External validity of Google search queries for seasonal and cross-annual trends of disease risks – An exploratory study using the example of the incidence of Lyme disease and hay fever

Claudia Kohring • Manas K. Akmatov • Jakob Holstiege • Jörg Bätzing

DOI: 10.20364/VA-25.02

Abstract

Background

With its online service *Google Trends*, Google offers the possibility of evaluating the frequency of search queries nearly in real time. Particularly in the context of changing health risks due to climate change, the timely detection of fluctuations in the number of new cases, both in terms of seasonality and global trends across years, would be highly relevant to healthcare. However, studies on the external validity of *Google Trends* data for this purpose are lacking. In this study, seasonal and annual *Google Trends* data are compared with the corre - sponding incidence trends for Lyme disease and hay fever to investigate whether *Google Trends* data are corresponding for depicting changing seasonal patterns and disease risks.

Methods

The reference data for the comparison with the *Google Trends* data to test the external validity are nationwide claims data from SHI - accredited physicians (according to § 295 Social Code Book 5, SGB V) and, if available, notification data from the Robert Koch - Institute (RKI) according to the corresponding state regulations. The cumulative incidence of diagnosed Lyme disease and hay fever in the period from the first calender quarter of the year 2013 to the fourth quarter 2021 was calculated from the claims data. After normalisation to values in the range 0 - 100, claims and notification data were compared with *Google Trends* data and visualised.

Results

The incidence of diagnosed Lyme disease and hay fever was 22 and 139 per 10,000 persons in 2021, respectively. While there were predominantly strong correlations in the seasonal patterns between the data sources for both diseases, the correlations between the annual trends were only weak.

Discussion

Google Trends data could be a possible source of information for monitoring changes in the start of the season or the period in which the number of new cases increases. The data do not have sufficient external validity to estimate cross-annual trends in the number of new cases in the population.

Corresponding author: Claudia Kohring
Central Research Institute of Ambulatory Health Care in Germany (Zi)
Salzufer 8 – 10587 Berlin – Tel. +49 30 2200 56133 – E-Mail: CKohring@zi.de

Report No. 25/02, published May 15, 2025

Key words

Claims data, Google, hay fever, incidence, infodemiology, Lyme disease, validation

Citation

Kohring C, Akmatov MK, Holstiege J, Bätzing J. External validity of Google search queries for seasonal and cross-annual trends of disease risks – An exploratory study using the example of the incidence of Lyme disease and hay fever. Central Research Institute of Ambulatory Health Care in Germany (Zi). Versorgungs-atlas-Report Nr. 25/02. Berlin 2025. URL: https://doi.org/10.20364/VA-25.02

Key messages

- The risk of newly diagnosed hay fever showed a strongly fluctuating trend, but on average an annual increase in incidence of 1.9 percent.
- The incidence of Lyme disease was initially characterized by marked decline and followed by a re-increase.
- *Google Trends* data captures the seasonality of hay fever and Lyme disease incidence, which is comparable to German physicians' claims data.
- Yearly trends, and thus longer cross-annual changes in disease risk in the population, were not reliably reproduced.

Background

Google's online service *Google Trends* is a tool for measuring the frequency of searches (1, 2), which has already been in use in the field of health research for a number of years (3). Options for regional differentiation and real-time availability where required make using *Google Trends* data particularly attractive (4). The underlying research approach is based on the assumption that there is a connection within populations between the health situation and patterns of information and communication online (5). One well-known practical application of Google data is its use in predicting the annual flu season (6-8). Using archived *Google Trends* data, the authors of a study (7) published in 2019 concluded that the inclusion of such data can improve the reliability of the traditional surveillance system in forecasting flu and other flu-like illnesses. In a review, Ziehfreund et al. (9) examined the potential of this data source for Germany in relation to various illnesses and concluded that under certain conditions and taking into account the inherent limitations, these data could be used to map current epidemiological developments over space and time.

The coronavirus pandemic in particular has highlighted importance of the availability of real-time data for mapping the health situation within the population, and it will continue to gain increasing significance for the early identification of public health action areas with respect to the possible consequences of climate change. *Google Trends* data could take on an important role in this respect as it can be used to pinpoint up-to-date information on morbidity trends within the population. Current health risks related to climate change that may require rapid action include the increase in respiratory illnesses and infectious diseases, for example. A retrospective investigation of pollen drift between 2001 and 2021 found increasing linear trends for allergy-relevant pollens in Germany with a small number of exceptions (10). As the number of people in Germany with sensitivities to ambrosia pollen could more than triple to 16 million over the next 20 to 40 years, this country is considered to be especially at risk (11). Climate change also presents survival advantages for the vectors that propagate vector-borne infectious diseases, which may increase their abundance and propagation areas, extend phases of activity and precipitate the immigration of new vectors, giving rise to new infectious agents as a result. This in turn increases the probability of human infection (12, 13). Transmitted through tick bites, Lyme disease is coming under particular scrutiny due to the significant burden of disease it causes in Germany (14-16). Measured on the basis of outpatient claims data or available notification data, Lyme disease and hay fever show strong seasonal variations within a calendar year (16-18).

The majority of the studies included in a 2014 systematic review (3) aimed to describe temporal or spatial trends or to examine potential causal inferences based on *Google Trends* data. Only about a third of these studies addressed the possibility of surveillance, and these could be sub-classified as prognostic or monitoring studies. In almost all of these studies, the *Google Trends* results were validated using external data. However, the quality of the reference data used was not the focus of the review. The authors of a subsequent review concluded that most of the included studies compared official data with other internet-based data series, but did not quantify or differentiate between them (19).

Nevertheless, it is crucial to evaluate the external validity of novel data sources, such as *Google Trends* data, by referring to established data. This is necessary to ascertain the basic suitability of this information source for studies examining changing morbidity trends over various time horizons. While acknowledging their inherent limitations, the scope and availability of routine and notification data renders them well-suited as comprehensive reference data in the sense of a gold standard for the external validation of *Google Trends* data. It has previously been demonstrated that long-term and medium-term trends can be identified through the analysis of physician claims data (20-28). If *Google Trends* data could be used to trace overall trends in the sense of cross-annual developments, and the extent of the seasonal occurrence and new cases of Lyme disease and hay fever over longer periods of time, further opportunities for estimating current morbidity developments in situations where established data sources are not yet accessible may also be conceivable. *Google Trends* data, which has been proven to accurately represent trend developments, could enable early reactions to changing disease patterns.

The aim of this exploratory-descriptive study is to compare annual and seasonal incidence trends for Lyme disease and hay fever diagnosed by SHI-accredited physicians with the corresponding Google search trends for these diseases between 2013 and 2021. The comparison for Lyme disease also draws on case numbers published by the Robert Koch-Institute (RKI) (29).

Methodology

Data and study population

The data basis for the present study consists of nationwide SHI-accredited physician claims data from the 17 Associations of Statutory Health Insurance Physicians regions in Germany in accordance with Section 295, Book 5 of the German Social Code (SGB V) for the years 2010 to 2021. The data include information about all individuals with statutory health insurance of any age who accessed SHI-accredited physician or psychotherapeutic care services at least once in a given calendar year (2021: N = 72,866,128). In addition to sociodemographic details such as age, sex and residential area, the dataset also comprises information on individual treatment cases including official physicians' fee scale items (Gebührenordnungspositionen, GOP) based on the Uniform Value Scale (Einheitlicher Bewertungsmaßstab, EBM), participating medical and psychotherapeutic service providers, and diagnoses based on the International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification (ICD-10-GM).

With its *Google Trends* service, Google offers the opportunity to query monthly search volumes for keywords within a specific period and a specific region (1). To this end, Google provides relative details within a value range of 0 to 100, based on the maximum regional value during the period under review (1, 2). For the present study, the monthly searches for the German search terms "borreliose" (Lyme disease) and "heuschnupfen" (hay fever) were requested on 30 June 2022 for the period 2013 to 2021, limited to the region of Germany. As far as possible, colloquial terms specific to both diseases were used in the selection of the search terms. Queries were performed separately to obtain normalised data for each individual search term; in addition, search terms were not placed in quotation marks, as this would have limited the search to exact matches. Data were not narrowed down by category or search type ("all categories" and "web search" settings). The *Google Trends* data were then aggregated per quarter and renormalised to values within the range 0 to 100. This aggregation was necessary because diagnoses in claims data of SHI-accredited physicians are only available at the quarterly level.

The weekly notification data on Lyme disease were accessed for the period 2013 to 2021 via the RKI's SurvStat@RKI 2.0 website (30). In accordance with supplementary state regulations, Lyme disease is a notified disease in nine German federal states (Bavaria, Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saarland, Saxony, Saxony-Anhalt, Thuringia and Rhineland-Palatinate). Cases of the disease are published on the basis of a uniform national case definition established by the RKI (29).

Case definition and incidence calculation based on outpatient diagnoses

A previous observation period of at least three years was applied as a prerequisite for the inclusion of patients in the study population for the incidence calculation. Individuals accessing SHI-accredited physician services during the respective reporting year were deemed to have fulfilled the inclusion criterion of a previous observation period of at least three years, if they also used SHI-accredited medical care at least once in the calendar year three years before. In this context, persons were considered newly infected if no corresponding diagnosis had been documented for them during these three years of previous observation. Cases of Lyme disease were identified based on the ICD-10 diagnosis code A69.2 and hay fever cases were identified using the diagnosis codes J30.1 and J30.2, which were documented respectively with the additional identifier "G" ("gesicherte Diagnose", or verified diagnosis). Using additional identifiers for disease certainty is mandatory in German outpatient care.

Newly diagnosed patients were used for this study in order to examine the informative value of the claims data as a data source with regard to changing disease risks over time, i.e. the development of new diseases in the population. We further assume that people search for complaints and symptoms on Google when these are new and they are unfamiliar with them. The annual cumulative incidence was calculated for the period 2013 to 2021; diagnoses of Lyme disease or hay fever had to be documented in at least one quarter of the respective calendar year (M1Q criterion). For outpatient diagnoses data, the data set used does not contain an exact diagnosis date, but only a quarterly reference as the smallest time unit. The quarterly cumulative incidence among the population receiving care from SHI-accredited physicians was calculated to match seasonal fluctuations. To match seasonal patterns within a specific year, new cases were allocated to the quarter in which the initial diagnosis occurred. The denominator populations were adjusted yearly for prevalent cases. In principle, reinfections of Lyme disease can occur. In accordance with the selected case definition, these were only considered as reoccurring incident cases if three years without a diagnosis had been recorded prior to the reinfection. Patients with hay fever can be re-assessed as incident cases as well, if they had another three years without a diagnosis.

Statistical analyses

To establish comparability with the *Google Trends* data, the quarterly cumulative incidence between 2013 and 2021 for Lyme disease and hay fever was converted to relative values from 0-100. As part of this process, the respective maximum value during the study period was set separately for each disease to 100 (Lyme disease: Q3/2020, hay fever: Q2/2013) and the lower values were expressed relative to this value. The notification data on Lyme disease were also normalised after quarterly aggregation.

To investigate linear associations between the *Google Trends* data, the SHI-accredited physician claims data and the RKI notification data, Pearson correlation coefficients were calculated for the quarterly and annual normalised values for the nine-year study period. Furthermore, a comparative visualisation was carried out for the normalised seasonal and annual trends from the various data sources.

Results

Annual and quarterly incidence trends

The population at risk of newly-diagnosed cases of Lyme disease in 2021 comprised approximately 66.5 million individuals with a statutory health insurance cover, of which 145,730 were identified as having incident Lyme disease (**Table 1**). This corresponds to a cumulative incidence of 22 cases of Lyme disease per 10,000 people in 2021. During the first three study years, a relatively significant decline was observed (2013 vs. 2015: -23.2 %), which was then followed by a further increase (**Table 1**). Throughout the study period, a slight upward trend in incidence was observed, with an average annual increase of 0.4 % (median: -5.5 %). The significant difference between the median and the average annual increase is due to the decline in incidence at the beginning of the study. When the quarterly incidence is considered (**Figure 1A**), a seasonal pattern emerges, with the lowest incidence occurring in the first quarter (January to March) of each year and the highest incidence in the third quarter (July to September). The extremal quotient was an average of 4.2 for the individual study years and 5.4 for the entire study period (highest incidence in the third quarter of 2020: 11 / 10,000; lowest incidence in the first quarter of 2016: 2 / 10,000).

In 2021, a new diagnosis of hay fever was documented among 845,537 individuals with statutory health insurance throughout Germany (population at risk 2021: 61 million individuals); the corresponding cumulative incidence was 139 cases per 10,000 people (**Table 1**). The incidence dropped by -14.3 % from 2013 to 2017 with a number of fluctuations in between and then rose by 32.4 % from 2017 to 2021. The average annual increase was 1.9 % for the entire period (median: 2.1 %). At the quarterly level, an annually recurring pattern was observed with the highest incident values regularly occurring in the second quarter (April to June) and the lowest values in the fourth quarter (October to December) (**Figure 1B**). The extremal quotient for the individual study years was an average of 3.0, while the value for the entire study period was 3.7 (highest incidence in the second quarter of 2013: 59 / 10,000; lowest incidence in the fourth quarter of 2017: 16 / 10,000).

Table 1: Size of the population at risk (N), absolute number of new diagnoses (n), cumulative incidence per 10,000 people and relative change compared to the previous year for Lyme disease and hay fever by calendar year for the period 2013 to 2021.

	Lyme disease				Hay fever			
Year	N	n	Incidence	Relative change to previous year (%)	N	n	Incidence	Relative change to previous year (%)
2013	62,004,150	140,503	23		59,497,513	726,816	122	
2014	62,657,786	121,146	19	-14.7	60,041,255	760,795	127	3.7
2015	63,268,862	110,085	17	-10.0	60,552,482	721,205	119	-6.0
2016	63,769,969	141,411	22	27.4	60,948,516	739,129	121	1.8
2017	64,080,337	134,903	21	-5.1	61,235,995	640,730	105	-13.7
2018	64,655,035	140,057	22	2.9	61,748,950	732,787	119	13.4
2019	65,030,482	132,579	20	-5.9	62,048,890	751,543	121	2.1
2020	65,177,549	155,778	24	17.2	62,161,689	769,141	124	2.2
2021	66,537,907	145,730	22	-8.4	61,012,766	845,537	139	12.0

Note: The cumulative incidence was rounded to whole numbers; the values for the relative change compared to the previous year refer to the incidence rounded to two decimal places.

Data basis: Statutory health insurance-accredited physician claims data according to Section 295, Book 5 of the German Social Code (SGB V)

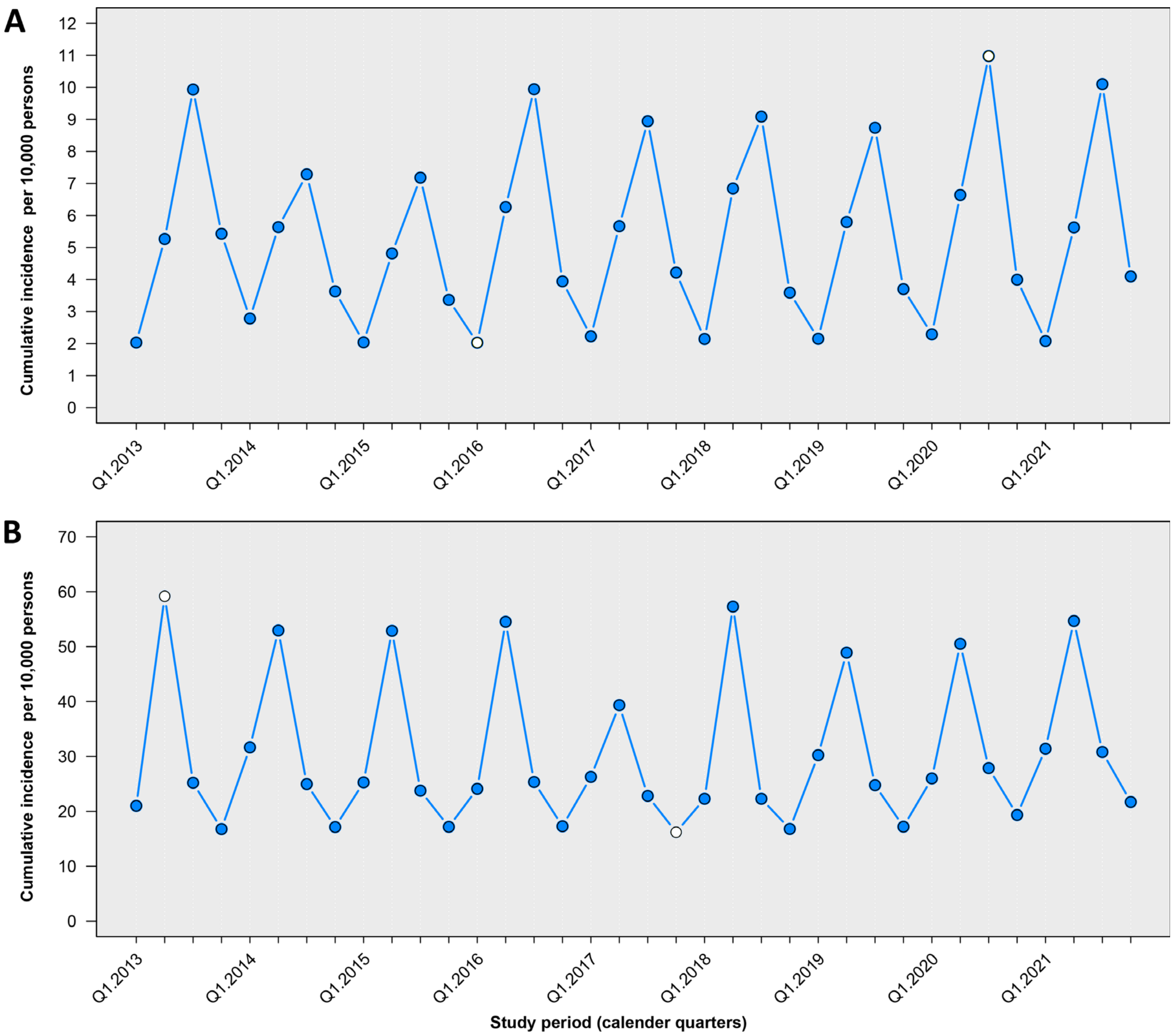

Figure 1: Temporal development of the cumulative incidence of Lyme disease (A) and hay fever (B) at quarterly level. Minimum and maximum over the entire period are highlighted in white.

Note: Q1, quarter 1, marks the first calender quarter of the respective year.

Data basis: Statutory health insurance-accredited physician claims data according to Section 295, Book 5 of the German Social Code (SGB V)

Comparison of the various data sources with respect to seasonal variations

For comparison of the claims data and *Google Trends* data, the quarterly incidence values determined using the claims data were first normalised as described above. Graphically consistent seasonal patterns emerged in both data sources for Lyme disease as well as hay fever. For Lyme disease, the maximum number of *Google Trends* data exceeded the maximum diagnosis frequency indicated in the SHI-accredited physicians' claims data in some years (**Figure 2A**); a strong positive correlation was observed between the claims data and the *Google Trends* data, with a Pearson correlation coefficient of 0.86. The correlation between the notification data and *Google Trends* data was markedly weaker, at r = 0.69. For hay fever, the patterns for the minimum and maximum values were in line with each other for all years investigated (**Figure 2B**); this strong positive association was confirmed by a correlation of 0.97.

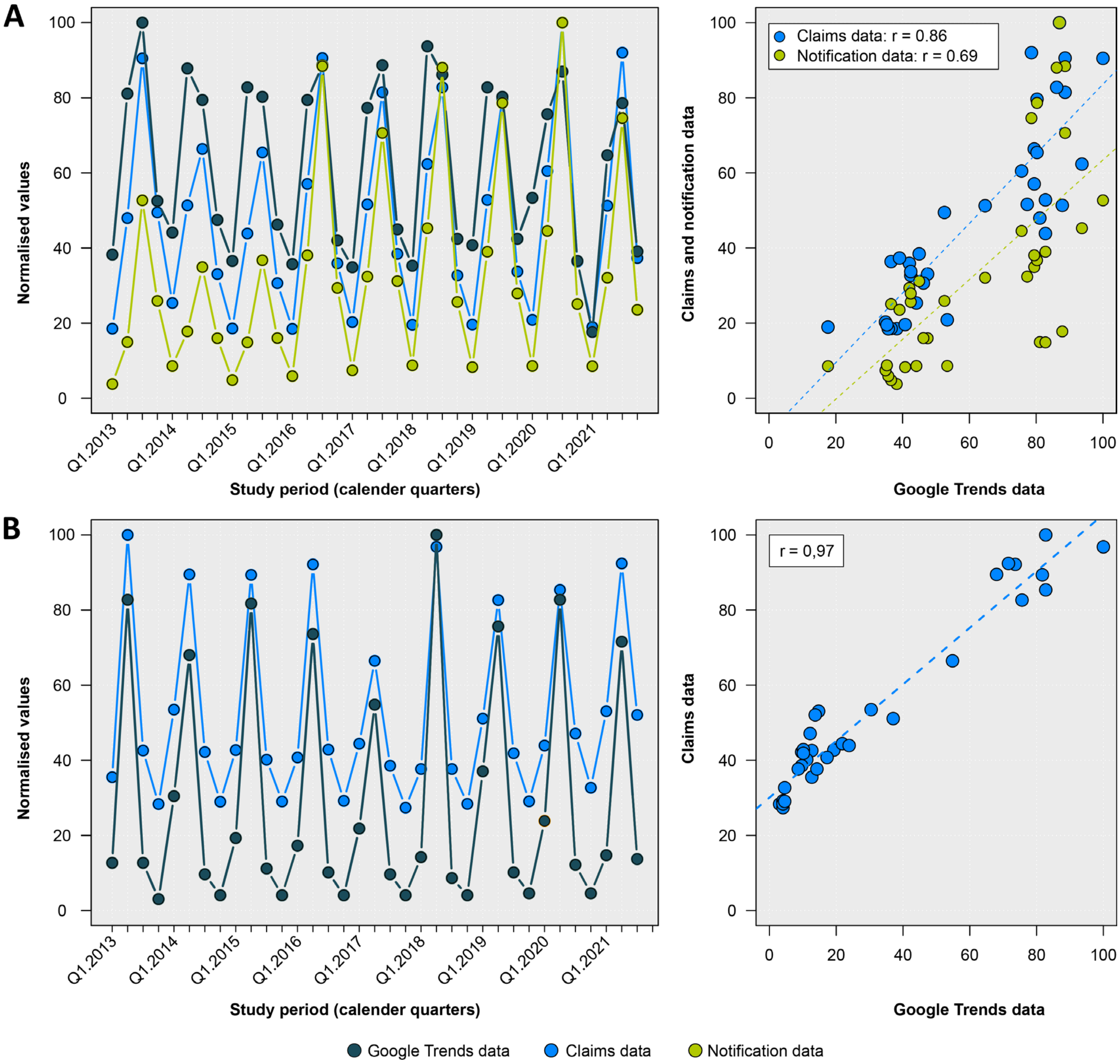

Figure 2: Comparison of the normalised values for the *Google Trends* data, the quarterly incidence and, where available, the number of cases reported to the Robert Koch - Institute shown on the left and on the right, the associated correlation diagrams including Pearson correlation coefficient (r) for Lyme disease (A) and hay fever (B) for the period Q1 / 2013 to Q4 / 2021 in Germany.

Note: Q1, quarter 1, marks the first calender quarter of the respective year.
Data basis: Statutory health insurance-accredited physician claims data according to Section 295, Book 5 of the German Social Code (SGB V)

Comparison of the various data sources with respect to annual trends

The annual trends for the normalised values from the three data sources for 2013 to 2021 are shown in **Figure 3A**. Greater deviations were observed before 2016 in particular; this applied above all to the normalised RKI notification data. The data trends converged between 2016 and 2020, but then diverged further in 2021; nevertheless, all data trends indicated a decline. There was no correlation between the *Google Trends* data and the claims data. A slight negative correlation of - 0.19 was observed between the *Google Trends* data and the RKI notification data. There was a weak positive correlation for hay fever at the annual level (r = 0.25). A similar trend was observed until 2017, with larger deviations in the subsequent years, which indicated conflicting trends after 2018 (**Figure 3B**).

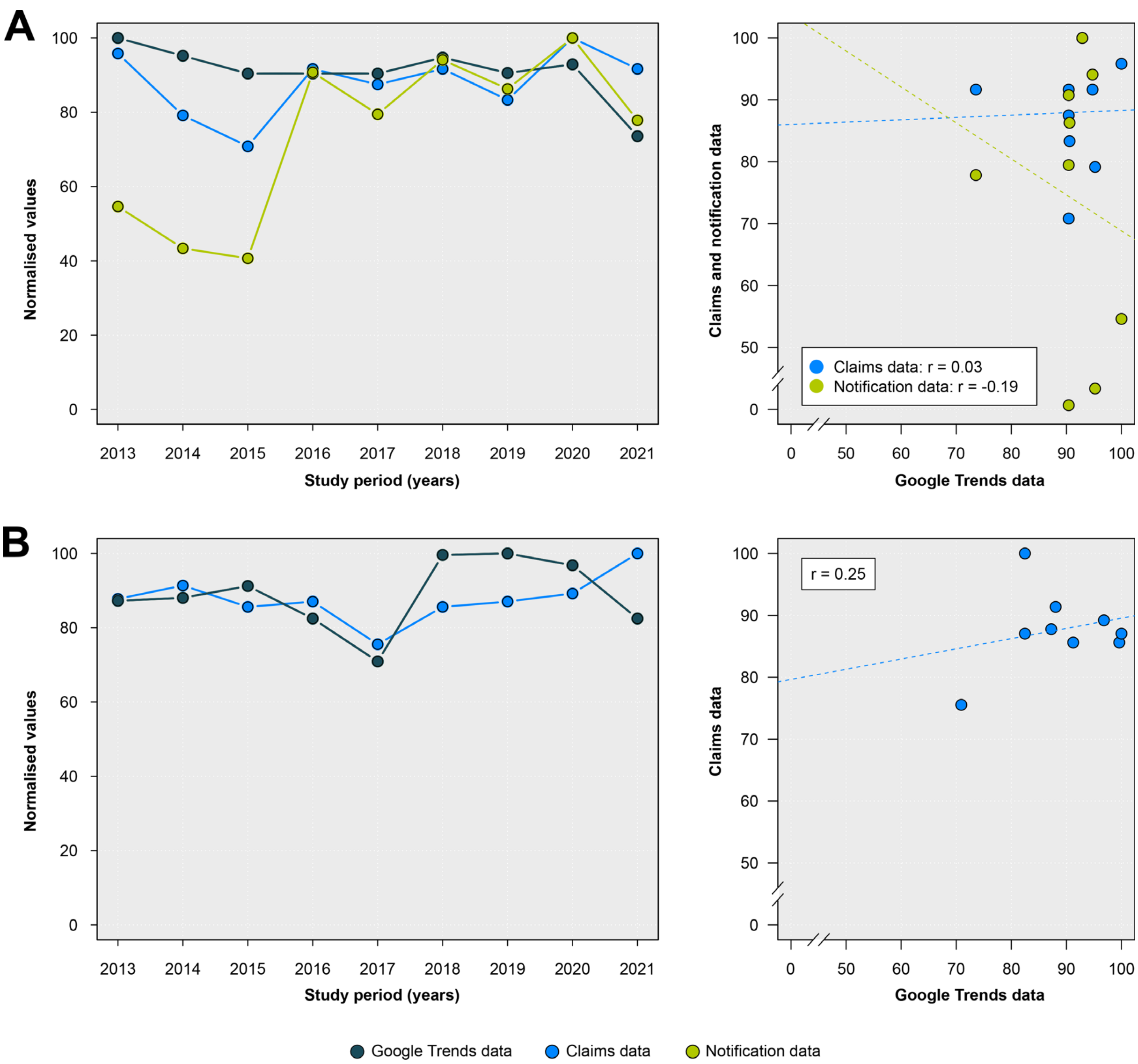

Figure 3: Comparison of the normalised values for the *Google Trends* data, the annual incidence and, where available, the number of notified cases shown on the left, and on the right the associated correlation diagrams including Pearson correlation coefficient (r) for Lyme disease (A) and hay fever (B) for the period 2013 to 2021 in Germany.

Note: Q1, quarter 1, marks the first calender quarter of the respective year. Data basis: Statutory health insurance-accredited physician claims data according to Section 295, Book 5 of the German Social Code (SGB V)

Discussion

To the best of the authors' knowledge, the present study was the first to map quarterly and annual incidence trends for Lyme disease and hay fever across a period of nine years for all age groups among the population of individuals with statutory health insurance in Germany [share of the total German population in 2020: 88.1 % (31)]. In the claims data, the annual risk for newly diagnosed hay fever demonstrated a highly fluctuating course over the 9-year study period, but on average a yearly increase of the incidence about 1.9 percent. In contrast, the incidence of diagnosed Lyme disease was characterized by a strong declince between 2013 and 2015, and a re-increase in subsequent study years. The seasonal pattern of online search behaviour associated with disease-specific search terms and the reference data from SHI-accredited physician claims data, and RKI notification data for Lyme disease, are highly correlated. On the other hand, the annual trends in the normalised values for the incidence in the claims data reveal only weak correlations. The same applies to the comparison of *Google Trends* data and notification data for Lyme disease.

In 2021 and 2022, comprehensive national prevalence and incidence data for both diseases was published for the first time on the Health Care Altas website (www.versorgungsatlas.de) and placed in the national and international context (15-17). Both the results published previously for Lyme disease and those published for hay fever are confirmed in the present study with extended study populations and case definitions. More recent studies based on notification and SHI routine data are available for Lyme disease, which cover shorter time periods [2015-2019 (32) and 2016-2020 (33), respectively], but consistent trends. National incidence trends up to and including 2021 are now available for Lyme disease and indicate a decline in incidence during the study period, although this decline is not constant and was disrupted by a sharp increase during the first year of the coronavirus pandemic in 2020. Due to the methodological approach using a diagnosis-free preobservation time of three years, it cannot be ruled out that reinfections were not captured in some cases, which may result in slight underestimation. For hay fever, we supplement the data on children and adolescents published on the Health Care Atlas website with incidence data on the entire population of individuals with statutory health insurance without age limitations, and observe corresponding seasonal trends in incidence.

Studies published to date on *Google Trends* data for Lyme disease and hay fever (or any other atopic diseases) have largely focused on estimating prospective developments in disease frequency based on users' online search behaviour. For example, Kapitány-Fövény et al. (34) investigated whether the use of *Google Trends* data could improve the prediction of Lyme disease incidence in Germany in addition to the notification data published by the RKI. Although this was not the case, a high degree of correlation was observed between both data sources for the retrospective values. This correlation between internet searches and Lyme disease case numbers published by the RKI was also analysed by Scheerer et al. (35) in addition to correlations between online searches and temperature data published by the German weather service. Strong correlations between the notification data and the temperature data were also observed in this study. In a previous investigation of Lyme disease (16), we demonstrated that although the number of Lyme disease diagnoses by SHI-accredited physicians was many times higher than the figures reported by the RKI, the development of both sets of figures over time was largely consistent. The deviations in the notification data at the beginning of the study period, which also arose in the present study, are probably caused by the under-documentation of notified cases. A study investigating the mapping of the seasonality of hay fever in Europe using *Google Trends* data performed by Bousquet et al. (36) revealed country-specific differences in the usage of various search terms. In a subsequent study (37), the French research team demonstrated clear correlations between the *Google Trends* data and data on pollen concentrations, although these correlations were not very pronounced. One commonality between the investigations (34, 35, 37) referenced above is that they did not focus on the external validation of *Google Trends* data.

The results of the present study showed very strong correlations between the *Google Trends* data and utilisation of SHI-accredited physician services in terms of the quarterly view and the seasonality it maps for Germany. The correlation between the claims and notification data for Lyme disease was also strong, while the correlation between *Google Trends* data and notification data was less pronounced, but still high. When the years in which notified cases were likely considerably under-documented are excluded, the positive correlation becomes markedly stronger (2016-2021: r = 0,70, data not shown). It should also be noted that *Google Trends* and RKI notification data could in principle also be compared on a weekly basis, but this was not the focus of this study. These results indicate that *Google Trends* data could act as a possible source of information concerning shifts in the beginning of the season with increased numbers of new diagnoses as well as changes in the duration of seasonal increases. The advantage and potential of *Google Trends* data over claims data is that it can also be retrieved on a monthly or even weekly basis and is always available in the most up-to-date form. In the comparisons carried out, changes in seasonal patterns may be concealed or only very imprecisely depicted as a result of the reference to calendar quarters as the smallest possible unit of time.

When only the overall trends at the annual level are compared, the correlation is markedly weaker. There is no reliable identification of years with relatively low or high numbers of new cases, which limits the scientific usefulness of *Google Trends* data for longer-term, cross-annual trends in the population. Consequently, *Google Trends* data do not appear fully suitable for use as a means of surveillance for the early estimation of disproportionate increases in the number of new cases, which are highly relevant with respect to the changes in environmental exposure (e.g. through plant pollen, vector spread) associated with the predicted climate conditions. The overall trends may be influenced to a greater extent by social factors such as one-off spikes in interest among the population (e.g. media reporting), but this rise is not necessarily accompanied by higher levels of interest due to seasonally varying health problems. At this point, the necessary concentration of the monthly *Google Trends* data at a quarterly and annual level and the related renormalisation of the values may cause imprecisions in the yearly mapping in particular. Although there may be many influencing factors in this regard, it is not possible to quantify their respective weights more specifically in the search results. Even if distinctive media coverage or other factors could be identified based on the research topic, it would not be possible to adjust the *Google Trends* data for it. In view of this, prospective longer-term wide-ranging applications, such as the targeted addressing of information requirements or even the regional distribution of resources for public health measures, as addressed by Ziehfreund et al. (9), should be discussed with caution. In addition, data on health complaints from the Structured Medical Assessment in Germany (SmED), recorded in the telephone assessments of the 116117 on-call medical service, provide a resource that is comparable in its temporal and spatial granularity and may be more suitable for such time series analyses due to its more specific and structured purpose. In early analyses, the SmED data already proved to be a valid indicator for the start and end of various acute respiratory infection waves (38, 39).

Strengths and limitations

The present study is based on German national SHI-accredited physician claims data for all individuals in Germany with statutory health insurance [2021: 88.1 % (31)]; individuals included in the study must have accessed SHI-accredited physician services at least once in the respective calendar year. The limitations of the SHI-accredited physician claims data should be taken into account when drawing conclusions concerning the external validation of *Google Trends* data with patient diagnoses as reference data. The diagnoses data from SHI-accredited physicians are not directly relevant for reimbursement between SHI-accredited physician and/or psychotherapeutic service providers and payers. Due to its primarily administrative nature, internal and external diagnosis validation is itself of central importance for any scientific secondary use of data within the framework of health services research (40-44). The restriction of defining ICD 10-coded diagnoses as the sole criterion for inclusion in the study limits the capture of cases and, in individual cases, may cause misclassification due to incorrect, unspecific or missing diagnoses. Furthermore, it should be noted that the SHI-accredited physician claims data does not contain any information on selective contract billing without the involvement of the SHI-accredited physicians' associations. This is especially relevant with respect to the selective contracts for GP-centered care in Baden-Wuerttemberg and Bavaria (45, 46). The potential of *Google Trends* data lies in its immediate and free availability. In addition, *Google Trends* queries can take into account differing regional development over time as well as regional language use where required. However, the methodology used by Google to provide the data is only very briefly outlined (1, 2) and it is unclear what influence the algorithms and any changes to them have. Events that temporarily lead to increased searches for certain keywords cannot be identified or calculated, but can lead to severe distortions. In addition, the provision of the service is entirely dependent on Google's business decisions. The data reported by the RKI are, in turn, subject to legal requirements and implementation by the reporting parties with regard to the type and depth of information and geographical validity.

Regarding the methodological approach of this study as a first exploration, it should be noted that only simple correlation coefficients were calculated, which have only limited explanatory power in the context of temporally autocorrelated data. For further analysis, methods should be used that can adequately represent the character and components of time series.

Conclusions

The findings of our study presented here clearly show that *Google Trends* data represents the seasonality emerging in SHI-accredited physician care in a comparable way, at least for Lyme disease and hay fever, although annual trends cannot be reliably mapped. *Google Trends* data therefore appears unsuitable for depicting long-term, cross-annual changes in disease risks in the population. However, in this exploratory analysis, it was shown that *Google Trends* data has the potential to generate early signals about changes in the start of the season or the period in which the number of new cases increases, particularly due to its very timely availability compared to other data sources. Studies using *Google Trends* as a data basis or as additional data points in the context of time series analyses should always aim to perform external validation in advance.

Acknowledgements

We would like to thank our colleagues Doreen Müller for additional comments and critical discussion, Christiane Gagelmann for editing as well as Camila Gonzatto da Silva and Melanie Nandzik for the layout on this work.

References

1. Google: Google Trends-Hilfe. Häufig gestellte Fragen zu Google Trends-Daten (Stand: 15.09.2022) https://web.archive.org/web/20220630014938/https://support.google.com/trends/answer/4365533?hl=de&ref_topic=6248052 (last accessed on 16.04.2025).

2. Simon R: Google News Lab - Medium. What is Google Trends data - and what does it mean? https://web.archive.org/web/20220911012356/https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8 (last accessed on 15.09.2022).

3. 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.

4. Mavragani A, Ochoa G: Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019; 5: e13439.

5. Eysenbach G: Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009; 11: e11.

6. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L: Detecting influenza epidemics using search engine query data. Nature 2009; 457: 1012-4.

7. Kandula S, Shaman J: Reappraising the utility of Google Flu Trends. PLoS Comput Biol 2019; 15: e1007258.

8. Eysenbach G: Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006: 244-8.

9. Ziehfreund S, Tizek L, Zink A: Websearch-Daten als Gesundheitsdaten? : Geografische Unterschiede, zeitliche Trends und Interessenschwerpunkte von Internetsuchmaschinenanfragen in Deutschland. Hautarzt 2022; 73: 53-60.

10. Stiftung Deutscher Polleninformationsdienst: Pollen- und Pilzsporenflug in Deutschland 2001-2021. Berlin2022.

11. Lake IR, Jones NR, Agnew M, et al.: Climate Change and Future Pollen Allergy in Europe. Environ Health Perspect 2017; 125: 385-91.

12. Semenza JC, Suk JE: Vector-borne diseases and climate change: a European perspective. FEMS Microbiol Lett 2018; 365.

13. Ogden NH, Ben Beard C, Ginsberg HS, Tsao JI: Possible Effects of Climate Change on Ixodid Ticks and the Pathogens They Transmit: Predictions and Observations. J Med Entomol 2021; 58: 1536-45.

14. Rizzoli A, Hauffe H, Carpi G, Vourc HG, Neteler M, Rosa R: Lyme borreliosis in Europe. Euro Surveill 2011; 16.

15. Akmatov MK, Holstiege J, Dammertz L, et al.: Epidemiology of Lyme borreliosis based on outpatient claims data of all people with statutory health insurance, Germany, 2019. Euro Surveill 2022; 27.

16. Akmatov MK, Holstiege J, Dammertz L, Kohring C, Heuer J, Bätzing J. Bundesweite und kleinräumige Kennzahlen zur Morbidität von Lyme-Borreliose in Deutschland anhand vertragsärztlicher Abrechnungsdaten, 2010 bis 2019. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 21/06. Berlin 2021. URL: https://doi.org/10.20364/VA-21.06

17. Holstiege J, Kohring C, Dammertz L, Heuer J, Akmatov MK, Bätzing J. Aktuelle Trends der Inzidenz diagnostizierter atopischer Erkrankungen bei Kindern und Jugendlichen in Deutschland. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 22/08. Berlin 2022. URL: https://doi.org/10.20364/VA-22.08

18. Binder K, Reich A, Sing A, et al.: LB-Meldedaten in Bayern (Kongressbeitrag). Gesundheitswesen 2015; 77: A34.

19. Mavragani A, Ochoa G, Tsagarakis KP: Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018; 20: e270.

20. Holstiege J, Dammertz L, Kohring C, Heuer J, Akmatov MK, Bätzing J. Bundesweite Inzidenztrends diagnostizierter Herzerkrankungen in den Jahren 2013 bis 2021. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 23/01. Berlin 2023. URL: https://doi.org/10.20364/VA-23.01.

21. Dammertz L, Holstiege J, Ng F, Kohring C, Heuer J, Akmatov MK, Bätzing J. Morbus Parkinson: Regionale Unterschiede der Diagnoseprävalenz und Komorbiditätsanalysen in der vertragsärztlichen Versorgung - Auswertung bundesweiter Abrechnungsdaten im Zeitraum 2010 bis 2019. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 22/01. Berlin 2022. URL: https://doi.org/10.20364/VA-22.01

22. Akmatov MK, Kohring C, Dammertz L, et al.: The Effect of the COVID-19 Pandemic on Outpatient Antibiotic Prescription Rates in Children and Adolescents-A Claims-Based Study in Germany. Antibiotics (Basel) 2022; 11.

23. Dammertz L, Kohring C, Heuer J, Akmatov MK, Bätzing J, Holstiege J. Inzidenztrends des diagnostizierten idiopathischen Parkinson-Syndroms in den Jahren 2013 bis 2019. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 22/06. Berlin 2022. URL: https://doi.org/10.20364/VA-22.06

24. Dammertz L, Schrag A, Bohlken J, et al.: Falling incidence of Parkinson's disease in Germany. Eur J Neurol 2023; 30: 3124-31.

25. Kohring C, Akmatov MK, Dammertz L, Heuer J, Batzing J, Holstiege J: Trends in incidence of atopic disorders in children and adolescents - Analysis of German claims data. World Allergy Organ J 2023; 16: 100797.

26. Tönnies T, Hoyer A, Brinks R, Kuss O, Hering R, Schulz M: Spatio-Temporal Trends in the Incidence of Type 2 Diabetes in Germany. Dtsch Arztebl Int 2023; 120: 173-9.

27. Rommel A, Deuschl G, Dodel R, et al.: Die Parkinsonkrankheit – Prävalenz, Trends und regionale Verteilung in Deutschland. Eine Auswertung auf Basis von GKV-Routinedaten. Journal of Health Monitoring 2025; 10: 1–9.

28. Rommel A, Gaertner B, Neuhauser H, et al.: Demenzerkrankungen – Prävalenz, Trends und regionale Verteilung in Deutschland. Eine Auswertung auf Basis von GKV-Routinedaten. Journal of Health Monitoring 2025; 10: 1–10.

29. Robert Koch-Institut: RKI-Ratgeber Lyme-Borreliose. Epidemiologisches Bulletin 2019: 137-43.

30. Robert Koch-Institut (RKI): SurvStat@RKI 2.0, Abfragedatum: 20.09.2022.

31. Verband der Ersatzkassen e. V. (vdek): vdek-Basisdaten des Gesundheitswesens in Deutschland 2022. Berlin 2022.

32. Brestrich G, Hagemann C, Diesing J, et al.: Incidence of Lyme Borreliosis in Germany: A retrospective observational healthcare claims study. Ticks Tick Borne Dis 2024; 15: 102326.

33. Skufca J, Tran TMP, Brestrich G, et al.: Incidence of Lyme Borreliosis in Germany: Exploring Observed Trends Over Time Using Public Surveillance Data, 2016-2020. Vector Borne Zoonotic Dis 2023; 23: 237-46.

34. Kapitány-Fövény M, Ferenci T, Sulyok Z, et al.: Can Google Trends data improve forecasting of Lyme disease incidence? Zoonoses Public Health 2019; 66: 101-7.

35. Scheerer C, Ruth M, Tizek L, Koberle M, Biedermann T, Zink A: Googling for Ticks and Borreliosis in Germany: Nationwide Google Search Analysis From 2015 to 2018. J Med Internet Res 2020; 22: e18581.

36. Bousquet J, Agache I, Anto JM, et al.: Google Trends terms reporting rhinitis and related topics differ in European countries. Allergy 2017; 72: 1261-6.

37. Bousquet J, Onorato GL, Oliver G, et al.: Google Trends and pollen concentrations in allergy and airway diseases in France. Allergy 2019; 74: 1910-9.

38. Zentralinstitut für die kassenärztliche Versorgung (Zi): Grafik des Monats November 2022. Anrufer:innen bei der 116117 mit Symptomen akuter Atemwegserkrankungen. https://www.zi.de/das-zi/medien/grafik-des-monats/detailansicht/november-2022 (last accessed on 16 April 2025).

39. Zoch-Lesniak B, Kroll LE, Czihal T, von Stillfried D. Die Einführung eines bundesweit einheitlichen, strukturierten Ersteinschätzungsverfahrens akuter Beschwerden – erste Ergebnisse für 2020 und 2021. 21. Deutscher Kongress für Versorgungsforschung, 05.-07.10.2022, Potsdam. https://www.egms.de/static/en/meetings/dkvf2022/22dkvf429.shtml (last accessed on 16 April 2025).

40. Neubauer S, Zeidler J, Lange A, Graf von der Schulenburg J-M: Grundlagen und Methoden von GKV-Routinedatenstudien. Diskussionspapier Nr. 534. Hannover: Leibniz Universität Hannover, Center for Health Economics Research Hannover (CHERH).

41. Schubert I, Ihle P, Köster I: Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. Gesundheitswesen 2010; 72: 316-22.

42. Horenkamp-Sonntag D, Linder R, Wenzel F, Gerste B, Ihle P: Prüfung der Datenqualität und Validität von GKV-Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D, (eds.): Routinedaten im Gesundheitswesen Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Verlag Hans Huber 2014/15; p. 314-30.

43. Swart E, Stallmann C, Powietzka J, March S: Datenlinkage von Primär- und Sekundärdaten : Ein Zugewinn auch für die kleinräumige Versorgungsforschung in Deutschland? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014; 57: 180-7.

44. Erler A, Beyer M, Muth C, Gerlach FM, Brennecke R: Garbage in - Garbage out? Validität von Abrechnungsdiagnosen in hausärztlichen Praxen. Gesundheitswesen 2009; 71: 823-31.

45. Lübeck R, Beyer M, Gerlach F: Rationale und Stand der hausarztzentrierten Versorgung in Deutschland. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015; 58: 360–6.

46. Bundesregierung: Deutscher Bundestag, 19. Wahlperiode, Drucksache 19/9503, 16.04.2019. Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Dr. Andrew Ullmann, Michael Theurer, Grigorios Aggelidis, weiterer Abgeordneter und der Fraktion der FDP – Drucksache 19/8823 – Qualität und Wirtschaftlichkeit der hausarztzentrierten Versorgung. Berlin.