topics "COPD", "chronic kidney diseases" and "breast cancer" displayed no significant change of RSV during the observation period. Interestingly, decomposition of the search topics time series "COPD", "pneumonia" as well as "breast cancer" into seasonal parts showed that trends and irregular components have an annual cycle with returning peaks and lows (online supplementary **Figures S1-S3**). November represented



the month with the highest "COPD"-related RSV (mean = 10.50,  $\pm$  0.65 SD), January was the month with the highest peak for "pneumonia" (mean = 24.33,  $\pm$  4.62 SD), whereas October represented the month with the highest "breast cancer" related RSV (mean 41.93,  $\pm$  3.85 SD). **(Figure, b)**. Subdivision of the "COPD" RSV into calendar quarters (Q) revealed that the median RSV of "COPD" was highest from January to March (Q1, mean = 9.84  $\pm$  1.04 SD) and lowest from July to September (Q3, mean = 8.02  $\pm$  0.83 SD) **(Figure, c)**.

## **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 diseases is low. To our knowledge, this is the first report highlighting that the topic "COPD" is underrepresented in *Google* search queries compared to other frequent but less preventable diseases. Accordingly, low awareness of COPD in the general population has been reported previously [13-15]. A recent investigation assessed awareness and understanding of COPD in smokers participating in a smoking cessation program. The authors showed that only 1 percent mentioned COPD as an example of a respiratory health problem, although more than two-thirds of the participants were presenting 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 RSV of "COPD" 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 lines [17,18].

In fact, highest efforts to raise awareness and knowledge on breast cancer have been undertaken over the last decades, conferring 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 RSV 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® program was initiated to raise awareness for COPD and ten years later in 2017, the National Institute of Health (NIH) and the Centers for Disease Control and Prevention (CDC) 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 "National Action Plan" proposes to evaluate the effectiveness of information campaigns. Accordingly, the herein presented GT 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 "COPD" 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 "COPD", that the awareness for other diseases grows at a much faster pace and that the RSV of the topic "lung cancer", a disease closely related to COPD and cigarette smoking, is constantly decreasing.

Interestingly, we observed a rise in the RSV of "COPD" during winter months, displaying an annual cycle. Similarly, a study by Kumar et al. reported seasonal patterns of online search trends related to cardiovascular health [23]. We hypothesize that our findings reflect higher rates of acute COPD exacerbations during winter months, closely related to infectious aetiologies such as *Streptococcus pneumoniae* and respiratory viruses including *influenza* [24,25]. The fact that GT is a powerful tool to analyse health seeking behaviour was previously demonstrated by Ginsberg et al., who geographically tracked the spread of influenza-like illness via GT [6]. The authors were able to show a correlation between web queries via *Google* and the number of physicians being consulted in a certain geographical 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 "COPD" revealed an awareness month-related RSV peak, a finding also described by Schootman et al.[9], who analysed public interest in cancer screening via GT and found a higher RSV during particular cancer awareness months.

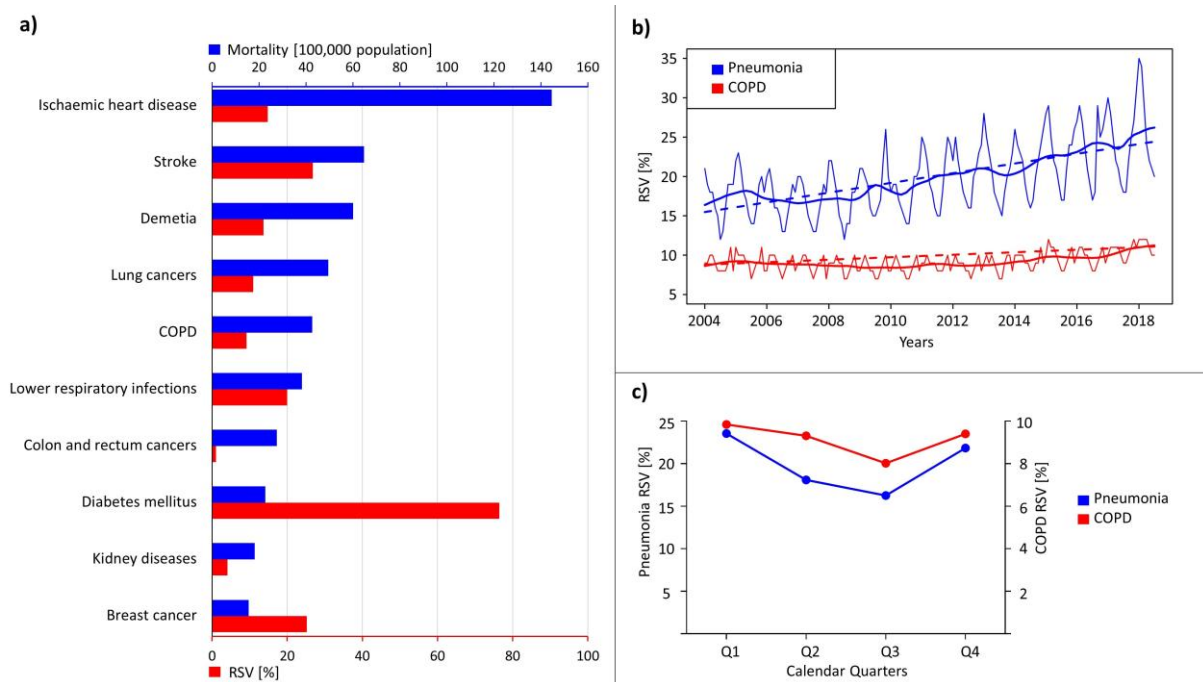
Accordingly, Jellison et al. [26] reported annual RSV peaks for "arthritis" during the National and World Arthritis Days.

Although the herein presented data show interesting new aspects related to the ten 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 RSVs. 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 health care 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 GT 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 the accuracy in defining the search queries is exemplified when searching GT for the topic "cough". Cough is a symptom frequently associated with COPD, however not specifically representing this disease (reported in the online supplement), 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 GT 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, GT 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, GT helps to collect, analyse, report, and disseminate COPD-related health data. Thus, GT may

drive change and track progress and may help to improve programs to counteract the current lack of public COPD awareness.



**Conflict of interest:** None.

**Funding:** this research did not receive any specific funding.

**Author's contribution:**

AB, AP, TS, CL, JLR, GW, IT conceived and designed the study

AB, AP drafted the manuscript

AB collected the data

AB, AP, CL, TS, IT analysed and interpreted the data

AB, AP, TS, CL, JLR, GW, IT revised and approved the manuscript

## References

- (1) Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DM, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Augusti A. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. *Eur Respir J* 2017; 49(3): 1700214.
- (2) World Health Organization (WHO). The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Date last updated: May 24 2018. Date last accessed: Dec 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(9): 691-706.
- (4) López-Campos JL, Tan W, Soriano JB. Global burden of COPD. *Respirology* 2016; 21(1): 14–23.
- (5) Comscore, Inc. (NASDAQ:SCOR). Comscore Releases February 2016 U.S. Desktop Search Engine Rankings. <https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings>. Date last updated: Mar 16 2016. Date last accessed: Dec 1 2018.
- (6) Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 457(7232): 1012–4.
- (7) Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, Murugiah K. The use of google trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583. Doi: 10.1371/journal.pone.0109583. [Epub 2014 Oct 22].
- (8) Support.google.com. Trends Help. <https://support.google.com/trends#topic=6248052>.

Date last updated: no date available. Date last accessed: Jan 18 2019.

- (9) Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, Davidson NO. The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* 2015; 5(6): e006678. Doi: 10.1136/bmjopen-2014-006678 [Epub 2015 June 8].
- (10) Realguess.net. Understand Google Trends Search Term and Topic. <https://realguess.net/tags/google-trends/>. Date last updated: Apr 27 2014. Date last accessed: Nov 18 2018.
- (11) Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2018). *Nlme: Linear and Nonlinear Mixed effects Models*. R package version 3. 1-137. <https://CRAN.R-project.org/package=nlme>.
- (12) Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeeen F (2019). *Forecast: Forecasting functions for time series and linear models*. R package version 8.5. <http://pkg.robjhyndman.com/forecast>.
- (13) Department of Health & Human Services USA, National Institution of Health, Centers for Disease Control and Prevention. COPD National Action Plan. [https://www.nhlbi.nih.gov/sites/default/files/media/docs/COPD National Action Plan 508\\_0.pdf](https://www.nhlbi.nih.gov/sites/default/files/media/docs/COPD_National_Action_Plan_508_0.pdf). Date last updated: May 22 2017. Date last accessed: Sep 16 2018.
- (14) Sikjær MG, Hilberg O, Fløe A, Dollerup J, Løkke A. 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(1): 1506235. Doi: 10.1080/20018525.2018.1506235 [Epub 2018 Aug 14].
- (15) Asai M, Tanaka T, Kozu R, Kitagawa C, Tabusadani M, Senjyu H. Effect of a chronic obstructive pulmonary disease (COPD) intervention on COPD awareness in a regional city in Japan. *Intern Med* 2015;54(2): 163–9.

- (16) Seo JY, Hwang YI, Mun SY, Kim JH, Kim JH, Park SH, Jang SH, Park YB, Shim JJ, Jung KS. Awareness of COPD in a high risk Korean population. *Yonsei Med J* 2015; 56(2): 362–7.
- (17) Wikipedia. Breast cancer awareness.  
[https://en.wikipedia.org/wiki/Breast\\_cancer\\_awareness](https://en.wikipedia.org/wiki/Breast_cancer_awareness). Date last updated: Jan 5 2019.  
Date last accessed: Jan 20 2019.
- (18) Wikipedia. COPD Awareness Month.  
[https://en.wikipedia.org/wiki/COPD\\_Awareness\\_Month](https://en.wikipedia.org/wiki/COPD_Awareness_Month). Date last updated: Jan 5 2019.  
Date last accessed: Jan 20 2019.
- (19) Breastcancerconsortium.net. Critical Thinking on Breast Cancer.  
<http://breastcancerconsortium.net/resources/topics/breast-cancer-movement/>. Date last updated: no date available. Date last accessed: Nov 15 2018.
- (20) American Psychological Association. October is National Breast Cancer Awareness Month. <https://www.apa.org/pi/women/resources/newsletter/2010/10/breast-cancer.aspx>. Date last updated: Oct 2010. Date last accessed: Nov 14 2018.
- (21) COPD Foundation. About The COPD Foundation. <https://www.copdfoundation.org>.  
Date last updated: no date available. Date last accessed: Jan 18 2019.
- (22) National Heart, Lung, and Blood Institute. Join the Breathe Better Network.  
<https://www.nhlbi.nih.gov/health-topics/education-and-awareness/copd-learn-more-breathe-better/join-breathe-better-network>. Date last updated: no date available. Date last accessed: Apr 09 2019.
- (23) Kumar N, Pandey A, Garg N, Sampene E, Lavie CJ. Seasonal and Geographic Patterns in Seeking Cardiovascular Health Information: An Analysis of the Online Search Trends. *Mayo Clin Proc* 2018;93(9): 1185–90.
- (24) Jenkins CR, Celli B, Anderson JA, Ferguson GT, Jones PW, Vestbo J, Yates JC, Calverley PM. Seasonality and determinants of moderate and severe COPD



exacerbations in the TORCH study. *Eur Respir J* 2012;39(1): 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(11): 704–8.

(26) Jellison SS, Bibens M, Checketts J, Vassar M. Using Google Trends to assess global public interest in osteoarthritis. *Rheumatol Int* 2018;38(11): 2133-6.

## Tables

**Table 1:** Relative search volume (RSV) of the topics used in *Google Trends* reflecting the top ten causes of death in high-income economies [2]. Data is shown as mean  $\pm$  standard deviation (SD)

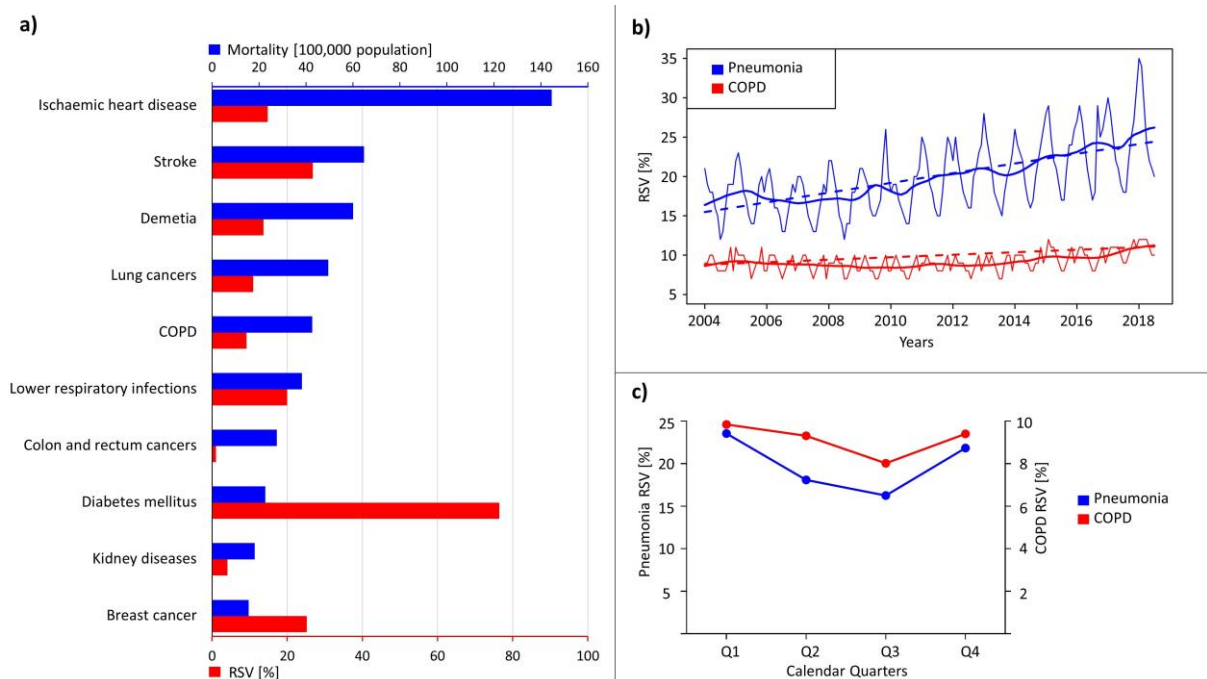
<b>Top ten causes of death in high-income economies 2016</b>	<b>Deaths per 100,000 population</b>	<b>Chosen topic in GT</b>	<b>Mean RSV</b>	<b>SD (<math>\pm</math>)</b>
ischaemic heart disease	146.7	myocardial infarction	14.81	1.85
stroke	63.1	stroke	26.78	3.44
Alzheimer disease and other dementias	61.1	dementia	13.70	2.50
trachea, bronchus, 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	1
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

**Table 2:** Changes of the RSV over the observation period using the generalized least squares model accounting for autocorrelation between residuals, adjusted for seasonal effects. Data is shown as beta estimate, standard error (se) and p-value.

	<b>beta</b>	<b>se</b>	<b>p-value</b>
diabetes mellitus	-0.039	0.012	< 0.002
myocardial infarction	0.021	0.002	< 0.001
stroke	0.037	0.005	< 0.001
dementia	0.038	0.003	< 0.001
lung cancer	-0.031	0.005	< 0.001
chronic obstructive pulmonary disease	0.013	0.009	0.13
pneumonia	0.051	0.005	< 0.001
chronic kidney disease	0.010	0.006	0.12
breast cancer	-0.067	0.040	0.1

## Figure legends

**FIGURE** a) Mortality rates and RSV of the top 10 causes of death in high-income countries. Blue bars depict absolute mortality rates per 100,000 population (top x-axis), red bars show RSV according to Google Trends analysis (bottom x-axis). b) Time-series for "pneumonia" and "COPD" as search topics (thin lines), the trend over time using LOESS (thick lines) and the linear trend estimated from generalized least squares models (dashed lines). c) Seasonal cycle of the RSV for "pneumonia" (blue) and "COPD" (red). Data is shown as mean (01/2004 – 07/2018).



## Online supplement

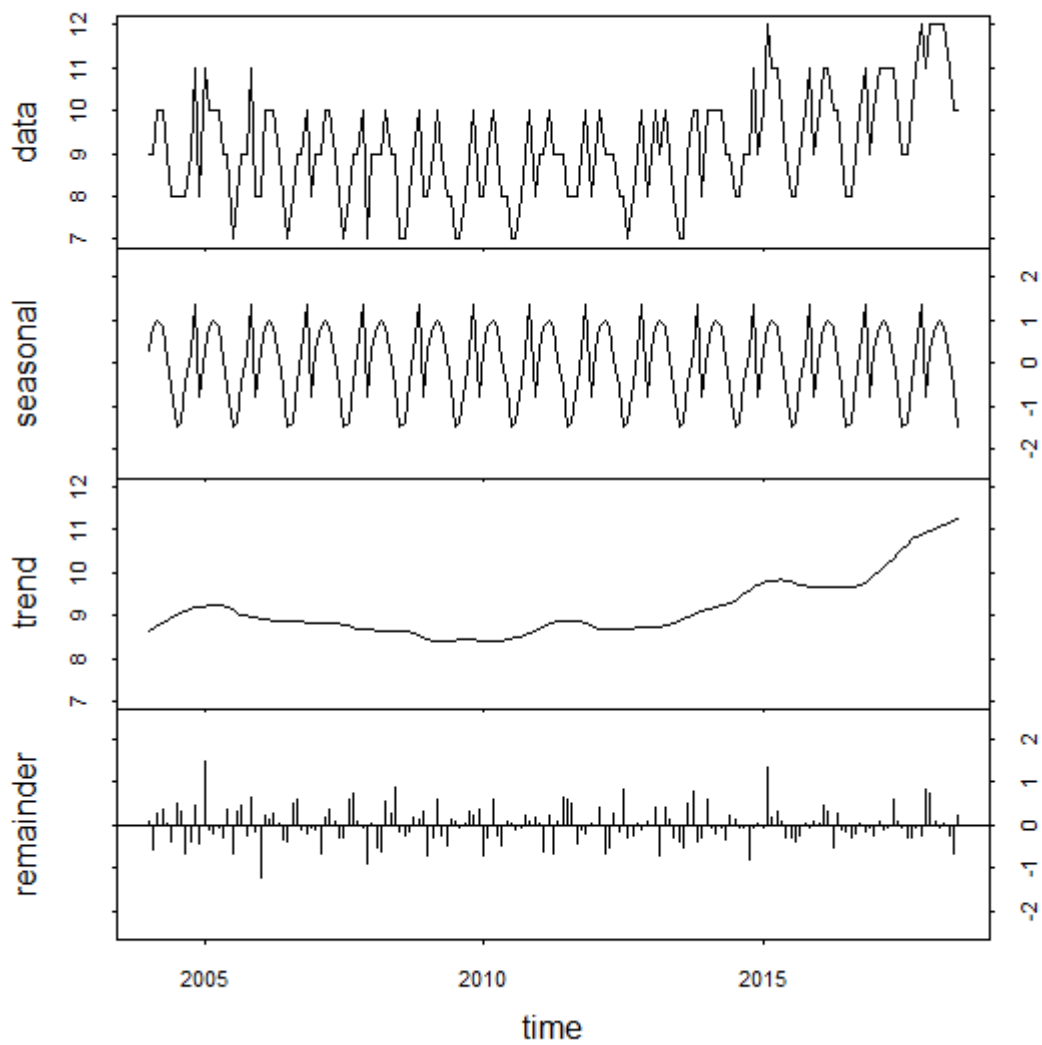
Anna Boehm, Alex Pizzini, Thomas Sonnweber, Judith Loeffler-Ragg, Claudia Lamina, Guenter Weiss, Ivan Tancevski

### Analysis of COPD-related terms and topics

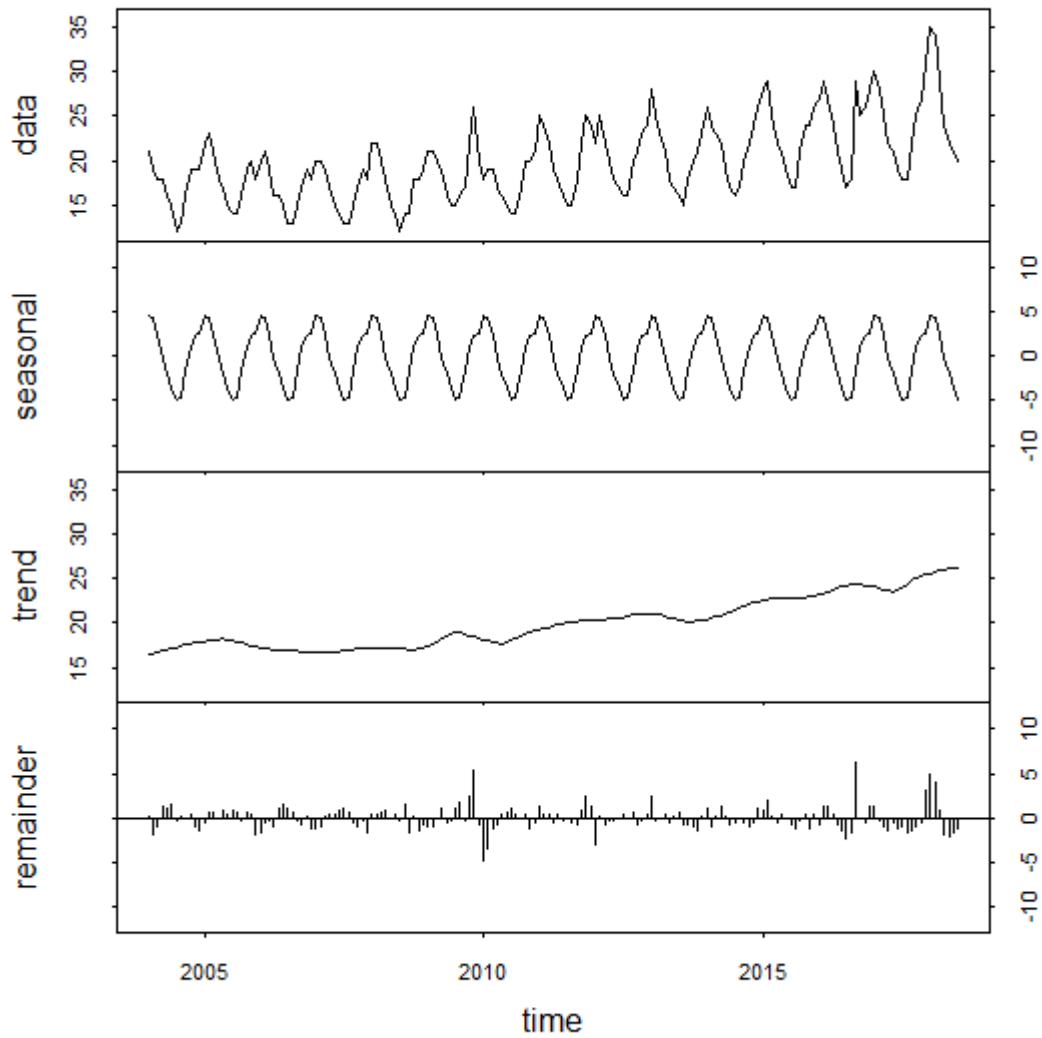
For a broader and more comprehensive analysis of the topic "chronic obstructive pulmonary disease", the following related terms were included in this online supplement: "asthma", "bronchitis" and "cough" (available as topics), and "emphysema", "chronic bronchitis" and "sputum" (only existing as terms on GT).

The "COPD"-related topics "asthma" and "bronchitis" revealed an RSV of  $22.17 \pm 3.47$  and  $10.42 \pm 2.81$  respectively (online supplementary **Figure S4**). The topic "cough" displayed an RSV which at some time-point even surmounted "diabetes", and in a direct comparison with "COPD" was searched 4.8 times more often.

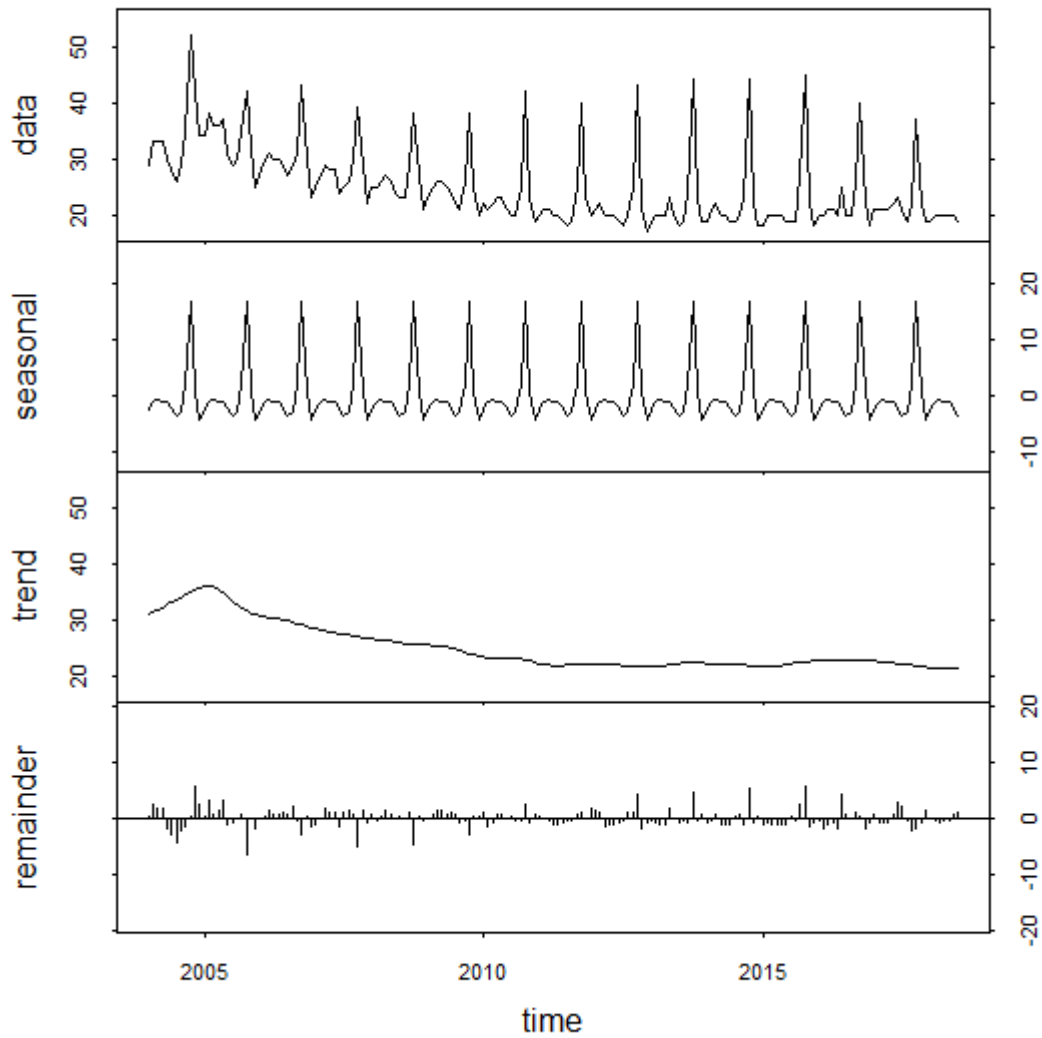
The terms "chronic bronchitis" and "emphysema" only existed as search terms and not as topics in GT, hence from a methodological point of view a direct comparison with "COPD" is not appropriate. However, if direct comparison is performed, "chronic bronchitis" displayed an RSV below 1.0 and "emphysema" showed an RSV of 1.0 to 2.0, with a single peak of 4.0 ("COPD" RSV in comparison: mean 9.15, SD 1.2).



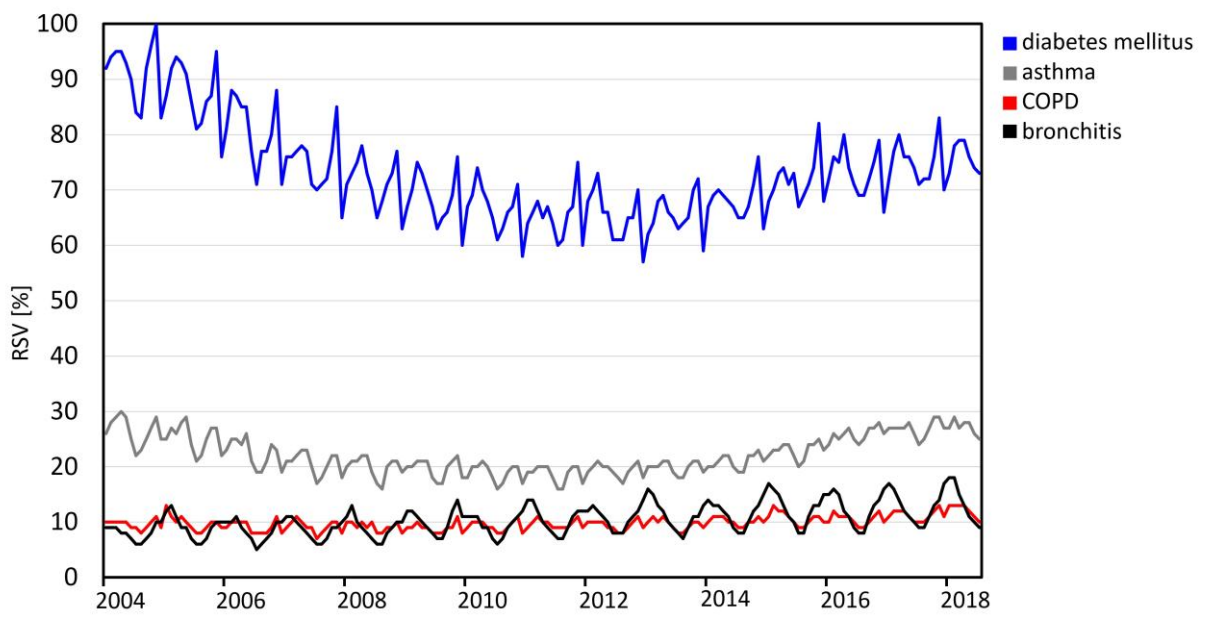
**Online Supplementary Figure S1:** Seasonal Decomposition of Time Series by Loess smoothing for **COPD**, showing (from top to bottom): the observed time series, the seasonal component, trend component and remainder. All y-axes are scaled to the same range.



**Online Supplementary Figure S2:** Seasonal Decomposition of Time Series by Loess smoothing for **Pneumonia**, showing (from top to bottom): the observed time series, the seasonal component, trend component and remainder. All y-axes are scaled to the same range.



**Online Supplementary Figure S3:** Seasonal Decomposition of Time Series by Loess smoothing for **Breast Cancer**, showing (from top to bottom): the observed time series, the seasonal component, trend component and remainder. All y-axes are scaled to the same range.



**Online Supplementary Figure S4:** Time-series for "diabetes mellitus", "asthma", "COPD" and "bronchitis" as search topics in GT.