Retracing the COVID-19 Pandemic in Germany from a Public Perspective using Google Search Queries Related to “coronavirus”

Aufarbeitung der COVID-19-Pandemie in Deutschland aus Bevölkerungssicht mithilfe von Google Suchanfragen zum Thema „Coronavirus“

Authors
Barbara Schuster¹, Linda Tizek¹, Maximilian C. Schielein¹, Stefanie Ziehfreund¹,², Kathrin Rothe³, Christoph D. Spinner³, Tilo Biedermann¹, Alexander Zink¹

Affiliations
1 Technical University of Munich, School of Medicine, Department of Dermatology and Allergy, Munich, Germany
2 Technical University of Munich, School of Medicine, Institute of General Practice, Munich, Germany
3 Technical University of Munich, School of Medicine, Institute for Medical Microbiology, Immunology and Hygiene, Munich, Germany
4 Technical University of Munich, School of Medicine, Department of Internal Medicine II, Munich, Germany

Key words
COVID-19, SARS-CoV-2, Pandemic, Google Trends, Health communication, public interest

Schlüsselwörter
COVID-19, SARS-CoV-2, Pandemie, Google Trends, Gesundheitskommunikation, Öffentliches Interesse

online publiziert 16.04.2021

Bibliography
Gesundheitswesen 2021; 83: e9–e14
DOI 10.1055/a-1398-5417
ISSN 0941-3790
© 2021. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence
Barbara Schuster
Klinik und Poliklinik für Dermatologie und Allergologie am Biederstein
Klinikum rechts der Isar der Technischen Universität München
Biedersteiner Straße 29
80802 München
Deutschland
barbara.schuster@tum.de

ABSTRACT

Aim of the study During pandemics, the whole population is simultaneously confronted with the same health threat, resulting in enormous public interest. The current COVID-19 pandemic has left the world in a unique state of crisis. The aim of this analysis was to explore whether Google searches can be used to retrospectively trace the COVID-19 pandemic in Germany and to detect local outbreaks by reflecting public interest in the virus.

Methods Google Trends was used to explore the relative search volume (RSV) related to “coronavirus” from January 2020 to July 2020 in Germany. The RSV ranging between 0-100 was compared to new SARS-CoV-2 infections per day on a national level and to the cumulative infection numbers on a state level, as well as to important infectiological and political events.

Results The most striking search peaks occurred after the first reported SARS-CoV-2 infection in Germany (January 27), during a major local outbreak in Heinsberg (February 25), after school closings (March 13) and the largest peak after nationwide contact restrictions (March 22) were announced. On a state level, peaks in RSV were observed after the first reported infection in each respective state. In addition, a higher RSV was recorded in states with higher numbers of infections (r = 0.6, p = 0.014) such as in Bavaria (RSV = 96, 391 infections/100,000 inhabitants) and Baden-Wüttemberg (RSV = 98, 340 infections/100,000 inhabitants). The lowest RSV (n = 83) and lowest number of infections (50 infections/100,000 inhabitants) was observed in Mecklenburg-Western Pomerania. Since the end of May, SARS-CoV-2 related RSV remained at low level even when numbers of infections were temporarily rising due to local outbreaks such as the outbreak in Gütersloh, North Rhine-Westphalia.

Conclusion RSV related to “coronavirus” precisely reflected public interest during the beginning of the COVID-19 pandemic. As public interest has strongly declined, information distribution regarding the newest developments over the entire course of the pandemic will be a major public health challenge.
Background

When at the end of 2019 the novel severe acute respiratory coronavirus 2 (SARS-CoV-2) first emerged in Wuhan, China [1, 2], the evolution of a global pandemic was underestimated. However, after the first cases of the virus were detected outside of China mid-January 2020, the World Health Organization (WHO) declared the outbreak a “Public Health Emergency of International Concern” on January 30, 2020 [3]. After renaming the new disease to coronavirus disease 2019 (COVID-19) on February 11, 2020 [4], the WHO declared a global pandemic on March 11, 2020 [5]. By the end of June 2020, more than 10 million documented cases of the disease were reported in more than 200 countries and territories, causing more than half a million deaths [6]. Globally, the pandemic caused social and economic disruptions, resulting in a global recession, and left the world in a unique state of emergency and uncertainty.

From a medical and public health perspective, epidemics and pandemics are some of the rare situations in which the whole population is confronted with the same health threat at the same time. As a consequence, in the first months of the pandemic, the “novel coronavirus” was the predominant topic in all conversations and media, ranging from newspapers, magazines, television, and radio to social media and blogs.

In Germany, 9 out of 10 inhabitants use the Internet and more than 60% of Germans use it to search for health-related information [7, 8]. Consequently, it is not surprising that during the COVID-19 pandemic, many people have turned to Google (Google LLC, Mountain View, California, USA) to obtain information about the virus. As a result, Google reported that the keyword “coronavirus” was searched for up to 4 times more frequently than the weather forecast during this period in Germany [9].

Prior studies have shown that analyzing Google searches can be beneficial for the detection of medical needs on a population level [10–12] and have even found correlations between the Google search volume for certain diseases and their incidence rates [13, 14]. In addition, it was suggested that Google searches can be a suitable tool for the early detection of disease outbreaks and epidemics [15–18]. However, studies further showed that Google searches are largely shaped by media coverage and political interventions, so they do not solely reflect the number of infections [18]. Some studies also found Google data to be unreliable for epidemiological problems [19].

Therefore, the aim of this analysis was to explore whether Google searches can be used to retrospectively retrace the COVID-19 pandemic and to detect local outbreaks by reflecting public interest in the virus.

Methods

Google Trends (GT) is a service by Google that provides information about the relative popularity of certain keywords or topics on Google over a certain period of time in selected regions. Data is available from 2004 to only a few days before the access date. For each day, GT provides the relative search volume (RSV) which represents the search results in proportion to the time point and location of a query. The search volume the topic of interest is divided into the total searches of the respective region to avoid that regions with the most search volume would always be ranked highest [20].

In the region with the highest relative popularity, the RSV is set to 100. In general, the RSV can range between 0 and 100. For example, a value of 50 means that the RSV is half as high as the RSV at its...
peak. Similarly, GT can be used to compare the RSV of different keywords or keywords in different regions; however, GT does not provide the exact search terms and their absolute frequency.

For this analysis, GT was used to explore the RSV related to the keyword “coronavirus” belonging to the topic “virus” from January 1 to June 30, 2020 in Germany, both on a national and on a state level. To examine the daily trends in RSV on a state level, 4 federal states were chosen based on the following reasons: Bavaria reported the first SARS-CoV-2 infection in Germany, North Rhine-Westphalia (NRW) experienced the first large German outbreak of COVID-19 in the community of Heinsberg, Saarland borders on the former high risk region Grand-Est in France [21], and Saxony-Anhalt was the last federal state to report its first case of COVID-19 in Germany. The RSV of Saxony-Anhalt, which had the highest overall peak in RSV, was used as a reference for all analyzes on a state level to keep the RSV values comparable. Thus, for the analyzes on a state level, a value of 100 corresponds to the highest overall peak in RSV observed in Saxony-Anhalt.

For comparison, the nationwide numbers of newly detected infections per day were extracted from the daily update of the Robert-Koch-Institut (RKI), the German national center for disease control [22]. In addition, the cumulative numbers of SARS-CoV-2 infections in the 16 German states were extracted from the RKI update from July 1, 2020 [23]. Based on these numbers and population data from the German Federal Statistical Office [24], the number of cumulative infections per 100 000 inhabitants was calculated for each state. The Pearson correlation between the numbers of infections per 100 000 inhabitants in the 16 German states and the respective RSV were calculated. The level of significance was set to α = 0.05. All statistical analyzes were conducted using IBM SPSS version 24 (IBM Corporation, Armonk, NY, USA).

As this analysis is solely based on publicly available data, institutional review board approval was not needed, and informed consent was not applicable.

Results

The RSV and the number of new infections over the course of the pandemic on a national level are displayed in Fig. 1. In the beginning of January 2020, when the outbreak was still restricted to China, the RSV was very low, resulting in a GT score of < 1. RSV started to increase around January 23, when the lockdown in Wuhan, China, was imposed. A few days later, on January 27, the first infection in Germany was reported in an employee near Munich, Bavaria, which led to the first peak in RSV (RSV = 24) one day later (January 28) following national news coverage. During the next week, more related cases were detected near Munich. However, as the chain of infection was reproducible and the outbreak was well contained [25], the RSV decreased steadily.

A second increase in RSV was observed starting on February 21, 2020, when the first European died of coronavirus in Italy, where the pandemic began to accelerate. The RSV peaked on February 28, 2020 (RSV = 58), during the course of the outbreak in Heinsberg, NRW, which began on February 24 with the first 2 reported cases [26] representing the beginning of a strong increase in infection rates in Germany. The third increase in RSV began on March 8, after the first infection had been reported in the ski resort town of Ischgl, Austria one day before. Along with increasing infection rates, a drastic increase in RSV could be observed (Fig. 1), which peaked on March 13, when 11 of the 16 German states announced the closing of schools and universities (RSV = 87). In the following days, RSV decreased at first but then peaked again on March 22, 2020, when nationwide contact restrictions were announced (RSV = 100). This day marked the overall peak of RSV during the observed peri-
od of time. The following “lockdown” period was characterized by a strong decrease in RSV, which was only interrupted 2 times correlating with announcements of continuing contact restrictions. In terms of infection numbers, the highest peak was observed on March 28 (6,294 new infections), almost one week after the contact restrictions had been announced, which was followed by a steady decline in infection rates. Since the end of May, both RSV and infection numbers remained at a low level (RSV around 5, numbers of new infections between 200 and 600), with only a small increase in infections observed following the local outbreak at a meat processing factory in the community of Gütersloh after June 17, 2020. Notably, RSV remained constantly low during this local outbreak.

On a state level, the Google search volume for 4 German states (Bavaria, North-Rhine Westphalia, Saarland, and Saxony-Anhalt) are displayed in Fig. 2. Additional to the general trends observed on national level, a strong increase in public interest was visible on the state level after the first case was reported in each respective state (Fig. 2). The first peak on January 28, following the first case near Munich, Bavaria, was especially high in Bavaria (RSV = 29), which was almost twice as high as in the other states. Similarly, there was a remarkably high peak in RSV in North-Rhine Westphalia (RSV = 59) after the detection of the first local case, which started the Heinsberg outbreak. In Saarland, the first case was detected on March 3, with a peak in RSV clearly visible the following day (RSV = 58), which was again not observed in the other states. There was also a noticeable peak (RSV = 22) in Saarland already on January 31, one day after 2 suspected cases were admitted to a hospital in Homburg, Saarland. In addition, the RSV was especially high in Saarland around March 15 (RSV = 78), the day the German borders were closed. Saxony-Anhalt was the last state to report its first infection of SARS-CoV-2 on March 10. However, as public interest in the virus was increasing all over Germany at this point in time, the increase in RSV is not as conspicuous. The RSV for all 16 German states can be found in the supplementary material (S1–S16).

![Fig. 2 Relative Search Volume of Google related to “coronavirus” in the course of the COVID-19 pandemic in 4 German states.](image)

![Fig. 3 Relative Search Volume (RSV) in Google related to “coronavirus” and infections per 100,000 inhabitants (i) in the 16 states of Germany.](image)
RSV and infections per 100,000 inhabitants over the whole study period in the 16 German states are displayed in ▶ Fig. 3. Over the whole observed period, RSV was highest in Saarland (RSV = 100) and lowest in Mecklenburg-Western Pomerania (RSV = 83; ▶ Fig. 3). When considering cumulative infections per 100,000 inhabitants (i), the highest infection rate was recorded in Bavaria (i = 391) and the lowest in Mecklenburg-Western Pomerania (i = 50). RSV and infection rates over the whole study period correlated moderately, with states that have a higher number of cumulative infections showing a higher RSV (r = 0.6, p = 0.014).

**Discussion**

The study results show that the relative search volume (RSV) related to “coronavirus” reflected the public interest during the pandemic in Germany very well on a national as well as on a state level. According to our results, public interest in the virus peaked on March 22, 2020, when the nationwide contact restrictions were announced. Since then, public interest has drastically decreased and has remained at a low level since the end of May.

We found a moderate correlation between the RSV and the number of infections over the whole study period in the 16 German states, suggesting that Google searches can be used as an indicator to identify regions particularly affected by epidemics. This is in line with previous studies which found that GT can be a suitable tool for the detection or monitoring of disease outbreaks [15–18]. However, public interest in COVID-19 was not driven by numbers of infections alone, as it was also largely shaped by media coverage and political interventions. Especially the news of school closings and contact restrictions resulted in very high RSV on the days of the respective announcements. In contrast, RSV was drastically decreasing after its overall peak on March 22, even though infection rates were at their maximum at that time. As decreasing RSV was also described for the Ebola pandemic in western Africa in 2014, it could be a trend during epidemics and pandemics that public interest declines after a few initial peaks [18]. This hypothesis is supported by the fact that the recent outbreak in Gütersloh did not lead to an increase in RSV, even though infection rates were rising. Fading public interest in the course of a pandemic, however, constitutes a major public health challenge, as it is important to keep people informed about the newest developments and interventions during a pandemic even after longer periods of time [27]. Low RSV during the course of a pandemic could signal that a more thorough communication strategy is necessary for the dissemination of important pandemic-related information and regulations in order to reach the whole population, like e.g. community-based approaches [28]. Monitoring RSV for related symptoms (e.g. fever, loss of smell) in addition to the RSV for the pandemic disease itself could help to allow detection of new outbreaks even in times of low public interest in the pandemic [29].

While in general the trends in RSV on a state level were comparable to the course of RSV on a national level (e.g. peaks after the first German case and after announcements of public measures), we found that not only infection rates but also regional events and special regional characteristics influenced RSV on a state level. For example, regional peaks in RSV were observed after the first infection in each state. As another example, when the German borders closed on March 15, RSV was especially high in Saarland, which has a large number of cross-border commuters due to its proximity to France and Luxembourg.

Several limitations apply when working with Google data. First, the data obtained by GT only reflect the RSV, which measures the share of queries related to the topic in selected regions, regardless of the number of actual search queries [20]. A higher RSV therefore does not necessarily mean an increase in search queries in absolute numbers, but only that the relative share of search queries was higher which could be due the fact that people searched for other topics less frequently as normally. In contrast to absolute numbers, a comparison with previous studies on other health topics is hardly possible when only having the RSV. Second, the estimations of the RSV is automatically provided by GT and do not contain any information on how many search terms were considered for the calculation [20]. Accordingly, the RSV might be different when different sets of search terms are considered and when more regions are investigated in the analysis. Finally, ecologic studies such as the present analysis are vulnerable for errors and incorrect conclusions, as they merely rely on aggregated data and correlation. Thus, they are not suitable to determine causality, but they can provide valuable clues about associations which should be further explored.

In conclusion, the study results show that Google search data reflect public interest during pandemics. While media coverage and political interventions seem to be the main drivers of public interest, a moderate correlation between the cumulative number of infections and the RSV on a state level during a pandemic was observed. This suggests that GT may be able to reflect and maybe also detect local outbreaks of infectious diseases. Naturally, public interest in COVID-19 seems to have drastically decreased. Thus, keeping people informed about the newest developments and aware of the ongoing risk of infection seems to be a major public health challenge in the later course of a pandemic, which needs to be addressed in favor of containment.

**Conflict of Interest**

The authors declare that they have no conflict of interest.

**References**


[22] Robert Koch Institut. Aktueller Lage-/Situationsbericht des RKI zu COVID-19 Im Internet https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html Stand: 24.06.2020

[23] Robert-Koch-Institut. COVID-19: Fallzahlen in Deutschland und weltweit. Im Internet: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html


