Social Science Research 126 (2025) 103099

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and recommendations

Johanna Hölzl [*], Florian Keusch, Christoph Sajons

University of Mannheim, Germany

ARTICLE INFO

Keywords:
Google Trends
Systematic literature review
Validity
Reliability
Generalizability

ABSTRACT

Researchers increasingly use aggregated search data from Google Trends to study a wide range of phenomena. Although this new data source possesses some important practical and methodological benefits, it also carries substantial challenges with respect to internal validity, reliability, and generalizability. In this paper, we describe and assess the existing applied research with Google Trends data in the social sciences. We conduct a systematic literature review of 360 studies using Google Trends data to (1) illustrate habits and trends and (2) examine whether and how researchers take the identified challenges into account. The results show that the large majority of the literature fails to test the internal validity of their Google Trends measure, does not consider whether their data are reliable across samples, and does not discuss the generalizability of their results. We conclude by stating practical recommendations that will help researchers to address these issues and properly work with Google Trends data.

While surveys are still the most commonly used data collection method in the social sciences (Sturgis and Luff, 2021), declining response rates (Schober et al., 2016) and concerns about erroneous responses (Jun et al., 2013) have led researchers to experiment with novel digital data about human thoughts and behavior (Salganik, 2017). One of these novel data sources is Google Trends. For any term of interest, Google Trends provides aggregated information on the relative frequency of searches across regions and over time based on a randomly drawn sample of all queries conducted on Google (Google, 2024; Google News Initiative, 2024b). The data are available at much more fine-grained time and geographic units than survey data (Ettredge et al., 2005; Jun et al., 2017). They are free of charge and can be retrieved immediately by everyone. Additionally, information from Google Trends is less prone to recall errors and social desirability, which can substantially bias traditional survey data, especially in the case of sensitive topics (DiGrazia, 2015; Stephens-Davidowitz, 2014, 2017). On the downside, some researchers point to issues of transparency and reliability of the sampling procedure and potential limitations in the generalizability of the obtained information (e.g., Behnen et al., 2020). It is also often unclear how validly the selected search terms reflect the constructs of interest (e.g., Mellon, 2013, 2014). Failing to account for these issues casts doubt on the accuracy and robustness of the results obtained from Google Trends.

Despite these severe concerns, little systematic research has been conducted on how researchers use Google Trends data in the social sciences, whether they account for the accompanying methodological challenges, and how they do so. Jun et al. (2017), Mavragani et al. (2018), and Nuti et al. (2014) take valuable steps in this direction though. Jun et al. (2017) provide a general overview of the topics examined with Google Trends data. Mavragani et al. (2018) and Nuti et al. (2014) review the use of Google Trends data in health studies. Nuti et al. (2014) additionally discuss the replicability of findings derived from Google Trends data in health research and create a checklist on what researchers should include and report so that their results can be replicated.

* Corresponding author.
E-mail address: johanna.hoelzl@uni-mannheim.de (J. Hölzl).

https://doi.org/10.1016/j.ssresearch.2024.103099
Received 8 January 2024; Received in revised form 18 October 2024; Accepted 20 October 2024
Available online 31 December 2024
0049-089X/

In this paper, we combine these approaches and extend them in several ways. In particular, we examine the following research questions:

RQ1: What are the main methodological challenges researchers face when using Google Trends data?
RQ2: How are Google Trends data used in the social sciences?
RQ3: How does the existing social science literature treat the methodological challenges of Google Trends data?
RQ4: What solutions can be implemented to address the limitations of Google Trends data?

In answering these questions, we contribute to the literature in four ways: (1) We describe in detail how Google Trends works, how its data can be retrieved, what benefits the data provide, and what problems emerge when doing research with it. (2) We conduct a systematic literature review of 360 English-language studies published until 2021 that use Google Trends data to answer social science research questions. The review documents the growing use of Internet search data in the social sciences and provides an overview of the current state of research. (3) Our review comprehensively and systematically examines how researchers (fail to) deal with the methodological challenges regarding keyword selection, construct validity, reliability, and generalizability. The results show that the large majority of Google Trends studies suffer from critical problems with respect to transparency and methodological rigor. Finally, we (4) use the findings of the systematic review and previous methodological work to develop recommendations on the proper use of Google Trends data to help researchers draw more reliable and valid conclusions in future applications. In sum, our paper should sensitize researchers to the challenges of Google Trends data and provide them with the necessary tools to deal with these challenges.

In the following paper, we start by outlining the main benefits and pitfalls of Google Trends data for the social sciences. We then describe the methodological approach for our systematic literature review and discuss its findings. We derive our recommendations from these results and present a step-by-step checklist for researchers who plan to work with Google Trends data. Eventually, we summarize our findings and point out areas where future research on the data quality of Google Trends is needed.

1. Description of Google Trends and its features

Google Trends provides aggregated data on the relative search frequencies of selected keywords for regions worldwide, going back to 2004. Researchers can retrieve search frequencies for up to five keywords or topics simultaneously. While selecting individual keywords only displays data for searches on this exact term, data for topics also encompass acronyms, misspellings, translations of the term in other languages, and other conceptually related terms (Google News Initiative, 2024b; Trends Help, 2024a). Data retrieved for individual keywords and for topics will therefore differ and may even produce different results and conclusions (see Online Appendix A for details). If enough people searched for a term in the selected region and time period, the search data can be grouped by location at the country, regional, and city level and by time at minute, hour, day, week, and monthly intervals (Google News Initiative, 2024a,b). Researchers can additionally see ‘related searches’ for their term of interest showing what other terms users frequently looked for in the same search session and which search terms had the highest growth rate in the respective time period (Trends Help, 2024b, 2024c). For more detailed information on the different functions of Google Trends see Online Appendix A.

According to the documentation provided by Google, the search data reflect the popularity of a keyword as the “search interest in a particular topic, in a particular place, and at a particular time” (Google News Initiative, 2024b). When retrieving the relative search frequencies for the keywords “Obama” as depicted in Fig. 1, researchers obtain data on the worldwide interest of Google users in the former U.S. president from 2004 until 2024.

Individuals can access and download the data via Google’s official website (google.com/trends). For automated access to the website, researchers can use a statistical software package such as *pytrends* for Python (DeWilde and Hogue, 2022) or *gtrendsR* for R (Massicotte and Eddelbuettel, 2022).

Importantly, Google provides the aggregated data not as absolute but relative search volume in the form of a normalized index from 0 to 100, the *Relative Search Index (RSI)*.[1] According to the company, Google provides the normalized index to allow for the comparison of a topic’s popularity over time without being biased by the overall massive growth of search queries since 2004 (Google News Initiative, 2024b; Rogers, 2016). Each data point of the index is adjusted to the absolute number of searches of a given day in the region of interest to account for seasonality in the number of queries. The maximum value of 100 gets assigned to the data point with the highest search interest for the selected time period, location, and keywords (when comparing multiple keywords simultaneously). Google then divides the search frequencies of all other data points by this maximum and multiplies the result by 100. As a consequence, each data point reflects the percentage of the search intensity relative to the maximum (a) within a specific time period (time series data, see Fig. 1 for an example), (b) for specific subregions (regional comparison data), or (c) of one term in comparison to other terms (Google News Initiative, 2024b; Rogers, 2016).

Google reports data points that have search volumes below a specific, unknown threshold as values of 0. According to Google, the company follows this approach to ensure the anonymity of their users (Google News Initiative, 2024b; Rogers, 2016). Thus, values of 0 can either mean that nobody searched for a specific term or that the search volume was too low to be depicted. Terms with overall low search volume therefore produce time series or geographic maps with many values of 0.

[1] Google does not provide a specific name for its index. Authors also use other terms to describe the index provided by Google Trends, including *(Google) Search Volume Index* (e.g., Jun et al., 2017; Lorenz et al., 2022), *Google Search Index* (Caporin and Poli, 2017), or *Google Trends Index* (e.g., Böhme et al., 2020; Golenvaux et al., 2020).

Fig. 1. Example trend for the intensity of worldwide Google searches for the term "Obama" from 2004 to today. Note: Screenshot of the output provided by the Google Trends tool on the monthly search intensity (y-axis) of the term "Obama" searched worldwide in all categories from January 1, 2004, to October 11, 2024 (x-axis). Text displayed when hovering over the two "Notes": "An improvement to our data collection system was applied from January 1, 2016 [January 1, 2022]." Source: Retrieved from the Trends tool provided by Google (Google Trends, 2024) accessed from Germany on October 11, 2024.

1.1. Benefits of Google Trends data for social science research

1.1.1. Practical benefits: Easy access, wide availability, long reach, and high frequency

Using Google Trends data for research provides several practical advantages. With a share of over 90 percent of Internet searches worldwide, Google holds a dominant position in the search engine market since 2009 (Statcounter, 2021; Timoneda and Wibbels, 2022). Compared to most surveys, Google Trends data are freely and immediately available at much finer-grained geographic levels and time units. Additionally, the search engine data can provide information on a wide range of topics (Ettredge et al., 2005; Jun et al., 2017). Researchers can thus use Google Trends to obtain insights on the societal salience of a wider array of issues in a more timely manner and for smaller subregions than surveys (e.g., Mellon, 2013, 2014).

Böhme et al. (2020) present an example of taking advantage of the immediate availability of Google Trends data to predict international migration flows. The authors forecast offline *behaviors* with trends for searches that they think represent the corresponding *intent to behave*. These trends include queries for "visa" or "immigration" as an indicator for migration. Other studies make use of the wide range of topics that can be examined with Google Trends. Gummer and Oehrlein (2022) monitor the *interest* in the survey FReDA by checking the search trends for the survey name after they sent out more than 50,000 letters inviting potential participants. Fang et al. (2021) also exploit Google Trends data for survey methodology by examining the time-varying salience for various events on survey non-response. Olzak (2021) augments her theoretical argument that protests on police brutality generate widespread public and media *attention* by the fact that the search volume on police violence rapidly increased following such demonstrations. We would not have been able to gain any of these insights without Google Trends providing prompt information on such a wide variety of topics.

1.1.2. Methodological benefits: Great honesty and low cognitive demand

Apart from these practical benefits, Google Trends comes with additional methodological advantages as the data are collected in a passive, non-reactive way. That is, they are based on the direct observation of individuals' search behavior. In consequence, Internet search data do not suffer from cognitive biases that may arise during the response process in surveys as they do not rely on individuals' comprehension of questions, accurate memories, or general willingness to provide a thoughtful answer (Salganik, 2017; Tourangeau et al., 2000).

Most importantly, however, Internet search data also do not suffer from social desirability bias (DiGrazia, 2015; Jun et al., 2013). This type of bias arises when survey respondents are hesitant to disclose their true beliefs or behaviors to researchers, particularly in face-to-face interviews and regarding sensitive topics. Instead, respondents provide socially acceptable or pleasing answers to present themselves in a more favorable light. In consequence, surveys frequently underestimate the prevalence of attitudes and behaviors that are considered socially undesirable (DiGrazia, 2015; Jun et al., 2013; Krumpal, 2013).

By contrast, people are assumed to be mostly alone and feel anonymous and unobserved when conducting a query on Google. They can therefore express taboo thoughts and honest feelings on socially sensitive topics more easily (DiGrazia, 2015; Stephens-Davidowitz, 2014, 2017). Users even have an incentive to tell Google the truth to retrieve the desired information from the search engine. As a result, we can observe frequent queries for sensitive topics related to pornography and health-related information (Stephens-Davidowitz, 2014). Making use of this unfiltered byproduct of human behavior is particularly interesting for attitudes research, where topics such as xenophobia, sexism, or homophobia are prone to social desirability bias and social censoring.

Of course, not all people conducting a particular query share the same specific mindset. However, individuals holding certain attitudes or planning a certain behavior are expected to be more likely to google for specific terms. By this means, the observed variation in aggregated search frequencies can hold a "high signal-to-noise ratio" (Stephens-Davidowitz, 2014:27) on the area level even if some individuals google for reasons unrelated to the construct of interest.[2]

The probably best-known study making use of Google Trends as honest data to measure *attitudes* is Stephens-Davidowitz (2014). In

[2] Stephens-Davidowitz (2014), for instance, finds that an area's search interest in the term 'God' highly correlates with the level of religiosity of that area. This correlation holds even though a large number of queries were made for the combined search for the video game 'God of War' (Stephens-Davidowitz, 2014).

this paper, the author uses search frequencies for the *N*-word, a common racial slur against Blacks in the U.S., to examine the impact of an area's racial animus on Obama's state-level vote shares in the 2008 presidential election. The crucial assumption here is that regions with more individuals holding racist opinions have higher search frequencies for the term. Following this logic, the author assumes that an area's aggregated search volume for the racial slur reflects that area's level of racism. While data from the General Social Survey show only a small racism effect on Obama's vote share, the Google Trends measure reveals a much larger impact of an area's racism level. Stephens-Davidowitz (2014) attributes this finding to the assumption that respondents identifying as Democrats do not disclose their Anti-African-American prejudices in a survey. In contrast, racist Democrats show their true beliefs on Google in the same way as racist Republicans do. Several studies have since used and validated Stephens-Davidowitz (2014) honest Google Trends measure of racism to examine the impact of racism on racial economic disparities (DiBartolomeo et al., 2021) and racial physical and mental health disparities (Chae et al., 2015; Isoya and Yamada, 2021; McKetta et al., 2017). Others made use of the indicator to investigate the effect of economic conditions on racial animus (Anderson et al., 2020; Connor et al., 2019).

In an even more recent example, Liu et al. (2023) demonstrate Google Trends' benefit of providing honest data in a very different research field. The authors estimate the incidence of traditionally underreported crimes including rape and theft. Their findings suggest that Internet search data depict these crime incidences more accurately compared to traditional data sources that are prone to considerable underestimation such as official records and surveys. Finally, researchers such as Brodeur et al. (2021) combine the timeliness and honesty benefits of Google Trends and use the data to assess levels of well-being within the population before and after COVID-19 lockdowns. The researchers assume that the promptly available search trends for terms such as loneliness, worry, or stress represent how well the public is feeling at a certain point in time.

1.2. Potential challenges of Google Trends data for social science research

Google Trends not only promises large benefits for social science research, but the data also come with certain challenges with respect to both internal and external validity (see Fig. 2). In this subsection, we discuss these challenges in more detail.

1.2.1. Challenge 1: Questionable internal validity

The use of Google Trends as an indicator of *public interest, attitudes,* and *behavior* raises concerns about internal validity, specifically construct validity. Does a selected keyword measure what it is supposed to measure? Or put differently: What does the search frequency for a specific term really tell us? It seems intuitive to see googling a term as a statement of *interest* as individuals searching information on a topic clearly want to know more about it. And if many individuals google the same term at the same time, the topic at first glance appears to be salient in society at that time. Mellon (2013, 2014), for instance, discovers that Google Trends can be used to measure the salience of a number of issues such as immigration, terrorism, and macroeconomics using carefully validated keywords. However, he also finds that many search terms that initially seemed plausible to measure public interest in a specific topic did not prove to be a valid measurement for salience of that issue in the end, as words often have multiple meanings and different terms can be used to search for the same topic. For instance, individuals wanting to know more about the COVID-19 pandemic might google either "coronavirus" or "covid" or "pandemic". Using just one of these terms to measure the public's interest in this issue over time may thus produce misleading conclusions if there is a shift in the way people talk about and search the phenomenon.

Establishing construct validity becomes even more challenging when researchers use Google Trends data to measure *attitudes* and *behavior* instead of mere *interest*. In this case, two major issues arise. First, researchers cannot determine the intent or motivation of the individuals behind the queries from Google Trends data alone. For instance, not everyone googling a racist slur is necessarily a racist. Maybe the person just wants to know what a word means or wants to find information on the history of the term. Individuals can google terms that are completely unrelated to their attitudes and intended behavior. Different groups of users may even search the same term for different and potentially opposing reasons (Fenga, 2020; Mukherjee and Jansen, 2017; Shahzad et al., 2017).

Second, Internet search behavior can also be impacted by various exogenous factors such as specific events, media coverage, or trends on social media (Fenga, 2020; Mukherjee and Jansen, 2017; Shahzad et al., 2017). Google Trends data therefore frequently show large fluctuations in search volume, while attitudes remain relatively stable over time (Kiley and Vaisey, 2020).

In light of these two problems, selecting the right search terms that validly measure the construct of interest poses a major challenge. In particular, for researchers to be able to measure attitudes or behavior with search data the following condition must hold: the variation in search volume over time or across regions must stem from individuals conducting a specific query because they themselves hold the attitude of interest or intend to behave in that way. If researchers simply select keywords ad-hoc, test numerous (combinations of) keywords, and decide on the one that best predicts their attitude or (offline) behavior of interest, there is a high probability of cherry-picking and spurious correlations (Memon et al., 2020; Stephens-Davidowitz and Varian, 2015).

1.2.2. Challenge 2: Intransparency and sample instability jeopardizing reliability

Google states that every time series retrieved from the tool represents a "random unbiased sample of Google searches" (Google News Initiative, 2024b). The company, however, does not disclose any information on the size of the sample, the exact sampling procedure, or the reference value used to scale the relative search index (RSI).[3] This intransparency leads to three interlinked major complications seriously questioning the reliability of the data source.

[3] We inquired at Google Trends via email about the details of the sampling procedure but did not receive a response.

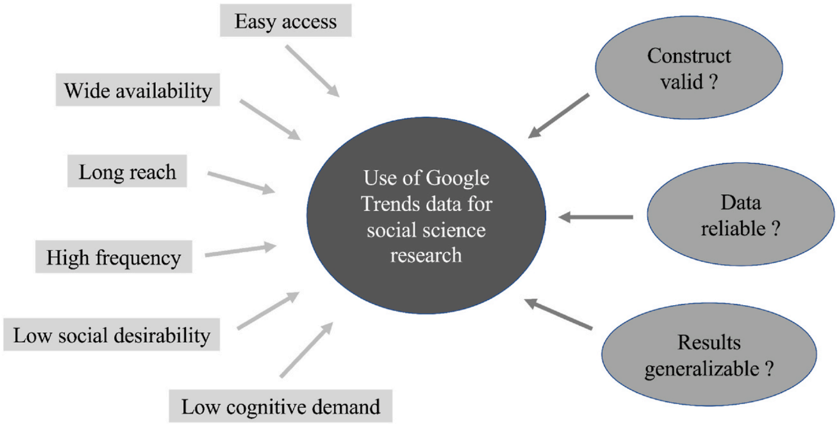

Fig. 2. Benefits and challenges of using Google Trends data for social science research.

(1) *Since the RSI is based on a sample of all Google searches, sampling variance can produce unreliable estimates from Google Trends.* We have to interpret the values of the RSI as a mere estimation of the true search volume and need to take its variation into account for statistical inference. However, Google Trends does not provide any measures of uncertainty such as standard errors or confidence intervals. Further complicating matters, the exact sampling procedure and underlying sampling algorithm are not transparent and might have changed (several times) over the years. Retrieving Google Trends data for the exact same parameters but on different days can therefore produce samples with widely different RSI values that may even trend in opposite directions (Behnen et al., 2020; Eichenauer et al., 2022; Franzén, 2023; Gummer and Oehrlein, 2022; Mavragani and Ochoa, 2019).[4]

The severeness of the sample instability seems to depend on the search volume for a specific term. Samples tend to produce consistent results for terms with a large search volume covering a time span of at least five years. For scarcely searched terms, however, large inconsistencies between samples retrieved at different points in time occur more frequently and impact the results more strongly. The values of the obtained RSI and conclusions derived from these samples may thus completely depend on the particular sample drawn on a certain day (Behnen et al., 2020). The same issue emerges for regions with low Internet penetration rates (Mavragani and Ochoa, 2019) and for searches from small countries (Eichenauer et al., 2022).

If researchers cannot obtain consistent values for many countries and subregions in the world or for many terms with only low search volume, some of the great benefits of Google Trends become obsolete. In that case, researchers cannot use the data at fine-grained regional and temporal units, for regions worldwide, and for such a wide range of topics. Sample inconsistencies thus pose a major obstacle for reliably using Google Trends data in research.

(2) *Changes in Google's internal algorithms represent further risks to the stability and comparability of the retrieved trends over time.* In 2008, Google introduced a feature that automatically recommends how a search request could be completed (Sullivan, 2008). These suggestions are partly based on what other users have already googled, leading to possible search externalities and making individual queries dependent on others' searches (Lazer et al., 2014). After the introduction of this feature, recommended queries potentially experienced a large surge in searches as people might have simply followed the suggestions. The auto-complete function therefore makes the comparison of search trends before and after 2008 difficult. In 2011, the company also substantially improved their geo-location algorithm, meaning that queries are mapped to a specific location more accurately (MacInnis and Gordon, 2015).

Google Trends also points to additional changes in the Google search algorithm in the time-series graphs it produces on their website (see Fig. 1). The two most recent changes concerned the improvement of their data collection system in 2016 and 2022 (Google, 2024). However, Google does not provide any further information on the specific details of these changes, and it is thus unclear how they affect any comparison over time or region. Importantly, we also do not know whether changes in the data retrieval algorithm at a certain point in time only influence the Google Trends numbers for searches conducted afterwards or for the whole time series. The latter would additionally hamper the replicability of results derived from Internet search data in the past. It is similarly intransparent how and based on which terms Google Trends calculates the RSI values when selecting a topic instead of a keyword and whether the calculation and search terms included have changed over time. Changes in the calculation of topics' RSI values would

[4] Researchers might also obtain different RSI values when downloading the data from different regions of the world or when using a VPN-connection. We conducted a small test using the same parameters to access Google Trends data on the popularity of the search term "Barack Obama" from IP addresses in the U.S., Germany, and Austria using a VPN. We found that the results differ especially for searches on more recent data ("Past day" vs. "2004-present") and when restricting the search to smaller markets (Austria vs. worldwide). We also do not know whether the samples include searches conducted in incognito mode.

further impede the replicability of Google Trends research when examining topic search trends instead of individual keywords.[5]

(3) *The intransparent scaling of Google Trends data leads to difficulties for comparisons across both time and region.* Google releases relative search frequencies scaled in reference to the maximum, namely the data point with the largest search volume for the selected term in the selected time range and region. Depending on the reference frame, the data are scaled based on the region or the time point with the highest number of searches. Researchers, however, cannot retrieve both regional comparison and time series data with the same reference maximum simultaneously. Google Trends also does not provide precise information on how they calculate the RSI and how the different trends relate to one another. If researchers thus wish to make a comparison across both regions and time simultaneously, they need to find a way to scale the data themselves by adjusting the regional and time reference frame relative to each other. A similar issue arises when researchers want to retrieve data for more than five keywords simultaneously or for small time units over a long time span. For example, Google restricts retrieval of daily time series to a maximum of about nine months.[6] Researchers thus need to adequately combine several monthly data sets if they are interested in a longer observation window of daily data (see, for example, Caporin and Poli, 2017).

1.2.3. Challenge 3: Unknown generalizability

Even if search frequencies prove to be a valid measure for a certain construct (*salience*, *attitude*, or *behavior*) and the retrieved samples are consistent, it is still unclear whether researchers can generalize the results beyond the population of Google users. Despite its extremely large user base, Google Trends data may, for instance, not generalize to the whole population in a specific region. Not everybody has access to or uses the Internet, and not every Internet user relies on Google (Lorenz et al., 2022; Schober et al., 2016). Google users thus represent a selective population and the frequency of searching information via Google depends on sociodemographic factors such as age and economic status (Mellon, 2013). As we do not know anything about the individuals conducting the queries, directly controlling for this selection is impossible. Likewise, we cannot obtain precise percentage estimates of how many individuals hold a certain view, because Google only provides aggregate and relative search frequencies instead of absolute numbers (DiGrazia, 2015). These issues of selection bias and relative aggregate data are even more problematic in regions with low Internet penetration rates (Koehler-Derrick, 2013).

Making matters worse, the composition of Google's user base has not remained constant over time. Instead, Google users represent "an always changing, self-selecting population of Internet users" (Fenga, 2020:282) whose exact composition is unknown and unstable. Without any information on the individuals conducting the searches and their unknown probability to be included in the sample, it is therefore impossible to adjust for the lack of population coverage and the selection bias of the data (Lorenz et al., 2022; Mavragani and Ochoa, 2019; Schober et al., 2016).

In conclusion, Google Trends have the potential to reflect people's interest in an issue and even certain attitudes and behaviors (e.g., DiGrazia, 2015; Schober et al., 2016; Stephens-Davidowitz, 2014). How well Internet search data reflect what a population thinks or does depends on the internal validity of the selected search terms, the reliability of the measure, and the generalizability of the results beyond Google users in each particular case.

2. Systematic literature review: Data collection

In the next step, we turn to the existing literature to get an overview of how researchers in the social sciences use Google Trends data in their studies and whether and how they deal with the above-mentioned challenges. To this end, we conduct a systematic literature review on the use of Google Trends closely following the PRISMA guidelines and recommendations for systematic reviews laid out by Page et al. (2021a).[7]

To get a comprehensive collection of existing studies using Google Trends data in the social sciences, we applied two search strategies. First, we conducted a systematic search for all English-language papers that contain the terms *"Google Trends"* or *"Google search"* in the title, abstract, or keywords,[8] on the databases *Web of Science, ProQuest, ScienceDirect, ArXiv*, and *OSF*. Second, we applied a snowball approach to check for completeness. Specifically, we went through the reference list of the papers that we found via the automated search and cross-checked on Google Scholar for related papers. With this approach, we retrieved other relevant papers that were not already included in our sample but fulfill the same inclusion criteria.

The data collection process took place in the first half of December 2021. Our corpus thus consists of papers that were published

[5] In an exploratory test conducted thanks to a reviewer suggestion (see Figure A.1 in Online Appendix A for details), we noticed another issue with the use of topics instead of keywords. When comparing the trends for the same term (i.e., using Google Trends' "+ Add Comparison") - once included as a topic and once as a single keyword - the topic's RSI values were in parts lower than the RSI values for the single keyword. This difference in the RSI values should not be possible as the RSI for topics should always entail more searches, i.e., searches for the term itself, misspellings, and translations, than the RSI for single keywords that should only entail the searches for this specific keyword.

[6] Similarly, Google Trends currently limits access to minute-level data to the past 4 hours, hourly data up to the past seven days, daily data up to about nine months, and weekly data up to five years.

[7] We rely on the PRISMA guidelines as they facilitate the comprehensive and transparent reporting of our reasoning behind the review, our methodological approach and decisions, and our results, so that other researchers can easily understand and replicate our work (Page et al., 2021a, b).

[8] This restriction of our analytical sample increases the probability that we capture papers that use Google Trends as a data source in their analysis. Please note, however, that other papers in the social sciences also use Google Trends as an addition to their main analysis without mentioning the data source in their abstract (see, e.g., Olzak 2021; Vasi et al., 2015).

online in peer-reviewed (e.g., articles in peer-reviewed journals) and non-peer-reviewed formats (e.g., preprints or working papers) from the introduction of the Google Trends tool in 2006 until mid-December 2021. Including both peer-reviewed and non-peer-reviewed papers in our review has several advantages. It enables us to (1) present the broadest possible picture of the quality of research using Google Trends data, (2) make an effort to avoid publication bias, and (3) incorporate more recent studies for completeness that have not (yet) been published in peer-reviewed journals at the time of our search.

Fig. 3 illustrates the detailed steps of our data collection process. In the screening phase, we applied the following inclusion criteria: (1) English-language studies that (2) use Google Trends as a data source in their empirical analysis to (3) address a core social science topic. More specifically, we removed all papers concerned with issues such as the prediction of stock market returns, the spread of diseases, and the analysis of geolocation images. The final corpus consists of 344 papers containing a total of 360 individual studies as some researchers conducted multiple distinct studies using Google Trends within the same paper (for details of our data collection and screening process, see Online Appendix B[9]).

We manually coded these studies with respect to the following categories[10]: characteristics of the publication outlet, field of study, constructs measured with Google Trends data, search strategy, combination with additional data sources, the type of statistical analyses, keyword selection, checks for construct validity and sample stability, and considerations of generalizability (see our Online Supplemental Material for the complete coding manual, the coded data set of all studies, and further replication material of our analyses). With these variables, we can describe the development and current state of the literature and contrast the methodological approaches used in different subgroups of studies, such as those using Google Trends data to measure salience, attitudes, and behaviors. For each paper, we looked at the main text as well as its footnotes, endnotes, and supplementary material. To test interrater reliability, we double coded a random set of 20 papers (21 studies). We calculated the resulting interrater reliability using Krippendorff's alpha. Across all variables, we achieved a value of 0.82 which is above the required threshold of $\alpha = 0.67$ for satisfactory agreement (Krippendorff, 2004). Only the category assessing whether Google Trends was used to compare search frequencies across time or region ($\alpha = 0.60$) and the category classifying the study's statistical approach as forecasting or modeling ($\alpha = 0.36$) were below this threshold. Due to its particularly low value, we excluded the latter category from our analyses.[11]

Lastly, we added information at the journal-level for those papers in our database that are published in peer-reviewed publication outlets. Specifically, we included information on the impact factor of the journal in the respective publication year and the journal's field of study based on the classification on Clarivate Analytics' (2020) Web of Science website.[12] For the non-peer-reviewed papers and any missing values for the peer-reviewed papers, we manually coded the fields of study based on the same categories and subcategories.

3. Results

By now, we discussed the main methodological challenges researchers face when using Google Trends data (RQ1). In this section, we first show descriptive statistics on how Google Trends data are used in the social sciences (RQ2). Next, we report how the existing literature treats the methodological challenges of Google Trends data (RQ3). Based on these results, we provide recommendations and best-practice examples based on our analyses in the next section to help researchers who want to work with Google Trends data reduce their limitations (RQ4).

3.1. *How are Google Trends data used in the social sciences?*

3.1.1. *Growth of Google Trends usage in the social sciences*

Fig. 4 documents the use of Google Trends as a data source for research in the social sciences since the introduction of the tool. Despite its introduction already back in 2006, the first papers using Google Trends in the social sciences were only published in 2010 and few authors worked with the aggregate search data in the following years. After 2014, the tool became increasingly popular with a steady increase from 17 papers published in 2014 to 40 papers published in 2019. In the last two years of our observation window, 2020 and 2021, Google Trends experienced a sharp increase in applications with 100 papers being published in 2021 alone.[13] This rapid rise in popularity points to the growing importance of Internet search data in the social sciences, especially during the pandemic. The increased usage also underlines the need for a systematic review of what has been done so far and how well the existing research accounts for the challenges that come along with using Google Trends data.

[9] Online Appendix B also provides the complete list of all papers included in the review.

[10] We partly based our coding scheme on the categories used by Mavragani et al. (2018). The majority of the coding was done by the first author, supported by two research assistants who obtained specific training for this task. When a paper consisted of multiple distinct studies using Google Trends, we coded each study separately.

[11] The particularly low value of this category results from the fact that Krippendorff's alpha is very sensitive to disagreement in minority cases where the large majority of observations take on the same value. In this case, the large majority of double coded studies did not use Google Trends for forecasting and the coding discrepancy resulted from disagreement whether a single study did do so or not.

[12] To better represent the range of journals, we added the categories Forecasting, Multidisciplinary Sciences, and Political Science to the list of fields of study provided by Clarivate Analytics' (2020) Web of Science website.

[13] Five papers in our sample were forthcoming at the time of our literature collection efforts in December 2021 and have been officially published and assigned a later publication year in the meantime. For the sake of consistency and comparability, we assigned these papers to 2021.

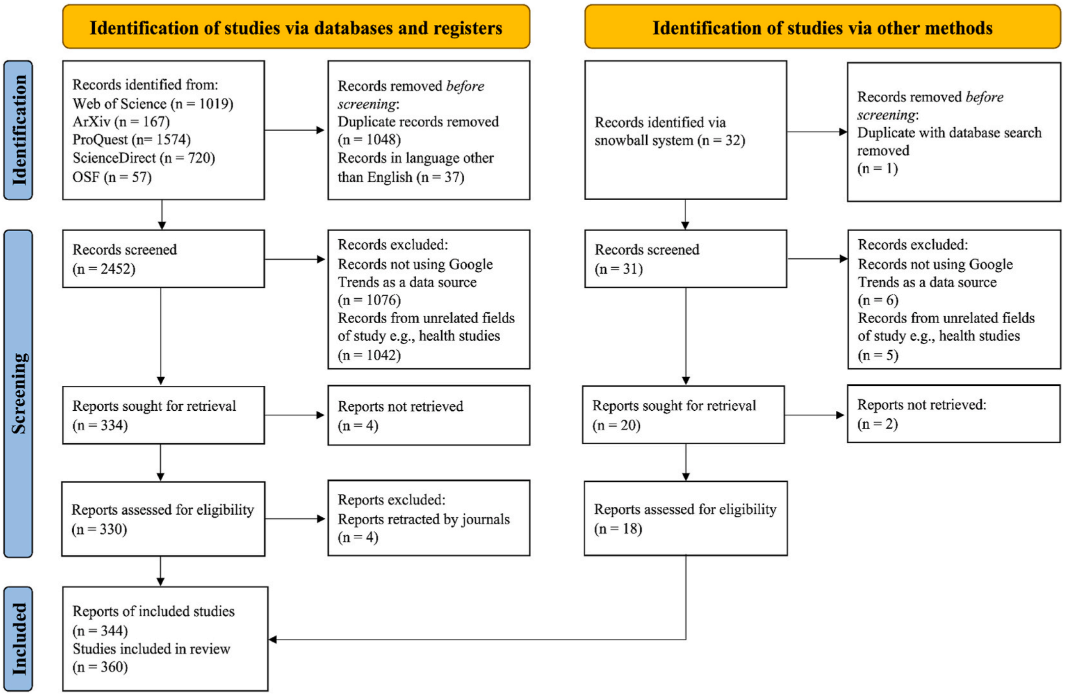

Fig. 3. PRISMA 2020 flow diagram for systematic reviews which include searches of databases and other sources. Source: Page et al. (2021a). For more information, visit: http://www.prisma-statement.org. Note: Individual reports (papers) can contain multiple distinct studies using Google Trends data.

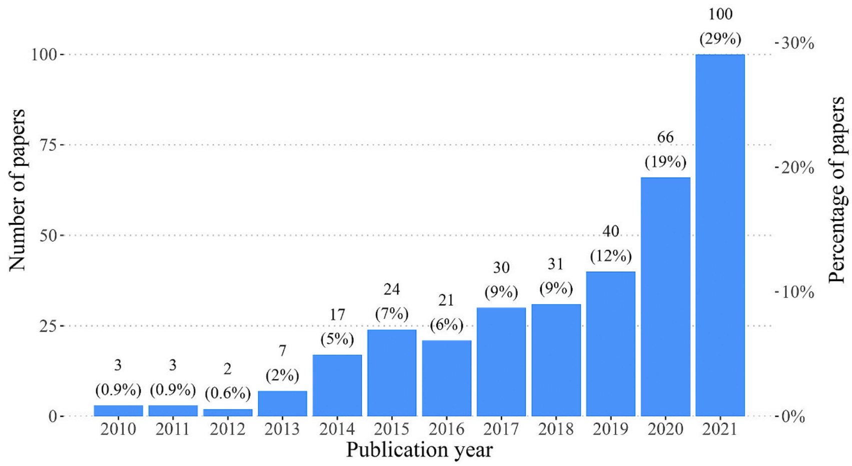

Fig. 4. Development of the number of papers using Google Trends data to examine a social science topic. Note: The bars indicate the distribution of English-language papers using Google Trends data for applied research in the social sciences (y-axis) from 2010 to 2021 (x-axis). Given that we searched the papers in late 2021, we assigned papers with an already scheduled publication in 2022 to 2021. N = 344.

3.1.2. *Google Trends research across publication outlets and fields of studies*

Table 1 describes the 344 papers included in our review in more detail. The large majority of papers are published in peer-reviewed journals (85 percent), with an average impact factor of 3.15 at the time of publication. The outlets with the highest numbers of publications in our database are (in descending order) *PLoS One*, the *Journal of Medical Internet Research*, the *Journal of Affective Disorders*, *Social Science Quarterly*, *IEEE Access*, and the *Journal of Public Economics.* This selection of journals demonstrates that although all papers in our database cover a generic social science topic in the broader sense, they often appear in journals that are classified as belonging to other academic disciplines (for a detailed description of these fields of studies see Online Appendix B).

Table 1
Descriptive statistics of papers included in the systematic literature review.

Publication outlet (N = 344):			**Broad WoS categories (N = 344):**[a]		
Non-peer reviewed	53	(15%)	Social Sciences	211	(61%)
Peer-reviewed	291	(85%)	Life Sciences & Biomedicine	100	(29%)
Total	344	(100%)	Technology	59	(17%)
			Multidisciplinary Sciences	22	(6%)
Impact factor of peer-reviewed papers (N = 291):			Physical Sciences	20	(6%)
Mean (Minimum, Maximum)	3.2	(0.1, 13.1)	Arts & Humanities	5	(2%)
Top outlets of peer-reviewed papers (N = 291):			**WoS subfields within the social sciences (N = 211):**[a]		
PLoS One	18	(6%)	Business & Economics	58	(27%)
Journal of Medical Internet Research	10	(3%)	Sociology	43	(20%)
Journal of Affective Disorders	6	(2%)	Political Science	37	(18%)
Social Science Quarterly	5	(2%)	Psychology	36	(17%)
IEEE Access	4	(1%)	Communication	34	(16%)
Journal of Public Economics	4	(1%)	Other Social Sciences	95	(45%)

Note: English-language papers using Google Trends data for applied research in the social sciences. Percentage values rounded to natural numbers.
[a] As each journal can be associated with up to four categories or subfields in Web of Science, the numbers add up to more than 344 (all papers) and 211 (all papers in the Social Sciences category) and 100 percent, respectively.

3.1.3. Constructs measured with search data: Salience, attitudes, or behavior?

Next, we leave the paper level and look at the specific content of each individual study. Table 2 reports the constructs researchers want to measure with Google Trends data. We distinguish between three main categories in our coding: (1) issue salience, (2) attitudes, and (3) behavior. In the case of *issue salience*, researchers used Google Trends as a general measure of public awareness or interest in a specific topic. As an indicator for *attitudes* or other internal states, Google Trends is intended to measure positive or negative feelings and sentiments towards, for instance, a topic, person, or issue. Lastly, the *behavior* indicator is supposed to show the proportion of a population actually doing something offline or intending to do something offline (behavior in the form of factual outcomes). Studies can measure several constructs simultaneously if they retrieve data from Google Trends to measure several variables such as the dependent and independent variable of one model.[14] Given the close relation between Internet searches and interest in the object of the enquiry, it is not surprising that more than half of the studies included in our literature review (60 percent) uses search data to measure issue salience (e.g., Pan and Siegel, 2020 studying interest of the Saudi Arabian public in imprisoned opinion leaders). A substantial fraction of studies attempts to go further than measuring mere issue salience, however. Specifically, 22 percent of the studies examine attitudes such as racism and sexism, and another 24 percent try to measure or predict behavior such as voting (see, for instance, Stephens-Davidowitz, 2014; Owen and Wei, 2021; and Askitas, 2015a,b). Fig. 5 depicts the trends in research with respect to these three constructs from 2010 to 2021. The general increase in the usage of Google Trends data over this period is equally driven by rising numbers of studies in each of the three categories. Additionally, we see that issue salience tops the list in all but one of the years. While researchers aimed for observing behavior earlier than for measuring attitudes, the number of studies examining attitudes overtook the respective number of studies focusing on behavior in 2018 and has remained as runner-up ever since.

3.1.4. Approach to collecting Google Trends data

Table 3 presents the details of how authors collect and analyze Google Trends data in their studies. Starting with the time and regional units[15] examined, we find that researchers primarily use Google Trends to retrieve monthly or weekly data (36 and 26 percent, respectively), either at the country level (61 percent) or the subregional level such as U.S. states (20 percent). With respect to the number of countries considered, two thirds of all studies look at single-country cases, 14 percent compare data for multiple countries, and 21 percent examine worldwide Google searches. The country receiving by far the most attention is the United States, with 47 percent of all studies either focusing exclusively on the U.S. or including them in multi-country analyses. For comparison, the second and third most frequently examined countries are the UK and Italy, which both appear in a bit less than nine percent of the studies in our dataset. In 48 percent of the studies, the researchers use Google Trends data to compare search frequencies across time, in 10 percent across regions, and in 40 percent simultaneously across both region and time. In accordance with the high prominence of the US and the UK among the examined countries, the majority of studies search for keywords in English (59 percent). The distant second and third most frequent languages are Spanish and German, for which keywords are retrieved in eight percent and six percent of the studies, respectively. No other language was found in more than six percent of studies. In 12 percent of studies, researchers do not extract data for a term in a specific language but download the trends for a topic so that the data include the search volume for that topic in all languages and with all possible spellings. Interestingly, almost half of the studies extract data either for a single keyword only (17 percent) or for a large number of 10 or more individual keywords (29 percent). Finally, we could not find the complete

[14] In cases of ambiguity about which construct the studies measured with Google Trends, all three authors discussed and decided together on the final coding.
[15] In cases where researchers downloaded Google Trends data for several regional or time units within the same study, we always coded the smallest unit.

Table 2
Constructs measured with Google Trends data - Descriptions and examples.

	Issue salience/Public interest	Attitude	Behavior
N (%)	215 (60%)	80 (22%)	86 (24%)
Description	Google Trends data used as a general measure of the population's interest in a specific topic or search term.	Google Trends data used as a measure of positive or negative feelings and sentiments towards a topic.	Google Trends data treated as an indicator for offline behavior (factual outcomes). Assumed to show the proportion of people doing something offline or having the intention to do something offline.
Typical examples	Searches for "coronavirus" as public awareness of the COVID-19 pandemic (Hou et al., 2020). Searches for "global warming" as salience of climate change (Anderegg and Goldsmith, 2014).	Searches for the n-word as an indicator for an area's prevalence of racial animus (Stephens-Davidowitz, 2014). Searches for "depression" or "anxiety" as an indicator for mental wellbeing (Banerjee, 2018).	Searches for "birth control" as an indicator of prevalence of contraceptive methods used in society (Guendelman et al., 2020). Searches for "vote yes" and "vote no" as indicators for voting behavior (Askitas, 2015b).

Note: N = 355 (English-language studies using Google Trends data for applied research in the social sciences). Individual papers can contain multiple distinct studies using Google Trends data. Percentage values rounded to natural numbers. We leave out 5 studies for which we could not determine the construct the author(s) of the paper tried to measure with Google Trends. Individual studies can be categorized in more than one construct if the author(s) used Google Trends data to measure more than one distinct variable. For that reason, the absolute and relative frequencies can add up to more than 355 and 100 percent, respectively.

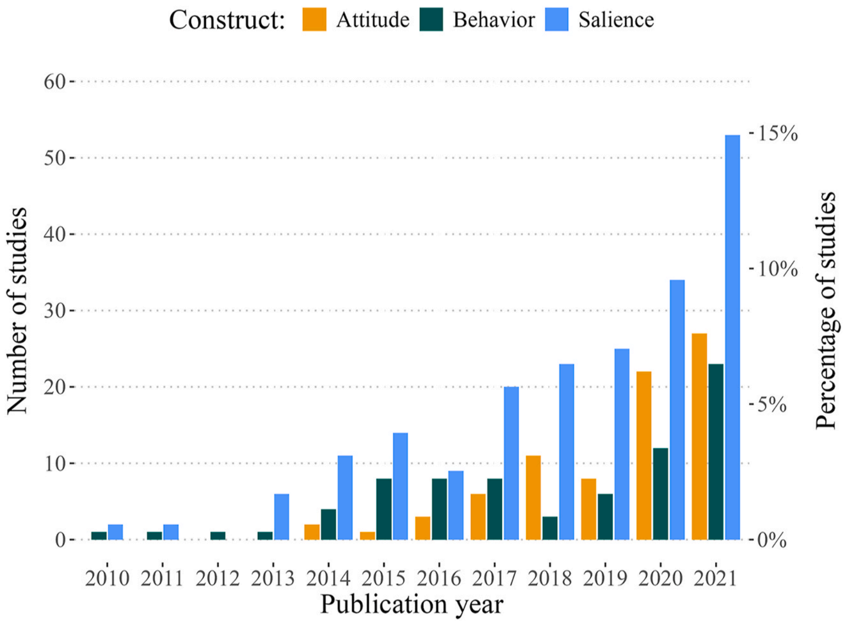

Fig. 5. Measurement of salience, attitudes, and behavior with Google Trends over time. Note: Bar chart indicating the distribution of studies measuring salience, attitudes, and behavior with Google (y-axis) from 2010 to 2021 (x-axis). Given that we searched the papers in late 2021, we assigned studies with an already scheduled publication in 2022 to 2021.

information about these basic parameters of data collection in 24 percent of the studies in our review, which is a worrying sign for the transparency and replicability of research conducted with Google Trends.

3.1.5. Statistical analyses researchers conduct with Google Trends data

In their statistical analysis, 70 percent of all studies combine Google Trends data with information from other sources. These other data sources include administrative data (29 percent) such as the *official number of COVID-19 cases* (e.g., Ma and Ye, 2021), survey data (21 percent) like the *American Community Survey* (e.g., Chae et al., 2018; Stephens-Davidowitz, 2014), and data from social media platforms (eight percent) such as Twitter/X (e.g., Rossouw et al., 2021). Additionally, almost half of the studies (48 percent) rely on a wide range of other sources of third-party information, chief among them, data obtained from private websites. Such third-party websites include *Wikipedia* (e.g., Bail et al., 2019), *Google Ads* (e.g., Behnen et al., 2020), or *America's Health Rankings* (e.g., Isoya and Yamada, 2021). Regarding the role of Google Trends in the analyses, the vast majority of reviewed studies (94 percent) use Google Trends data as one of their main data sources. That is, these studies include a variable based on Google Trends data as the dependent, independent, or control variable in their main analysis of interest and not to merely check the validity of another data source (e.g.,

Table 3
Data collection and analysis.

A. Data collection parameters:

Time unit:		
Minute	2	(1%)
Hourly	9	(3%)
Daily	69	(19%)
Weekly	92	(26%)
Monthly	128	(36%)
Yearly	10	(3%)
Unclear	50	(14%)
Total	360	(102%)
Regional unit:		
City	10	(3%)
Metro	15	(4%)
Subregion	73	(20%)
Country	218	(61%)
Worldwide	20	(6%)
Unclear	24	(7%)
Total	360	(101%)
Number of countries:		
1	236	(66%)
2	13	(4%)
3	8	(2%)
4	6	(2%)
≥5	22	(6%)
Worldwide	75	(21%)
Total	360	(101%)

Region:[a]		
US	170	(47%)
UK	32	(9%)
Italy	31	(9%)
Worldwide	75	(21%)
Unclear	23	(6%)
Level of comparison:		
Region	36	(10%)
Time	171	(48%)
Both region and time	143	(40%)
Across terms	2	(1%)
Unclear	8	(2%)
Total	360	(101%)
Search term language:[a]		
English	211	(59%)
Spanish	28	(8%)
German	20	(6%)
Topic	42	(12%)
Unclear	17	(5%)
Number of keywords:		
1	61	(17%)
2–9	184	(51%)
≥10	104	(29%)
Unclear	11	(3%)
Total	360	(100%)

Missing information on at least one data collection parameter: 87 (24%)

B. Statistical analyses:

Combination with another data source:[a]		
Administrative data	105	(29%)
Survey data	76	(21%)
Social media data	27	(8%)
Other third party data	171	(48%)
Any combination with another data source	252	(70%)
Google Trends as (one of) the main data source:		
Main data source	338	(94%)
Additional data source	22	(6%)
Total	360	(100%)

Univariate, bivariate, or multivariate analysis (N = 338):[b]		
Univariate	74	(22%)
Bivariate	84	(25%)
Multivariate	179	(53%)
Unclear	1	(0.3%)
Total	338	(100%)
Google Trends as the dependent or independent variable in multivariate analyses (N = 179):[c]		
Dependent	59	(33%)
Independent	72	(40%)
Independent & dependent	44	(25%)
Does not apply[d]	4	(2%)
Total	179	(100%)

Note: Descriptive statistics of studies included in the systematic literature review.
N = 360 (English-language studies using Google Trends data for applied research in the social sciences) if not indicated differently in the table note. Individual papers can contain multiple distinct studies using Google Trends data. Percentage values rounded to natural numbers.
[a] Each study could be assigned to more than one category.
[b] Variable only coded if Google Trends is used as (one of) the main data sources (N = 338).
[c] Variable only coded for multivariate analyses (N = 179).
[d] In these papers, another multivariate analysis was applied such as exploratory factor analysis.

Twitter data) by comparing it to Google Trends data. In the latter cases, Google Trends data serve as an additional data source (e.g., Rill et al., 2014; Shahzad et al., 2017). From the studies using Google Trends as one of their main data sources, around a quarter each conducts univariate and bivariate analyses, while a bit more than half run multivariate models. In 33 percent of these multivariate models, Google Trends is included as the dependent variable, in 40 percent as the independent variable, and in around 25 percent as both.

In conclusion, the systematic review shows that researchers increasingly leverage the practical and methodological benefits of Google Trends data to examine a wide variety of research topics at different regional levels, taking advantage of the immediate availability and high frequency of the data, as well as the honesty of search queries. In the next subsection, we examine whether and how well researchers acknowledge and address the associated methodological challenges of internal validity, reliability, and generalizability, which are the necessary conditions for deriving meaningful conclusions from the obtained results.

3.2. How does the existing research treat the methodological challenges of Google Trends data?

3.2.1. Challenge 1: Do researchers check the internal validity of their Google Trends measure?

Downloading data from Google Trends is simple, but knowing what the data actually tell us is difficult. To systematically document and assess authors' efforts to choose appropriate search terms and verify that their selected keywords actually reflect what they want to measure, we code two separate aspects for each study in our review: (1) Did authors justify their choice of keywords ex ante? And (2), did authors check the validity of the selected keywords in any statistical way ex post? The upper half of Table 4 reports the results for these two categories of construct validity. The findings show a bleak picture of the state of the literature. Despite the importance of keyword selection, only 58 percent of included studies provide any reasoning (independent of the quality of that reasoning) for why the authors chose these keywords and why they think these keywords reflect their construct of interest. Maybe even more alarming, only a third (35 percent) conducted any checks for internal validity in the attempt to support their keyword selection and verify that their keywords actually measure the intended theoretical construct. Thus, almost two thirds fail to validate their selected terms at all, potentially severely restricting the meaningfulness of the obtained results. Thus, a large number of studies implicitly require the readers to simply trust the presented measurement and derived conclusions.

Looking more closely at the strategies the authors used to justify their keyword selection, the single most frequent explanation in our review is the popularity of the keyword (16 percent of all studies). In this approach, researchers tested several potential keywords and decided for the most popular search term with the highest search frequency and the lowest number of values of 0. In second place, 14 percent of the studies adopted the operationalization from previous studies by selecting the same search terms other researchers had used to measure the same construct with Google Trends. Another popular way to find inspiration for keywords is to make use of the 'related searches' feature provided by Google Trends (10 percent). Few studies based their decision on theoretical considerations or items in existing surveys (seven and two percent, respectively). The lion's share of explanations, however, consists of a large variety of types of qualitative reasoning (39 percent). For instance, Barnes et al. (2015) rely on thesaurus.com to search for synonymous keywords. Baram-Tsabari and Segev (2015) argue that they decided for Nobel laureates' names to study their global prominence as the authors assume that the names are written the same in every language. In other studies, the authors simply stated that they deemed the selected keywords relevant for their construct of interest (e.g., Barros et al., 2019).

In regards to the type of validity check conducted, 21 percent of all studies compare the results coming from Google Trends with measurements of the same phenomenon or theoretical construct from another data source (criterion validity).[16] Mellon (2013, 2014), for instance, tests several Google Trends indicators for the salience of certain topics against responses to a Gallup survey question about the most important problem the country currently faces. To this end, the author correlates the time series from Google Trends with the repeated cross-sectional survey data using OLS regressions after checking for time series stationarity and seasonality and determining validity based on the regression coefficient's significance and the model's overall R^2. Another eight percent of the studies in our review simply deem their search terms as valid, for instance, by looking at what else individuals googled when searching for their selected keyword. Seven percent of the studies compare the Google Trends measurements with indicators from another data source for a different (social) phenomenon that is thought to be theoretically related to the Google Trends construct of interest (another form of criterion validity). One example here is Hamamura and Chan (2018), who examine whether Google Trends data can be used to measure anxiety levels within the Japanese population. The authors apply a regression discontinuity design by comparing anxiety-related searches before and after an earthquake and use the resulting effect size and significance of the discontinuity variable as an indicator for validity. Finally, two percent of the studies check the selected keywords' discriminant validity, for instance, by testing whether their correlations with an external benchmark are stronger than the correlations of a set of random keywords with that same benchmark (e.g., Lorenz et al., 2022). Eight percent conduct other validity checks.

Unfortunately, we do not see any trend of improvement over the years and across constructs. The proportion of studies providing any reasoning for their keyword selection stays relatively stable over time while the number of studies conducting validity checks even declines over the years. Whereas 44 percent of studies checked the internal validity of their keywords in 2014, only 33 percent did so in 2021. Likewise, we cannot attribute this lack of reasoning for keyword selection and validity checks solely to the subgroup of studies that aim to measure issue salience. In this group of studies, indeed only 47 percent explain their decision for specific keywords and as few as 22 percent conduct any validity check. However, we also only find any reasoning for keyword selection in 75 and 68 percent of studies measuring attitudes and behaviors, respectively, and validity checks in about half of the studies examining one of these two constructs. These findings show that there is much room for improvement across all constructs of interest.

3.2.2. Challenge 2: Do researchers ensure that their measures are reliable?

For reliability, the picture that emerges from our systematic literature review is even more troubling (see lower left corner of Table 4).[17] Only 26 studies, or seven percent of our database, try to account for potential instability of their Google Trends sample by downloading several samples and, for example, taking their average for the analyses (e.g., Adamczyk et al., 2021; DiBartolomeo et al.,

[16] Note that researchers can conduct more than one validity check in a study.

[17] As researchers have no influence on, for instance, any changes of Google's search algorithm or the way Google calculates the RSI, we focus on what researchers can do here, namely, somehow account for the variance of the samples drawn from Google Trends.

Table 4
How researchers deal with the methodological challenges of Google Trends data.

A. Construct validity:					
Reasoning:[a]			**Validity checks:**[a]		
GT 'related searches' function	36	(10%)	On the same construct	77	(21%)
Popularity of the keyword	59	(16%)	On another construct	25	(7%)
Theoretical considerations	25	(7%)	Face validity	30	(8%)
Survey items	6	(2%)	Discriminant validity	8	(2%)
Previous literature's GT keywords	52	(14%)	Other validity check	29	(8%)
Other reasoning	140	(39%)	*Any validity check*	125	(35%)
Any reasoning	208	(58%)			
B. Reliability:			**C. Generalizability:**		
Sample stability checks:			**Considerations of generalizability:**		
Fewer than seven samples	9	(3%)	Representativeness mentioned	77	(21%)
Seven samples or more	11	(3%)	Representativeness explicitly not seen as a concern	22	(6%)
Multiple samples, but exact number unclear	6	(2%)	*Representativeness seen as a problem*	55	(15%)
Any sample stability check	26	(7%)			

Note: Descriptive statistics of studies included in the systematic literature review.
N = 360 (English-language studies using Google Trends data for applied research in the social sciences). Individual papers can contain multiple distinct studies using Google Trends data. Percentage values rounded to natural numbers.
[a] Each study could be assigned to more than one category.

2021; Lui et al., 2011; Medeiros and Pires, 2021). More specifically, nine studies retrieve less than seven samples, 11 studies download seven and more samples, and six studies use multiple samples but do not state the exact number. The large majority, however, does not consider the issue of reliability at all. The only bright spot is an improvement over time: In the earlier years, not a single study checked the stability of its sample. This changed in 2014, when Anderegg and Goldsmith (2014) were the first to do so by collecting several samples on the same parameters within one month. In 2020 and 2021, we already observe nine and 10 studies, respectively, that examine the stability of their sample. This increase constitutes a statistically significant rise compared to the period up to 2019 (p-value $<$0.01).[18] Overall, however, research using Google Trends still shows large deficits when it comes to ensuring the stability of the samples used.

3.2.3. Challenge 3: Do researchers consider the (lacking) generalizability of their data?

Regarding issues of generalizability of Google Trends data, the situation is slightly better than for reliability, but still highly concerning. As can be seen in the lower right corner of Table 4, only 77 studies or 21 percent of our database talk about the external validity of their Google Trends data (or lack thereof) at all.[19] Out of these 77 studies, only 55 (15 percent of all studies considered) acknowledge that their results may not extrapolate to the general population due to the characteristics of Google's user base. The other 22 studies, however, state that they see the Internet search data as generalizable without any further concerns. Similarly to what we found for the issue of keyword validation, we also do not observe a clear positive trend with respect to generalizability over the years.

3.2.4. Overall assessment of how the literature deals with the methodological challenges

Fig. 6 sums up the findings of our systematic review with respect to the specific challenges of using Google Trends data. Many studies do not provide any reasoning for why they selected specific keywords, do not conduct any checks on whether these keywords actually measure what they are supposed to measure (internal validity), rarely ever undertake steps to ensure that their samples are stable (reliability), and do not discuss the generalizability of their data. More specifically, only 208 studies (58%) explain their keyword selection reasoning, 125 (35%) check the validity of these keywords, 26 (7%) retrieve several samples to check their sample reliability, and 55 (15%) studies see Google Trends' generalizability as an issue of the data.[20] Overall, of the 360 studies in our review, only four (!) pay attention to all methodological challenges presented (Connor et al., 2019; Kalmoe, 2017; Knipe et al., 2020; Lee, 2020). We also do not have much reason to assume that this situation will improve by itself: Although slightly more studies explain their keyword selection and account for the stability of their sample over time, we do not see such a trend for keyword validation and considerations around the generalizability of the data.

[18] To check for time trends on how studies dealt with the identified methodological challenges, we conducted two-sided t-tests comparing studies published up to 2019 with those appearing in 2020/21. With the exception of sample stability, we do not find any statistically significant differences on the 5 percent level (see also Online Appendix C).
[19] Specifically, we checked whether studies mention the terms "represent*", "generaliz*", "self-select", "select*", "external validity," and "coverage" in the proper context.
[20] Readers interested in seeing which studies accounted for the challenges in what way can find more information in our data set provided in the Online Supplemental Material on OSF.

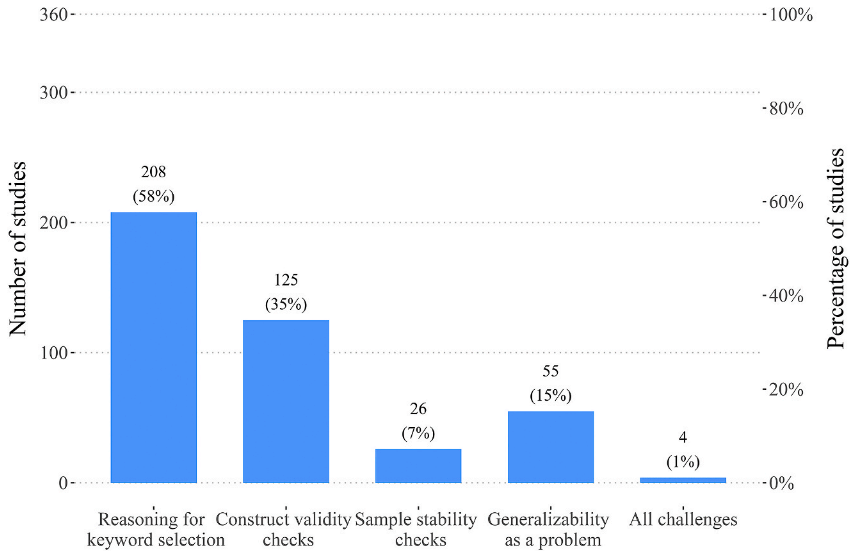

Fig. 6. Number of studies considering the challenges. Note: Bar chart indicating the distribution of studies that consider a specific challenge, namely providing reasoning for the selected search terms, checking whether the Google Trends measure is a valid measure of the construct of interest, ensuring the reliability of the retrieved data from multiple samples, and considering the potentially lacking generalizability of the data.

Interestingly, we also do not find any significant differences in dealing with these methodological issues between peer-reviewed and non-peer-reviewed studies in our data set (see Online Appendix C for details), which suggests that reviewers and editors may not be deeply familiar with Google Trends data and its features and pitfalls.

The four studies that do account for all the identified challenges of Google Trends, however, show that it is possible to do both: advance social science research on a wide range of topics by making use of the data source's benefits while also exhibiting methodological rigor. For the field of political communication, Kalmoe (2017) finds evidence on the impact of high-profile events on public news attention at the example of drone strikes. Coming from psychology, Connor et al. (2019) shed more light on a long contested research field as they find only a small association between income inequality and racial animus within a state. In the field of Life Sciences and Biomedicine, Google Trends shows particular relevance paired with methodological rigor by being able to track the public's mental health concerns. Here, Knipe et al. (2020) examine the impact of the COVID-19 pandemic and the subsequent public health measures on the public's mental health and financial worries as expressed in their Google searches. Similarly, Lee (2020) makes use of search queries to predict suicide rates potentially advancing suicide prevention. These four studies are the glaring exception in the overall picture of the studies examined in the review. It is therefore high time for a comprehensive debate about rigorous standards for using Google Trends data in the social sciences.

4. How to use Google Trends data better for research: Practical recommendations

Table 5 presents a first suggestion for such standards, presented as a list of practical aspects researchers should consider when working with Google Trends data and our recommendations on how they can account for these aspects. The recommendations are derived from our assessment of the challenges of Google Trends data, the findings of our systematic literature review, and previous methodological work on the use of the search volume data, for example, by Behnen et al. (2020), Nuti et al. (2014), and Stephens-Davidowitz and Varian (2015). Table 5 also includes Nuti et al.'s (2014) guidelines to ensure the replicability of results based on Google Trends data. Furthermore, we highlight papers that provide examples of how to address issues of keyword selection, internal validity, reliability, and generalizability.[21]

Before we dive into the details, however, a final note of caution: Even if researchers follow all the recommendations as presented in this chapter, they still need to be careful when their aim is to draw causal conclusions. In this case, researchers need to implement an appropriate identification strategy that allows for causal inference and we would like to refer interested researchers to the work by Angrist and Pischke (2009), Antonakis et al. (2010, 2014), and Wooldridge (2010) at this point.

[21] In our recommendations, we focus on how researchers can deal with the methodological peculiarities of Google Trends to ensure that their measures are transparent, replicable, reproducible, and valid. Researchers interested in a step-by-step introduction on the basics of how to download data from the Google Trends website should consult the instructions on the use of Google Trends provided by the Google News Initiative (2024c).

Table 5
Checklist - Recommendations on using Google Trends for research.

	Topic	Recommendation	Potential values/examples
Replicability and reproducibility	Access date(s)	Provide the date(s) of last accessing Google Trends and downloading the sample(s) (Nuti et al., 2014).	Day, month, year
	Location(s) of access	State the location(s) from where Google Trends data have been accessed. If you used a VPN connection, state that, too.	City, country; VPN-connection to City, country
	Time period	State the precise time frame(s) for which you retrieved the data (Nuti et al., 2014).	Day, month, year
	Time unit	State the time unit for which you retrieved the data.	Every minute, hourly, daily, monthly, yearly
	Region	State the country or region for which you retrieved the data.	Name of the country or region; worldwide
	Regional unit	State for which regional unit/level you retrieved the data.	City, media market level, subregion, country, worldwide
	Query category	State whether you downloaded the sample for your keyword searched for across all categories or whether you restricted your download to specific categories (Nuti et al., 2014).	"All query categories were used" (Nuti et al., 2014:45); Specific category name (e.g., News).
	Full search input	State the full search input (all keywords or topics and their combinations, use of quotation marks, minus or plus signs) (Nuti et al., 2014).	"term 1 + term 2 - term 3″
	Data and code dissemination	Make the retrieved data and code publicly available.	FAIR principles (Wilkinson et al., 2016).
Challenges	Internal validity		
	Rationale for keyword selection	Provide the reasoning behind your choice of keywords or topics. Per default, use keywords instead of topics.	Keyword selection based on theoretical considerations on the latent construct of interest; based on the operationalization and validation of previous studies using Google Trends to measure the same construct.
	Construct	State the theoretical (latent) construct you want to measure with Google Trends.	Salience, attitude, or behavior.
	Validity check	Check the validity of your construct (Mellon, 2013, 2014).	Using another data source to compare Google Trends data to, such as survey or administrative data; checking the search terms' discriminant validity via a comparison with a set of random keywords.
	Reliability	Download samples from Google Trends several days in a row for the exact same parameters and compare the time series you retrieve. Seriously reconsider using Google Trends as a data source for your research application if the samples largely differ between download times. If the samples differ only slightly from each other, take the average of the retrieved samples and provide information on the magnitude of variation among the samples (Behnen et al., 2020; Mavragani and Ochoa, 2019; Medeiros and Pires, 2021).	Samples downloaded on 20 consecutive days and use of average values from the samples for further analyses. Confidence intervals have been calculated using bootstrapping.
	Generalizability	Consider for whom your data might be generalizable (see e.g., Schober et al., 2016).	Google Trends data only generalizable to searches of Google users in the specific region and at the time of interest.

4.1. Getting started: Make sure that your research using Google Trends is replicable, reproducible, and transparent!

Scientists need to ensure that consumers of their work can understand their decisions, replicate their results, and scrutinize their conclusions. To make the data collection transparent and replicable, researchers should retrieve their Google Trends samples using a statistical software package such as *pytrends* for Python (DeWilde and Hogue, 2022) or *gtrendsR* for R (Massicotte and Eddelbuettel, 2022) and provide the code instead of manually downloading the data from the Google Trends website. In particular, researchers need to provide information on when and where they retrieved their sample(s), for which categories, time range(s), and time unit(s), and for which region(s) and regional unit(s) (see also Nuti et al., 2014). They should further present details on the keyword or topic selection process by clearly stating which keywords or topics they retrieved, in which language and 'category' from Google Trends (e.g., "All categories" or "Food and Drink"), whether and how they combined several keywords, and why they chose these specific keywords or topics. We strongly encourage researchers to select keywords instead of topics as it is unknown how and based on which searches Google combines the search volume of different terms into the RSI for topics as well as whether the aggregation method has changed over time. As there are multiple options for specifying keywords, researchers should also report their full search string in detail (Nuti et al., 2014). This way, decisions become fully transparent with respect to either using quotation marks to restrict the results to exact search phrases or including every search that contains the keywords in any order. Likewise, researchers need to indicate the use of options like plus and minus signs to include or exclude specific terms from the search string (Google News Initiative, 2024a,b;

Stephens-Davidowitz and Varian, 2015).

Finally, as for all data sources, researchers should adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) principles in their management and dissemination of Google Trends data to make their analyses reproducible (for a detailed overview on how to apply the FAIR principles, see Wilkinson et al., 2016). As Google Trends is operated by a private company and given the data source's reliability issues, researchers cannot simply trust that the service will continue to provide the data in the same way in the future. It is thus ever more important that researchers make the information on their data collection parameters publicly available together with their retrieved data in an easily findable, accessible, interoperable, and reusable format.

4.2. *Refine the data appropriately for the planned analysis!*

Google Trends data come in a specific structure and with several technical limitations. Depending on their intended analysis, researchers may therefore need to employ a particular strategy to circumvent problems of values of 0, to compare more than five search terms at the same time, or to rescale and refine the data downloaded from Google Trends. In this section, we list suggestions from the literature that appear plausible to us on how to deal with these issues that may arise in the data collection process with Google Trends. Note, however, that these approaches have not been thoroughly examined so far, which means that we know only little about their potential impact on the results. It is therefore even more important that researchers are transparent about what they do, why they do it, and how that may affect their findings.

(1) *Analysis of data with values of 0*: One problem researchers might face in their data collection process is samples with many values of 0, as Google assigns a 0 to data points that have low search volumes below a specific, unknown threshold. In this case, researchers can either choose another search term or try to circumvent the problem with an approach suggested by Stephens-Davidowitz (2014) and refined by DiGrazia (2015). They propose to combine the search terms of interest with a benchmark word such as 'noodle' or 'weather'. The benchmark term should have relatively consistent search frequencies over time and meet the minimum search threshold throughout the whole observation period. Researchers should download the data for a search string with just the benchmark term and another time series for the combined search term of interest and the benchmark term using the "+" sign. The difference between the combined time series and the one just for the benchmark term approximates the time series for the search term of interest without values of 0 (see DiGrazia, 2015 for the details of this method, including the necessary intermediate steps and final adjustments).[22]

(2) *Analysis of more than five search terms*: When researchers want to simultaneously compare the search trends of more than the five terms allowed by Google Trends, they need to extract multiple separate data sets (e.g., for eight keywords at least two) and combine them. Because Google Trends scales the RSI values of each data set relative to the search term with the highest frequency within the data set, the results are not directly comparable across data sets. To bring them on a common scale, one can retrieve the trends on all terms of interest in comparison to a control term (e.g., two data sets including four keywords of interest each and the same benchmark; see Fowle, 2020). Importantly, this control term should not only have a stable trend but also greater popularity than any of the terms of interest so that the maximum within each data set gets assigned to the control term. In that case, the values of all terms of interest now share the same relative scale and are comparable.[23]

(3) *Analysis of high frequency data over longer periods of time:* Google Trends also limits the availability of high frequency data to certain maximum time periods (e.g., daily data for up to around nine months). If the goal is to retrieve higher frequency data for a longer time span than Google Trends currently allows, researchers can implement the approach proposed by Caporin and Poli (2017). The authors use two sets of time series data as starting points: the disjoint one-month time series with the daily RSI values they want to combine (e.g., one data set for each calendar month of 2022) and the single time series covering the complete time frame of interest with monthly RSI data (in our example, the 12 monthly values for 2022). Next, researchers need to follow a three-step procedure: First, calculate the daily share of searches within each month. Second, weight them by the RSI value of the respective month within the whole observation period. Third, normalize the resulting data points by dividing each data point by the highest overall value and multiplying them with 100. The calculated values represent a daily time series comparable across the complete observation window (e.g., the 365 days of 2022) with one maximum as the reference point on the typical scale from 0 to 100. For a more detailed description of this approach, we refer interested readers to the original paper by Caporin and Poli (2017) where the authors provide easy to follow step-by-step instructions. Similarly, researchers can expand data at the minute, hour, or week level beyond the limits Google Trends sets per default.

Specifically for daily data, Eichenauer et al. (2022) provide the *trendecon* package in R that offers researchers an easy alternative to retrieve consistent daily time series from weekly and monthly Google Trends data by applying a disaggregation routine. Korab (2022) presents a step-by-step example on how to use this package.

(4) *Analysis of panel data:* As another limitation, Google Trends does not offer a full-fledged panel option that allows researchers to directly compare search frequencies across both time and regions. Researchers interested in panel analysis therefore again need to extract and match different data sets with different relative scales, in this case, regional and time series data. To do so, researchers can use the same approach as suggested for high frequency time series (see Caporin and Poli, 2017), just with different inputs: on the one hand, separate data sets for each individual region for the same time period (e.g., each U.S. state from 2010 to 2020), and on the other hand, a single data set comparing the average search intensities across regions over the time period under consideration (e.g., regional

[22] However, at the time of submission, the feature to use the plus sign as a Boolean "OR" in a query does not seem to work. This approach to deal with values of 0 is therefore at least temporarily not possible.

[23] To facilitate the implementation of this approach, Fowle (2020) provides the Python code used in her article.

comparison data for the U.S. also from 2010 to 2020).

Another option to build panel data is the spatio-temporal approach presented in Memon et al. (2020). This method additionally takes differences across regions in the distribution and growth of population and Internet access into account. As this information is often not available, though, its applicability is limited to special circumstances.

If the aim of the examination is more targeted towards relative changes of search volume rather than overall highest search frequencies, researchers could apply the approach used by Algan et al. (2016, 2019). Here, each regional value is normalized to have a mean of zero and a standard deviation of one.

4.3. Provide the reasoning for your keyword selection!

In line with a replicable description of the data collection and refinement process, researchers need to be clear about what theoretical construct their Google Trends indicator is supposed to measure and based on which criteria they selected their keywords. Additionally, researchers should provide information on whether they checked for and why they discarded other possible keywords along the way.

As the existing research has not yet systematically evaluated the quality of the various approaches of keyword selection, there is no general consensus on how to best select search terms for a specific research purpose. To us, a good starting point is to generate a list of potential keywords from various sources including previous literature, relevant theories, websites on the topic, the online thesaurus (www.thesaurus.com), the website Semantic Link (www.semantic-link.com), qualitative interviews, or Google services such as Google Correlate, Google's search recommendations, and the 'related searches' feature (see, for example, Askitas, 2015a; Böhme et al., 2020; El Ouadghiri et al., 2021; Owen and Wei, 2021; Wilde et al., 2020). In a second step, one can narrow this list of keyword candidates by using appropriate external information. Banerjee (2018) and Moshontz et al. (2019), for instance, conduct surveys asking respondents which keyword(s) on the list they would insert into Google if they had a specific search interest. If researchers want to do international comparisons and need to obtain equivalent keywords for different languages in a transparent and replicable manner, a good way is to use a common and accessible translation software that includes all these languages (Lin et al., 2020).

4.4. Ensure that your Google Trends indicator validly measures the construct of interest!

After having carefully selected their keywords, researchers need to test whether the chosen terms actually reflect their construct of interest. If researchers aim at measuring society's interest in a topic, the theoretical link between searching for a specific keyword and the measured construct is rather direct and plausible. As the motivation behind a query is unknown, that connection is much more questionable and needs more credible validation if we want to interpret the action of searching for a keyword as the intent to behave or holding an attitude.

To validate their keywords, researchers should compare the search data to an indicator for the same or a related phenomenon coming from an external data source. Probably the most systematic approach using survey data is a step-by-step procedure originally proposed by Mellon (2013, 2014). In this approach, researchers manually check whether the chosen term corresponds with the concept of interest by looking at Google's top related searches associated with the keyword test. Keywords that pass this test for content validity are then used to predict the survey measure in an out-of-sample data (having accounted for issues of time series data including seasonality and non-stationarity). If they do so reasonably well,[24] they are considered valid keywords. Arguably, validation against survey data has its limitations for sensitive behaviors and attitudes, as we cannot know whether observed differences between the Google Trends measure and the external indicator stem from actual differences or social desirability and other forms of measurement error in the survey (Stephens-Davidowitz, 2014).

As an alternative to survey data, researchers can also compare their Google Trends measure to indicators from other external sources. For issue salience, such an external validation source could be media coverage (e.g., Adam-Troian et al., 2022; Ripberger, 2011; Scharkow and Vogelgesang, 2011; Yeung, 2019). For the validation of Google Trends as an indicator of behavior, the most direct way is to predict the offline behavior of interest, for instance, suicide attempts and migration decisions (e.g., Avramescu and Wiśniowski, 2021; Chai et al., 2021; Fantazzini, 2014; Lee et al., 2016; McCarthy, 2010; Nakamura and Suzuki, 2021). Likewise, researchers might use external behavioral data to validate Google Trends measures for attitudes if they can make a strong theoretical argument for that relationship.

As a robustness check, researchers can additionally determine discriminant validity by comparing the correlations of their keywords of interest with the external data to the correlations of a set of random keywords with the same external data. To be deemed valid, the selected keywords should show stronger correlations with the external benchmark data than the random keywords. This approach helps to assess the chance of spurious correlations (see, for example, Yeung, 2019).

Once a search term is validated to measure a certain construct with one or several of the suggested approaches, researchers can continue to use the indicator also at the more fine-grained temporal and geographical units than the initial comparison data.

[24] Mellon (2013, 2014) uses the significance of regression coefficients and the regressions' R^2 to determine a keyword's validity. There are currently no general standards, however, that specify at what threshold a keyword can be considered valid. At the least, researchers should thus set the threshold used in their validation approach before conducting the analysis and make their chosen threshold transparent.

4.5. Account for potentially low reliability of your data!

Previous work has shown that both single values and time trends of Google Trends samples retrieved for the same parameters but at different points in time can vary substantively (Behnen et al., 2020; Mavragani and Ochoa, 2019). To increase the stability of Google Trends data, we follow the lead of several authors and recommend retrieving multiple samples for keywords with high search volume on different days and averaging the search frequencies of the individual samples (see Behnen et al., 2020; Eichenauer et al., 2022; Mavragani and Ochoa, 2019; Stephens-Davidowitz and Varian, 2015; and a particularly diligent application with 28 different samples for the same time range and search terms by Medeiros and Pires, 2021). The mean value of samples downloaded at multiple points in time may still contain some error, but the more samples are drawn, the more likely their mean approximates the true population value. As every sample retrieved from Google Trends comes with sampling variation to some extent, however, researchers should not only use the average value of their samples but also indicate the dispersion around this mean value, for example, by calculating confidence intervals via bootstrapping. And if the values differ strongly between samples or even trend in opposite directions, researchers should seriously reconsider using Google Trends as a data source at all.

Finally, researchers need to pay attention to any changes in Google's algorithm that might have occurred during their observation window. We recommend, for instance, to be particularly careful when comparing search trends before and after 2008 as Google introduced the "recommended searches" feature in 2008.

4.6. Be aware of the (lack of) generalizability of your data!

Researchers need to be aware that the user base of Google does not represent the general population, most likely not even the population of all Internet users. In addition, Google Trends only provides aggregate instead of individual data, which makes it impossible to control for and weigh the samples based on individual-level characteristics. Nevertheless, big data sources can have topic coverage instead of population coverage if the trends depicted in the search frequencies represent the salient topics in a society at a specific point in time (Schober et al., 2016). Researchers thus need to decide from case to case whether Google Trends data can be used to draw conclusions about the general population. However, we advise against using Google Trends to infer point estimates for regions with low overall Internet penetration rates or a small market share of Google. Due to the change of Google's user base over time, we also recommend being cautious with the interpretation of long-term trends in the search index (see also Stephens-Davidowitz and Varian, 2015).

5. Conclusions

In this paper, we presented the benefits and challenges of using Google Trends data for scientific work and critically assessed how the existing literature in the social sciences deals with the identified challenges. Google Trends provides immediate and free access to honest data on a wide range of topics across regions worldwide going back to 2004. The tool thus has the potential to complement traditional data sources like surveys and administrative records. However, Google Trends also comes with severe limitations. First, aggregated Internet search data do not tell us anything about the intent behind an individual query. Researchers therefore cannot easily determine what the search for a specific keyword can actually tell us beyond mere interest in a topic. Second, the values from Google Trends are based on a sample of Google searches. The resulting sample instability raises severe concerns about the reliability of Google Trends data for research, especially regarding keywords with overall low search volumes, for countries with low Internet penetration rates, and for small regions. Third, results from Google Trends might not be generalizable as not everyone has access to and uses the Internet, and not every Internet user relies on Google to search for information online. Results based on the frequencies of Google enquiries may therefore not extrapolate to other groups of individuals. In summary, researchers need to carefully weigh the benefits and challenges of using Google Trends data for each specific project.

To document and critically assess the usage of Google Trends data in the social sciences, we conducted a systematic review of the existing literature up until the end of 2021. Our results show a large increase in the use of Google Trends for research purposes over the past decade. The reviewed studies take advantage of the different benefits of Google Trends data to examine a large variety of topics, geographical regions, and time periods. The large majority, however, fails to adequately account for or at least acknowledge the methodological challenges inherent to Google Trends data. In total, only four out of the 360 studies examined (just 1.1 percent of our database!) deal with all of the issues discussed above in a satisfactory way.

Given the sobering findings of our systematic review, we call for the discipline to take a step back and try to improve the work with Google Trends data in three directions. First, before using Google Trends to answer substantive research questions, researchers should focus on laying the methodological groundwork to assess under what circumstances Google Trends provides reliable and valid measures. For instance, a systematic evaluation of the different ways of selecting and validating keywords as an indicator for attitudes would be a great step forward in this respect. We also need more systematic research on how different factors affect the reliability of Google Trends measures, including potential inconsistencies of samples depending on the location of data retrieval and how many samples are necessary to obtain stable mean search frequencies. Finally, researchers need to develop strict quality standards for future applications.[25]

[25] Another important step in this direction would be more transparency from Google Trends itself on the way the search enquiries are stored and the sampling algorithm used for the Trends tool.

Second, applied researchers should only consider Google Trends as a data source for their work if they are confident that they can satisfactorily address all the challenges presented in this paper (see our checklist in Table 5) and comply with further emerging quality standards. Researchers need to carefully select and validate their keywords of choice, especially when measuring attitudes and behavior. They should retrieve Google Trends data across multiple days and average the results to account for instability in samples. Finally, they have to consider the generalizability of Google search data in the interpretation of their findings. When following the recommendations presented in this paper, measuring issue salience may be possible if the specific conditions of stable and generalizable samples are met. Even then, we are still concerned about the valid measurement of behavior and especially attitudes with Internet search data.

Third, journals should support and enforce this direction by looking more closely at how authors deal with the identified challenges when they assess the quality of submitted papers working with Google Trends data. The recommendations put forth in this paper may serve as a helpful tool for researchers, editors, and reviewers along this way.

Online supplemental material

Further replication materials can be found on OSF (DOI 10.17605/OSF·IO/CTN63).

Funding

Funding for this study was provided by the Fritz Thyssen Foundation (Az. 10.21.1.010SO).

CRediT authorship contribution statement

Johanna Hölzl: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Florian Keusch:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Christoph Sajons:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for feedback on earlier drafts of this paper from Anne-Sophie Oehrlein, Jonathan Vincent, Marcel Kappes, Malte Grönemann, Hannah Reitz, and the members of the FK^2 Research Group. We thank Leila Gushakowski and Jonathan Vincent for research assistance. We thank two anonymous reviewers for their extremely helpful feedback.

Appendix A, B, and C. Additional information on Google Trends, the data collection for the systematic literature review, and the detailed description of our robustness checks

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssresearch.2024.103099.

References

Adamczyk, Amy, Liu, Yu-Hsuan, Scott, Jacqueline, 2021. Understanding the role of religion in shaping cross-national and domestic attitudes and interest in abortion, homosexuality, and pornography using traditional and Google search data.". Soc. Sci. Res. 100, 102602. https://doi.org/10.1016/j.ssresearch.2021.102602.

Adam-Troian, Jais, Eric Bonetto, Arciszewski, Thomas, 2022. Using absolutist word frequency from online searches to measure population mental health dynamics.". Sci. Rep. 12 (1). https://doi.org/10.1038/s41598-022-06392-4.

Algan, Yann, Beasley, Elizabeth, Guyot, Florian, Higa, Kazuhito, Murtin, Fabrice, Senik, Claudia, 2016. Big data measures of well-being: evidence from a Google well-being index in the United States. OECD Statistics Working Papers 2016 4 (3), 6–36. https://doi.org/10.1787/18152031.

Algan, Yann, Murtin, Fabrice, Beasley, Elizabeth, Higa, Kazuhito, Senik, Claudia, 2019. Well-being through the lens of the Internet. PLoS One 14 (1), e0209562. https://doi.org/10.1371/journal.pone.0209562.

Anderegg, William R.L., Goldsmith, Gregory R., 2014. Public interest in climate change over the past decade and the effects of the 'climategate' media event. Environ. Res. Lett. 9 (2014), 054005. https://doi.org/10.1088/1748-9326/9/5/054005.

Anderson, D. Mark, Crost, Benjamin, Rees, Daniel I., 2020. Do economic downturns fuel racial animus? J. Econ. Behav. Organ. 175, 9–18. https://doi.org/10.1016/j.jebo.2020.04.004.

Angrist, Joshua, D., Pischke, Jörn-Steffen, 2009. Mostly harmless econometrics: an empiricist's companion. Princeton University Press.

Antonakis, John, Bendahan, Samuel, Jacquart, Philippe, Lalive, Rafael, 2010. On making causal claims: a review and recommendations. Leader. Q. 21 (6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010.

Antonakis, John, Bendahan, Samuel, Jacquart, Philippe, Lalive, Rafael, 2014. Causality and endogeneity: problems and solutions. In: The Oxford Handbook of Leadership and Organizations, pp. 93–117.

Askitas, Nikos, 2015a. Calling the Greek referendum on the nose with google trends. Forschungsinstitut zur Zukunft der Arbeit/Institute for the Study of Labor (IZA).

Askitas, Nikos, 2015b. Predicting the Irish "Gay Marriage" referendum. Forschungsinstitut zur Zukunft der Arbeit/Institute for the Study of Labor (IZA).

Avramescu, Andreea, Wiśniowski, Arkadiusz, 2021. Now-casting Romanian migration into the United Kingdom by using Google search engine data. Demogr. Res. 45, 1219–1254. https://doi.org/10.4054/DemRes.2021.45.40.

Bail, Christopher A., Brown, Taylor W., Wimmer, Andreas, 2019. Prestige, proximity, and prejudice: how Google search terms diffuse across the world. Am. J. Sociol. 124 (5), 1496–1548. https://doi.org/10.1086/702007.

Banerjee, Snehasish, 2018. How does the world google the Internet, anxiety, and happiness? Cyberpsychol., Behav. Soc. Netw. 21 (9), 569–574. https://doi.org/10.1089/cyber.2018.0206.

Baram-Tsabari, Ayelet, Segev, Elad, 2015. The half-life of a 'teachable moment': the case of Nobel laureates. Publ. Understand. Sci. 24 (3), 326. https://doi.org/10.1177/0963662513491369.

Barnes, Christopher M., Gunia, Brian C., Wagner, David T., 2015. Sleep and moral awareness. J. Sleep Res. 24 (2), 181–188. https://doi.org/10.1111/jsr.12231.

Barros, Joana M., Melia, Ruth, Francis, Kady, Bogue, John, Sullivan, Mary O., Young, Karen, Bernert, Rebecca A., Rebholz-Schuhmann, Dietrich, Duggan, Jim, 2019. The validity of Google Trends search volumes for behavioral forecasting of national suicide rates in Ireland. Int. J. Environ. Res. Publ. Health 16 (17), 3201. https://doi.org/10.3390/ijerph16173201.

Behnen, Philipp, Kessler, Rene, Kruse, Felix, Marx Gómez, Jorge, Schoenmakers, Jan, Zerr, Sergej, 2020. Experimental evaluation of scale, and patterns of systematic inconsistencies in Google Trends data. In: Koprinska, I., Kamp, M., Appice, A., Loglisci, C., Antonie, L., Zimmermann, A., Guidotti, R., Özgöbek, Ö., Ribeiro, R.P., Gavaldà, R., Gama, J., Adilova, L., Krishnamurthy, Y., Ferreira, P.M., Malerba, D., Medeiros, I., Ceci, M., Manco, G., Masciari, E., Ras, Z.W., Christen, P., Ntoutsi, E., Schubert, E., Zimek, A., Monreale, A., Biecek, P., Rinzivillo, S., Kille, B., Lommatzsch, A., Gulla, J.A. (Eds.), ECML PKDD 2020 Workshops, Communications in Computer and Information Science, vol. 1323. Springer International Publishing, Cham, pp. 374–384.

Böhme, Marcus H., Gröger, André, Stöhr, Tobias, 2020. Searching for a better life: predicting international migration with online search keywords. J. Dev. Econ. 142 (2020).

Brodeur, Abel, Clark, Andrew E., Fleche, Sarah, Powdthavee, Nattavudh, 2021. COVID-19, lockdowns and well-being: evidence from Google Trends. J. Publ. Econ. 193, 104346. https://doi.org/10.1016/j.jpubeco.2020.104346.

Caporin, Massimiliano, Poli, Francesco, 2017. Building news measures from textual data and an application to volatility forecasting. Econometrics 5 (3), 35. https://doi.org/10.3390/econometrics5030035.

Chae, David H., Clouston, Sean, Hatzenbuehler, Kramer, Michael R., Cooper, Hannah L.F., Wilson, Sacoby M., Stephens-Davidowitz, Seth I., Gold, Robert S., Link, Bruce G., 2015. Association between an Internet-based measure of area racism and Black mortality. PloS One 10 (4), e0122963. https://doi.org/10.1371/journal.pone.0122963.

Chae, David H., Clouston, Sean, Martz, Connor D., Hatzenbuehler, Mark L., Cooper, Hannah L.F., Turpin, Rodman, Stephens-Davidowitz, Seth, Kramer, Michael R., 2018. Area racism and birth outcomes among Blacks in the United States. Soc. Sci. Med. 199, 49–55. https://doi.org/10.1016/j.socscimed.2017.04.019.

Chai, Daniel, Dai, Mengjia, Gharghori, Philip, Hong, Barbara, 2021. Internet search intensity and its relation with trading activity and stock returns. Int. Rev. Finance 21 (1), 282–311. https://doi.org/10.1111/irfi.12268.

Clarivate Analytics, 2020. Web of science core collection help. Clarivate Analytics. https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html. (Accessed 20 July 2022).

Connor, Paul, Sarafidis, Vasilis, Zyphur, Michael J., Keltner, Dacher, Chen, Serena, 2019. Income inequality and White-on-Black racial bias in the United States: evidence from Project Implicit and Google Trends. Psychol. Sci. 30 (2), 205–222. https://doi.org/10.1177/0956797618815441.

DeWilde, Burton, Hogue, John, 2022. Pytrends: Pseudo API for Google Trends.

DiBartolomeo, Jeffrey A., Kothakota, Michael G., Parks-Stamm, Elizabeth J., Tharp, Derek T., 2021. Racial animosity and Black financial advisor underrepresentation. Preprint.

DiGrazia, Joseph, 2015. Using Internet search data to produce state-level measures: the case of tea party mobilization. Socio. Methods Res. 1–28. https://doi.org/10.1177/0049124115610348.

Eichenauer, Vera Z., Indergand, Ronald, Martínez, Isabel Z., Sax, Christoph, 2022. Obtaining consistent time series from Google Trends. Econ. Inq. 60 (2), 694–705. https://doi.org/10.1111/ecin.13049.

El Ouadghiri, Imane, Khaled Guesmi, Peillex, Jonathan, Ziegler, Andreas, 2021. Public attention to environmental issues and stock market returns. Ecol. Econ. 180, 106836. https://doi.org/10.1016/j.ecolecon.2020.106836.

Ettredge, Michael, Gerdes, John, Karuga, Gilbert, 2005. Using web-based search data to predict macroeconomic statistics. Commun. ACM 48 (November 2005), 87–92.

Fang, Qixiang, Joep, Burger, Meijers, Ralph, van Berkel, Kees, 2021. The role of time, weather and Google Trends in understanding and predicting Web survey response. Survey Research Methods 15 (1), 1–25. https://doi.org/10.18148/srm/2021.v15i1.7633.

Fantazzini, Dean, 2014. Nowcasting and forecasting the monthly food stamps data in the US using online search data. PLoS One 9 (11). https://doi.org/10.1371/journal.pone.0111894.

Fenga, Livio, 2020. Filtering and prediction of noisy and unstable signals: the case of Google Trends data. J. Forecast. 39 (2), 281–295. https://doi.org/10.1002/for.2626.

Fowle, Carrie, 2020. Using Google Trends at scale. Medium. https://towardsdatascience.com/using-google-trends-at-scale-1c8b902b6bfa. (Accessed 17 November 2022).

Franzén, Alexandra, 2023. Big data, big problems: why scientists should refrain from using Google Trends. Acta Sociologica 000169932211511. https://doi.org/10.1177/00016993221151118.

Golenvaux, Nicolas, Gonzalez Alvarez, Pablo, Kiossou, Harold Silvère, Schaus, Pierre, 2020. An LSTM approach to forecast migration using Google Trends. Association for computing machinery, San Diego, California, USA.

Google, 2024. FAQ about Google Trends Data. https://support.google.com/trends/answer/4365533?hl=en. (Accessed 18 October 2024).

Google News Initiative, 2024a. Advanced Google Trends. https://newsinitiative.withgoogle.com/resources/trainings/advanced-google-trends/. (Accessed 18 October 2024).

Google News Initiative, 2024b. Basics of Google Trends. https://newsinitiative.withgoogle.com/resources/trainings/basics-of-google-trends/. (Accessed 18 October 2024).

Google News Initiative, 2024c. Google Trends lessons. Google News Initiative Training Center. https://newsinitiative.withgoogle.com/resources/products/google-trends/. (Accessed 18 October 2024).

Google Trends, 2024. Obama. Google Trends. https://trends.google.de/trends/explore?date=all&q=Obama&hl=en. (Accessed 11 October 2024).

Guendelman, Sylvia, Elena, Yon, Pleasants, Elizabeth, Hubbard, Alan, Ndola, Prata, 2020. Shining the light on abortion: drivers of online abortion searches across the United States in 2018. PLoS One 15 (5). https://doi.org/10.1371/journal.pone.0231672.

Gummer, Tobias, Oehrlein, Anne-Sophie, 2022. Using Google Trends data to learn more about survey participation. Social Science Computer Review 089443932211291. https://doi.org/10.1177/08944393221129179.

Hamamura, Takeshi, Chan, Christian, 2018. Googling away our fear: big data analyses on anxiety. Preprint. Open Science Framework.

Hou, Zhiyuan, Du, Fanxing, Zhou, Xinyu, Jiang, Hao, Sam Martin, Heidi Larson, Leesa, Lin, 2020. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. Journal of Medical Internet Research 22 (8). https://doi.org/10.2196/21143.

Isoya, Hidehito, Yamada, Hiroyuki, 2021. Hidden Internet-based measure of discrimination and mental health of Black people. Soc. Sci. Med. 273, 1. https://doi.org/10.1016/j.socscimed.2021.113776.

Jun, Seung-Pyo, Park, Do-Hyung, Yeom, Jaeho, 2013. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. Technol. Forecast. Soc. Change 1–17. https://doi.org/10.1016/j.techfore.2013.10.021.

Jun, Seung-Pyo, Yoo, Hyoung Sun, Choi, San, 2017. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. Technol. Forecast. Soc. Change 130, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009.

Kalmoe, Nathan, 2017. Digital news-seeking during wartime: unobtrusive measures of Pakistani and American attention to drone strikes. J. Inf. Technol. Polit. 14 (1), 16–33. https://doi.org/10.1080/19331681.2016.1261265.

Kiley, Kevin, Vaisey, Stephen, 2020. Measuring stability and change in personal culture using panel data. Am. Socio. Rev. 85 (3), 477–506. https://doi.org/10.1177/0003122420921538.

Knipe, Duleeka, Evans, Hannah, Sinyor, Mark, Niederkrotenthaler, Thomas, Gunnell, David, Ann, John, 2020. Tracking online searches for emotional wellbeing concerns and coping strategies in the UK during the COVID-19 pandemic: A Google Trends analysis [version 1; peer review: awaiting peer review]. Wellcome Open Research. https://doi.org/10.12688/wellcomeopenres.16147.1.

Koehler-Derrick, Gabriel, 2013. Quantifying anecdotes: Google search data and political developments in Egypt. PS Political Sci. Polit. 46 (2), 291–298. https://doi.org/10.1017/S1049096513000267.

Korab, Petr, 2022. Fixing Google Trends data limitations. Medium. https://towardsdatascience.com/trendecon-essential-r-package-to-construct-time-series-from-google-trends-36365ea48382. (Accessed 24 November 2022).

Krippendorff, Klaus, 2004. Content Analysis: an introduction to its methodology, second ed. Sage, Thousand Oaks, Calif.

Krumpal, Ivar, 2013. Determinants of social desirability bias in sensitive surveys: a literature review. Qual. Quantity 47 (4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9.

Lazer, David, Kennedy, Ryan, King, Gary, Vespignani, Alessandro, 2014. The parable of Google flu: traps in big data analysis. Science 343 (March 14), 1203–1205.

Lee, Donghyun, Kim, Minki, Lee, Jungyoun, 2016. Adoption of green electricity policies: investigating the role of environmental attitudes via big data-driven search-queries. Energy Pol. 90, 187–201. https://doi.org/10.1016/j.enpol.2015.12.021.

Lee, Joo-Young, 2020. Search trends preceding increases in suicide: a cross-correlation study of monthly Google search volume and suicide rate using transfer function models. J. Affect. Disord. 262, 155–164. https://doi.org/10.1016/j.jad.2019.11.014.

Lin, Yu-Hsuan, Chiang, Ting-Wei, Lin, Yu-Lun, 2020. Increased Internet searches for insomnia as an indicator of global mental health during the COVID-19 pandemic: multinational longitudinal study. J. Med. Internet Res. 22 (9), e22181. https://doi.org/10.2196/22181.

Liu, Yu-Hsuan, Wolff, Kevin T., Tzu-Ying, Lo, 2023. Big data in crime statistics: using Google Trends to measure victimization in designated market areas across the United States. Methodological Innovations 16 (3), 341–358. https://doi.org/10.1177/20597991231183962.

Lorenz, Reilly, Jacob, Beck, Horneber, Sophie, Keusch, Florian, Antoun, Christopher, 2022. Google Trends as a tool for public opinion research: an illustration on perceived threats of immigration. In: Migration Research in a Digitized World, IMISCOE Research Series. Springer, Cham, pp. 193–206.

Lui, Catherine, Metaxas, Panagiotis T., Mustafaraj, Eni, 2011. On the predictability of the U.S. elections through search volume activity. Avila, Spain.

Ma, Mac Z., Ye, Shengquan, 2021. The role of ingroup assortative sociality in the COVID-19 pandemic: a multilevel analysis of Google Trends data in the United States. Int. J. Intercult. Relat. 84, 168–180. https://doi.org/10.1016/j.ijintrel.2021.07.010.

MacInnis, Cara C., Gordon, Hodson, 2015. Do American states with more religious or conservative populations search more for sexual content on Google? Arch. Sex. Behav. 44 (1), 137–147. https://doi.org/10.1007/s10508-014-0361-8.

Massicotte, Philippe, Eddelbuettel, Dirk, 2022. Package 'gtrendsR'.

Mavragani, Amaryllis, Ochoa, Gabriela, 2019. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health and Surveillance 5 (2). https://doi.org/10.2196/13439.

Mavragani, Amaryllis, Ochoa, Gabriela, Tsagarakis, Konstantinos P., 2018. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J. Med. Internet Res. 20 (11), 1–20. https://doi.org/10.2196/jmir.9366.

McCarthy, Michael J., 2010. Internet monitoring of suicide risk in the population. J. Affect. Disord. 122 (3), 277–279. https://doi.org/10.1016/j.jad.2009.08.015.

McKetta, Sarah, Hatzenbuehler, Mark L., Pratt, Charissa, Bates, Lisa, Link, Bruce G., Keyes, Katherine M., 2017. Does social selection explain the association between state-level racial animus and racial disparities in self-rated health in the United States? Ann. Epidemiol. 27 (8), 485–492.e6. https://doi.org/10.1016/j.annepidem.2017.07.002.

Medeiros, Marcelo, Pires, Henrique F., 2021. The proper use of Google Trends in forecasting models. https://doi.org/10.48550/ARXIV.2104.03065.

Mellon, Jonathan, 2013. Where and when can we use Google Trends to measure issue salience? PS: political. Science & Politics 46 (2), 280–290. https://doi.org/10.1017/S1049096513000279.

Mellon, Jonathan, 2014. Internet search data and issue salience: the properties of Google Trends as a measure of issue salience. J. Elections, Public Opin. Parties 24 (1), 45–72. https://doi.org/10.1080/17457289.2013.846346.

Memon, Shahan Ali, Razak, Saquib, Weber, Ingmar, 2020. Lifestyle disease surveillance using population search behavior: feasibility study. J. Med. Internet Res. 22 (1), e13347. https://doi.org/10.2196/13347.

Moshontz, Hannah, Vuletich, Heidi A., Payne, Keith, Marsh, Elizabeth, 2019. Inequality's association with poor educational outcomes in U.S. states relates to interest. Academic Shortcuts. preprint. PsyArXiv.

Mukherjee, Partha, Jansen, Bernard J., 2017. Conversing and searching: the causal relationship between social media and Web search. Internet Res. 27 (5), 1209–1226. https://doi.org/10.1108/IntR-07-2016-0228.

Nakamura, Nobuyuki, Suzuki, Aya, 2021. COVID-19 and the intentions to migrate from developing countries: evidence from online search activities in Southeast Asia. J. Asian Econ. 76, 101348. https://doi.org/10.1016/j.asieco.2021.101348.

Nuti, Sudhakar V., Wayda, Brian, Ranasinghe, Isuru, Wang, Sisi, Dreyer, Rachel P., Chen, Serene I., Murugiah, Karthik, 2014. The use of Google Trends in health care research: a systematic review. PLoS One 9 (10), 1–49. https://doi.org/10.1371/journal.pone.0109583.

Olzak, Susan, 2021. Does protest against police violence matter? Evidence from U.S. cities, 1990 through 2019. Am. Socio. Rev. 86 (6), 1066–1099. https://doi.org/10.1177/00031224211056966.

Owen, Ann L., Wei, Andrew, 2021. Sexism, household decisions, and the gender wage gap. Lab. Econ. 72, 102062. https://doi.org/10.1016/j.labeco.2021.102062.

Page, Matthew J., McKenzie, Joanne E., Bossuyt, Patrick M., Boutron, Isabelle, Hoffmann, Tammy C., Mulrow, Cynthia D., Shamseer, Larissa, Tetzlaff, Jennifer M., Akl, Elie A., Brennan, Sue E., Chou, Roger, Glanville, Julie, Grimshaw, Jeremy M., Hróbjartsson, Asbjørn, Lalu, Manoj M., Li, Tianjing, Loder, Elizabeth W., Mayo-Wilson, Evan, McDonald, Steve, McGuinness, Luke A., Stewart, Lesley A., Thomas, James, Tricco, Andrea C., Welch, Vivian A., Whiting, Penny, Moher, David, 2021a. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ n71. https://doi.org/10.1136/bmj.n71.

Page, Matthew J., Moher, David, Bossuyt, Patrick M., Boutron, Isabelle, Hoffmann, Tammy C., Mulrow, Cynthia D., Shamseer, Larissa, Tetzlaff, Jennifer M., Akl, Elie A., Brennan, Sue E., Chou, Roger, Glanville, Julie, Grimshaw, Jeremy M., Hróbjartsson, Asbjørn, Lalu, Manoj M., Li, Tianjing, Loder, Elizabeth W., Mayo-Wilson, Evan, McDonald, Steve, McGuinness, Luke A., Stewart, Lesley A., Thomas, James, Tricco, Andrea C., Welch, Vivian A., Whiting, Penny, McKenzie, Joanne E., 2021b. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ n160. https://doi.org/10.1136/bmj.n160.

Pan, Jennifer, Siegel, Alexandra A., 2020. How Saudi crackdowns fail to silence online dissent. Am. Polit. Sci. Rev. 114 (1), 109–125. https://doi.org/10.1017/S0003055419000650.

Rill, Sven, Reinel, Dirk, Scheidt, Jörg, Zicari, Roberto V., 2014. PoliTwi: early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis. Knowl. Base Syst. 69, 24–33. https://doi.org/10.1016/j.knosys.2014.05.008.

Ripberger, Joseph T., 2011. Capturing curiosity: using Internet search trends to measure public attentiveness. Pol. Stud. J. 39 (2), 239–259. https://doi.org/10.1111/j.1541-0072.2011.00406.x.

Rogers, Simon, 2016. What is Google Trends data — and what does it mean? Medium. https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8. (Accessed 18 October 2024).

Rossouw, Stephanie, Greyling, Talita, Adhikari, Tamanna, 2021. Happiness lost: was the decision to implement lockdown the correct one? S. Afr. J. Econ. Manag. Sci. 24 (1), a3795. https://doi.org/10.4102/sajems.v24i1.3795.

Salganik, Matthew J., 2017. Bit by bit: social research in the digital age. Princeton University Press, Princeton, NJ.

Scharkow, Michael, Vogelgesang, Jens, 2011. Measuring the public agenda using search engine queries. Int. J. Publ. Opin. Res. 23 (1), 104–113. https://doi.org/10.1093/ijpor/edq048.

Schober, Michael F., Pasek, Josh, Guggenheim, Lauren, Lampe, Cliff, Conrad, Frederick G., 2016. Research synthesis. Social media analyses for social measurement. Publ. Opin. Q. 80 (Spring 2016), 180–211. https://doi.org/10.1093/poq/nfv048.

Shahzad, Basit, Lali, Ikramullah, Saqib Nawaz, M., Aslam, Waqar, Mustafa, Raza, Mashkoor, Atif, 2017. Discovery and classification of user interests on social media. Information Discovery and Delivery 45 (3), 130–138. https://doi.org/10.1108/IDD-03-2017-0023.

Statcounter, 2021. Search engine market share worldwide. StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share. (Accessed 13 July 2022).

Stephens-Davidowitz, Seth, 2014. The cost of racial animus on a Black candidate: evidence using Google search data. J. Publ. Econ. 118 (2014), 26–40. https://doi.org/10.1016/j.jpubeco.2014.04.010.

Stephens-Davidowitz, Seth, 2017. Everybody lies: big data, new data and what the Internet can tell us about who we really are. HarperCollins, New York.

Stephens-Davidowitz, Seth, Varian, Hal, 2015. A hands-on guide to Google data.

Sturgis, Patrick, Luff, Rebekah, 2021. The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015. Int. J. Soc. Res. Methodol. 24 (6), 691–696. https://doi.org/10.1080/13645579.2020.1844896.

Sullivan, Danny, 2008. Google.com finally gets Google Suggest feature. Search Engine Land. https://searchengineland.com/googlecom-finally-gets-google-suggest-feature-14626. (Accessed 5 December 2021).

Timoneda, Joan C., Erik, Wibbels, 2022. Spikes and variance: using Google Trends to detect and forecast protests. Polit. Anal. 30 (1), 1–18. https://doi.org/10.1017/pan.2021.7.

Tourangeau, Roger, Rips, Lance J., Rasinski, Kenneth, 2000. The psychology of survey response. Cambridge University Press, Cambridge, UK.

Trends Help, 2024a. Compare trends search terms. https://support.google.com/trends/answer/4359550. (Accessed 18 October 2024).

Trends Help, 2024b. Find related searches. https://support.google.com/trends/answer/4355000?hl=en#. (Accessed 18 October 2024).

Trends Help, 2024c. Refine trends results by category. https://support.google.com/trends/answer/4359597?hl=en&ref_topic=4365530. (Accessed 18 October 2024).

Vasi, Ion Bogdan, Walker, Edward T., Johnson, John S., Tan, Hui Fen, 2015. 'No fracking way!' Documentary film, discursive opportunity, and local opposition against hydraulic fracturing in the United States, 2010 to 2013. Am. Socio. Rev. 80 (5), 934–959. https://doi.org/10.1177/0003122415598534.

Wilde, Joshua, Chen, Wei, Lohmann, Sophie, 2020. COVID-19 and the future of US fertility: what can we learn from Google? Preprint. SocArXiv.

Wilkinson, Mark D., Dumontier, Michel, Jan Aalbersberg, IJsbrand, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, Bonino Da Silva Santos, Luiz, Bourne, Philip E., Bouwman, Jildau, Brookes, Anthony J., Clark, Tim, Crosas, Mercè, Dillo, Ingrid, Dumon, Olivier, Scott, Edmunds, Evelo, Chris T., Finkers, Richard, Gonzalez-Beltran, Alejandra, Gray, Alasdair J.G., Groth, Paul, Goble, Carole, Grethe, Jeffrey S., Heringa, Jaap, 'T Hoen, Peter A. C., Hooft, Rob, Kuhn, Tobias, Kok, Ruben, Kok, Joost, Lusher, Scott J., Martone, Maryann E., Mons, Albert, Packer, Abel L., Persson, Bengt, Rocca-Serra, Philippe, Roos, Marco, Van Schaik, Rene, Sansone, Susanna-Assunta, Schultes, Erik, Sengstag, Thierry, Slater, Ted, George, Strawn, Swertz, Morris A., Thompson, Mark, Lei, Johan Van Der, Van Mulligen, Erik, Velterop, Jan, Waagmeester, Andra, Wittenburg, Peter, Wolstencroft, Katherine, Zhao, Jun, Mons, Barend, 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3 (1), 160018. https://doi.org/10.1038/sdata.2016.18.

Wooldridge, Jeffrey M., 2010. Econometric analysis of cross section and panel data. MIT press.

Yeung, Timothy Y.C., 2019. Measuring Christian religiosity by Google Trends. Rev. Relig. Res. 61 (3), 235. https://doi.org/10.1007/s13644-019-00379-w.