Title
Harnessing Google Health Trends API Data for Epidemiologic Research

Journal
American Journal of Epidemiology, 192(3)

ISSN
0002-9262

Authors
Neumann, Krista
Mason, Susan M
Farkas, Kriszta
et al.

Publication Date
2023-02-24

DOI
10.1093/aje/kwac171

Peer reviewed

American Journal of Epidemiology **Submitted Manuscript**

Title: Harnessing Google Health Trends API Data for Epidemiologic Research

Authors: Krista Neumann, Susan M. Mason, Kriszta Farkas, N, Jeanie Santaularia, Jennifer Ahern, Corinne A. Riddell

Correspondence Address: Correspondence to Krista Neumann, Division of Epidemiology, School of Public Health, University of California, Berkeley, Room #5404, 2121 Berkeley Way West, Berkeley, California, 94720 (email: krista.neumann@berkeley.edu)

(The author reviewing editorial queries is Corinne Riddell (c.riddell@berkeley.edu) as Krista Neumann is on parental leave.)

Affiliations: Division of Epidemiology, School of Public Health, University of California, Berkeley, United States (Krista Neumann, Krista Farkas, Jennifer Ahern, Corinne Riddell); Division of Epidemiology and Community Health, University of Minnesota, Minnesota, United States (Susan M. Mason, Krista Farkas, Jeanie Santaularia); Division of Biostatistics, School of Public Health, University of California, Berkeley, United States (Corinne Riddell)

Funding: This work was funded by two pilot grants: one from the National Institute for Child Health and Human Development via the Berkeley Population Center (grant P2CHD073964) and the other from The Hawley Family Fund through the Division of Epidemiology and Community Health at the University of Minnesota. This work was also supported by the National Institute for Child Health and Human Development via grants from the Minnesota Population Center and the Interdisciplinary Population Health Scient Training Program at the University of Minnesota (T32HD095134 and P2C HD041023)

Data Availability Statement: The dataset is restricted and not publicly available.

Thanks: N/A

Conference presentation: This work was presented as a poster presentation at the Interdisciplinary Association for Population Health Science conference (held virtually), October 19-21, 2021

Preprint Information: N/A

Disclaimer: N/A

Conflict of Interest: All authors report no conflicts.

Running Head: Google Health Trends for Epidemiologic Research

Key words: Google, child abuse, abuse

Abbreviations: GHT-API, Google Health Trends Application Programming Interface; GT, Google Trends; DMA, designated market area

Abstract

Interest in using internet search data, such as that from the Google Health Trends Application Programming Interface (GHT-API), to measure epidemiologically relevant exposures or health outcomes is growing due to their accessibility and timeliness. Researchers input search term(s), geography and time period, and the GHT-API returns a scaled probability of that search term, given all searches within the specified geo-time period. In this study, we detail a method for using these data to measure a construct of interest in five iterative steps: first, identify phrases the target population may use to search for the construct of interest; second, refine candidate search phrases with incognito Google searches to improve sensitivity and specificity; third, craft the GHT-API search term(s) by combining the refined phrases; fourth, test search volume and choose geographic and temporal scales; and fifth, retrieve and average multiple samples to stabilize estimates and address missingness. An optional sixth step involves accounting for changes in total search volume by normalizing. We present a case study examining weekly state-level child abuse searches in the United States during the COVID-19 pandemic (January 2018-August 2020) as an application of this method and describe limitations.

Introduction

There is growing interest in using internet search data to characterize epidemiological patterns of exposure and disease because they are accessible, free, and near-real-time. The Google Health Trends Application Programming Interface (GHT-API) is one source of such data. To access these data, after obtaining an API key (1), researchers specify the search term(s), geographic region, and time period of interest, and the GHT-API returns an estimated scaled probability of the search term(s) given a random sample of all Google searches within the specified geo-time period. Google searches, which can be accessed through either the GHT-API or a separate, publicly available Google Trends website (GT) (2) (or its associated GT API) have been used to measure variables that are difficult to capture via traditional data sources, such as abuse (3–5), racism (6), and public sentiment around drinking water contamination and birth control (7, 8). These data can also be used to examine trends when real-time data are beneficial, such as during influenza seasons (9–11).

The GHT-API is distinct from GT (2) and less cited in academic research — PubMed returned only 5 results for "Google Health Trends" compared to 780 results for "Google Trends". While both extract a random sample of all Google searches and allow comparisons of multiple search terms over geographic-temporal periods, their outputs differ (12). GT rank-orders search volumes within the specified geo-time period and returns a search volume indexed between 0 and 100 representing the relative popularity of the search term(s) (13); GHT-API returns the probability of the specified search terms, based on a random sample of all Google searches in the specified geo-time period, scaled by 10 million for readability (2020 *Google Health Trends API Getting Started Guide* [unpublished document provided with API key]). Note that Google does not disclose the total number of searches used to calculate this probability, and thus returned results can only be interpreted as relative volume with an unknown denominator. The GHT-API has advantages over the GT since values are not scaled to the highest result, permitting comparison of search data extracted across different points in time. In order to compare trends over time using the GT, the entire time period of interest needs to be extracted at once so that the scaling doesn't change. In contrast, the GHT-API allows you to compare trends across different time periods, regardless of the time interval for which the data was extracted. This is useful for expanding the time period of interest at later points.

There is little formal guidance about how to craft a GHT-API search strategy to accurately measure a construct of interest (12,14–17). Searches that are too broad can yield high search volumes, but may capture many searches that are less relevant. Searches that are too narrow might be highly specific but result in missing values because the GHT-API suppresses data when the number of searches is below an undocumented threshold. Additionally, sampling variation needs to be considered as the GHT-API estimates probabilities from a random sample that is updated once per day (2020 *Google Health Trends API Getting Started Guide* [unpublished document provided with API key]).

The objective of this paper is to describe best practices when using the GHT-API to measure a construct of interest, using a motivating case study.

Case Study

This manuscript builds on previous work by our team (18). In that study, we examined child abuse and neglect during the COVID-19 pandemic. Typically, abuse and neglect incidences are assessed locally through Child Protection Services data or nationally via the National Child Abuse and Neglect Data System. However, both datasets may be influenced by factors affected by the pandemic, including: opportunities for children to interact with mandated reporters (e.g., through schools); adequate budget for the child protection workforce; circumstances that allow child protection workers to thoroughly investigate reports (e.g., via in-home visits); and timely release of data. Pandemic-related school closures, recession-related budget and workforce cuts, and social distancing mandates that limit in-home visits may converge to reduce the number of child abuse and neglect cases detected through child welfare services during the pandemic, even if incidence of abuse and neglect increased. We therefore investigated GHT-API as an alternate data source that may not be subject to these concerns.

Methods

To measure a construct of interest using GHT-API data, we recommend using the following iterative search strategy approach (Figure 1).

Step 1: Identify how Google users search for the construct of interest

The first step is to ascertain how target internet users might describe the construct of interest in their Google searches and generate a comprehensive list of potential search phrases. A literature search and expert opinions may be useful approaches for identifying relevant search phrases.

Case Study Application:

To determine how children might describe abuse and neglect in their searches, we performed a qualitative literature search and reviewed validated survey instruments (19, 20) to identify first-hand descriptions of abuse and neglect.

Next, we generated a list of specific verbs used to describe physical, emotional, and sexual abuse as well as neglect (e.g., 'hit', 'threaten', ‘molest’, ‘left alone’), and identified the most common and relevant abuse perpetrators. We then created a series of abuse phrases we thought were likely to be searched by children using the general format of ‘*[perpetrator] + [verb] + [victim]’* (e.g., 'mom hit me', 'dad threatened me', 'mom’s boyfriend molested me'). We opted for broader searches such as ‘mom hit me’ rather than more specific searches like 'mom hit me in the face' to try to capture as many cases of abuse and neglect as possible. Since returned GHT-API results include searches that contain all words specified in each phrase in addition to other words (see Web Appendix 1 for details), the results for the former phrase will encapsulate the latter, more specific phrase. We also included the term *'child abuse hotline'* to capture searches from those looking to report or find child abuse resources.

Step 2: Use incognito searches to improve sensitivity and specificity

After enumerating a comprehensive list of phrases related to the construct of interest, the second step is to refine this list so that phrases are broad enough to encompass the intended construct but narrow enough to limit the number of false positives. It is up to the researcher to determine what level of specificity is desired, the degree to which false positives will impact interpretability, and ultimately which search phrases capture the construct they intend to measure based on context expertise and existing literature. One way to do this is to perform Google searches for each potential search phrase using a Google Chrome incognito browser (14). The incognito browser ensures that information about the researcher does not influence the search results, thereby presenting a more generalized set of results. By reviewing returned results, phrases with results that are not relevant to the desired construct can be identified. This strategy can also lead to the identification of additional phrases missed in the first step through Google's autocomplete feature and through returned results.

Case Study Application:

For our outcome of interest, we wanted to capture searches motivated by child abuse or neglect incidence. We thus sought to minimize false positives by excluding search phrases in which the searcher was not likely to be experiencing or witnessing abuse. We assessed the relevance of specific search phrases by performing an incognito search and examining their results; if the top results were unrelated to experienced or witnessed abuse, we deemed the phrase too broad for inclusion. Thus, broad terms like *'child abuse'* were discarded because results revealed that they were used by people looking for general information on the topic, as well as those responding to child abuse coverage in mass media. In contrast, the phrase '*child abuse hotline'* generated results consistent with those looking to report suspected or experienced child abuse (for example, results included resources and guidance for reporting child abuse), and thus we deemed it specific enough to retain.

When trying to capture searches related to neglect, we considered phrases indicative of problematic levels of drug and alcohol use by a caregiver. We initially considered phrases ranging from '*[perpetrator] is high on drugs'* to *'[perpetrator] drinks wine'* to capture a broad spectrum of substance use. We discarded terms such as *'mom drinks wine*' since it returned numerous articles relating to "wine mom" culture and merchandise. Most of the search phrases we identified to describe neglect were too broad in nature and we removed many of them from consideration due to misclassification concerns.

Incognito searches also revealed additional perpetrators of abuse that we had not initially considered, including "parent", "parents" and "babysitter." We expanded our search phrases to include these additional perpetrators.

Step 3: Craft search term(s)

After finalizing relevant search phrases (i.e., one phrase on its own, such as ‘mom hit me’), the third step is to determine how to combine them into a search term for querying the GHT-API. A search term is defined as: (1) A single search phrase on its own, or (2) a combination of search phrases combined using “+” and entered into the GHT-API (e.g., ‘*mom hit me + mother hit me + stepmom hit me + stepmother hit me*’ are the first four search phrases in the physical abuse search term). If a search phrase by itself is commonly searched and adequately captures the construct of interest, then it may not need to be combined with other phrases and can be used as a search term on its own. However, certain phrases may be rarely searched at the desired geo-time scale (see Step 4), potentially leading to suppressed data and the inability to answer the research question. Combining related search phrases into a more comprehensive search term may be necessary to obtain sufficient search volume. Researchers should weigh the benefits of combining search phrases such that zero-values no longer hinder analysis, with the risk that the meaning of the construct might be lost. Content expertise is again helpful for making these decisions, as are incognito searches to assure that search terms lead to results that are relevant to the outcome of interest. Note that when combining search phrases, the final GHT-API result will not distinguish search volume between the individual phrases. Further, GHT-API interprets spaces between individual words as Boolean AND, and the operator “+” as Boolean OR, with AND taking priority over OR; this syntax cannot be altered with parentheses. Web Appendix 1 contains more details on how to combine search phrases into search terms for querying the GHT-API (21).

There does not appear to be a limit to the number of search phrases that can be combined into a search term, although the API is slower to return results when more than 100 phrases are included. It is possible that there is a limit to the number of words or characters, but this is not documented.

Case Study Application:

In our setting, each search phrase on its own — with the exception of *'child abuse hotline'* — did not return enough search volume to avoid significant missing data at the specified geo-temporal level. Since the number of zero-values precluded meaningful analysis, and state-weekly data was needed to answer our research question, we combined all abuse phrases within an abuse subtype (i.e., physical, emotional, sexual, neglect) into four GHT-API search terms using the Boolean OR operator "+". These phrases included common misspellings as well as various conjugations and tenses of verbs. This left us with four mutually exclusive search terms corresponding to each abuse subtype, each comprising between 700 and 2,100 individual phrases, and a fifth search term consisting only of the single search phrase *'child abuse hotline'*. Combining phrases did not compromise measurement of our construct because we were interested in measuring the overall experience of abuse and neglect, and not in specific manifestations of abuse.

Step 4: Determine feasible geographic and temporal scales

The GHT-API allows researchers to specify both geographic and temporal resolutions. Within the US, queries can be made at the national, state, or designated market area (DMA) level, where DMAs are smaller geographic regions grouped by the Nielsen Company for television and radio ratings (22). Timescales include day, week, month, and year.

When selecting a geographic and temporal resolution it is important to note that, to ensure anonymity, GHT-API suppresses values which fall below an unspecified threshold (2020 *Google Health Trends API Getting Started Guide* [unpublished document provided with API key]). The threshold value is not documented and may vary by geographic region and population size (Figure 2). Suppressed values are returned as zeros, thus any zeros may indicate no search volume at that geo-time scale or data suppression. A search term may return all zeros at the DMA-daily level (the smallest level of geo-time), but could return non-zero results for larger geographical or temporal units. Thus, to overcome a high degree of missing data, a larger geographic and/or time scale may be specified.

The search term(s) may also need to be expanded by returning to steps 1 through 3 if zero-values are common at the desired geo-time scale. The goal is to identify a geo-time unit that is large enough to minimize suppression, without hindering the ability to study the research question of interest. Search terms can be tested by making a GHT-API query at the desired geo-time scale to assess sufficiency of data. Testing can also identify whether specific search phrases have enough volume to be used as a search term alone or in combination with other phrases. Iterate Steps 3 and 4 as needed to achieve sufficient search volume at the desired geo-time scale, such that analysis is not hampered by the number of zero-values, the construct of interest can still be measured, and the geo-time scale chosen is granular enough to answer the research question.

Case Study Application:

For our application, we needed data at the state-week level (i.e., we could not shift to a larger time or geographic scale). To get an initial sense of the search volume, we queried each of our five search terms (i.e., physical abuse, emotional abuse, sexual abuse, neglect, child abuse hotline) as well as a combination term that included all search phrases except for neglect-related phrases, at the national-weekly level (Web Figure 1). We observed very low search volume for the neglect search term at the national level, which is a level larger than our desired state-level. Given this, and concerns about misclassification (see

Step 2), we excluded neglect search phrases from our analysis. We then tested the remaining four search terms at the state-week level and found missingness to be an issue for small population states, whose zero-values made up more than half of their state-week observations. Therefore, to overcome missingness, we combined all 3,484 separate abuse phrases, including *'child abuse hotline'*, into a single search term (the combination term) to capture overall child abuse for the analysis. Combining all abuse phrases into a single search term precluded us from studying individual domains of abuse (i.e., physical, emotional, and sexual). Our research question aimed to measure overall abuse experience, and this tradeoff in specificity was reasonable. Furthermore, while we intended to study neglect as part of abuse overall, this abuse domain could not be studied because of very low search volume and was subsequently dropped from the outcome of interest.

Step 5: Retrieve and average multiple samples to stabilize estimates and address missingness

GHT-API results are based on a unique-session-generated random sample which is cached for 24 hours (14). This means that the same GHT-API query returns a different scaled probability when made on a different day. Performing multiple queries (over multiple days) and thus generating different random samples, which are then averaged, stabilizes estimation of the true underlying search probability. The number of samples needed will vary depending on search volume. For example, if the search term is popular or queried on a larger geo-temporal scale, fewer samples will be needed to obtain a stable average compared to a search term which has low volume or high variability. The goal is to draw additional samples until adding more samples does not materially change the estimated average.

Extracting multiple samples can also help to address data suppression not remedied by Steps 3 and 4. Since GHT-API results are a scaled probability, missingness is a function of both how common the search term is (the numerator), as well as the total search volume (the denominator). The denominator varies according to the underlying population size for the unit of analysis. By leveraging the fact that GHT-API results that are close to, but below, the suppression threshold in one sample may fall above the threshold in another sample, averaging over multiple non-zero samples reduces suppressed geo-time data points. However, this approach requires caution because it only captures the portion of the underlying data distribution that falls above the suppression threshold, and thus the resulting estimate created by averaging non-zero samples is likely biased upwards compared with the true value if suppression were not applied.

Case Study Application:

We found that the proportion of suppressed observations across all ten samples of our final combination search term decreased with state population size (Figure 3). There was no suppression in the most populous states and substantial suppression in the least populous states (see Web Figure 2 for details on potential suppression thresholds). We decided to exclude states with more than 55% of their data missing due to concerns about validity, as we were worried about bias due to truncation.

We determined that there was decreased utility in including more than ten samples by plotting averaged estimates across five, seven and ten samples for a subset of states to determine stability of sample averages (Web Figure 3). We also chose ten samples to balance measure stability with time constraints; since the GHT-API limits users to 5,000 queries a day, and each state-week sample counts as a single query to draw 10 samples of weekly-level data for 51 states (including DC), over 190 weeks required 96,900 queries. At the 5,000 per day query limit, it took 19.4 days to gather the GHT-API data.

After dropping eight states that had more than 55% of their data missing, we calculated missingness for the remaining state-weeks: over all ten samples, search volume was missing for 4,874 of the 57,540 state-weeks (approximately 8%). After averaging over non-zero samples, only 62 of the 5,754 state-weeks remained missing (approximately 1%).

Step 6 (Optional): Normalize to account for changes in total search volume

Since returned GHT-API values are an estimated scaled probability of the search term within the specified geo-time period, changes in the returned value over time can reflect changes in either the search term of interest (the numerator) or changes in total searches (the denominator). This may be problematic if the researcher suspects that the total number of Google searches, and thus the denominator of the search probability, may be changing over the study period. Normalization has been suggested as a way to reduce changes in the outcome measure due to changes in total search volume (4). The idea is to choose a normalizing search term for which searches would be expected to remain relatively constant in absolute number (the numerator of the probability) over the study period, and then divide the search volume for the term of interest by the search volume for the normalizing term; this "cancels out" the total (changing) search volume in the denominator, such that the search volume of interest becomes relative to the number of (constant) searches for the normalizing term (See Web Appendix 2 for details).

While normalization could be useful in theory, it is impossible in practice to determine if the absolute number of searches for a selected normalizing term remains constant over the study period, as GHT-API results for any candidate normalizing term are also scaled probabilities of total searches. Thus, it is not possible to disentangle search volume for the normalizing term from total search volume and ascertain whether it is constant over time. Further, normalization may cloud interpretations as results are now relative to a potentially arbitrary search term, making comparisons between similar studies challenging. For these reasons, normalization is not recommended as standard practice, but may be explored through sensitivity analyses

Case Study Application:

Our case study was focused on changes in child abuse-related searches during the COVID-19 pandemic — a time period during which Google searches overall may have increased in line with measured increases in internet traffic (23) and use of popular websites (24).

While we hypothesized that total searches increased, and thus had a strong rationale for normalizing our data, it was difficult to identify a normalizing term that would not be affected by the pandemic. For example, we hypothesized that searches for “the” and “a” would increase proportionately with a change in total search volume. We conducted a sensitivity analysis in which we normalized by “and”, but were unable to determine whether it helped overcome the challenge introduced by potential major changes in search behavior due to the pandemic.

Discussion

We have outlined a series of steps to craft an effective GHT-API search strategy that takes into account sensitivity, specificity, and missing data due to data suppression. We demonstrated these steps and their utility in a case study examining state-level Google searches for child abuse during the COVID-19 pandemic.

There are several limitations to GHT-API data that should be considered. First, people who perform Google searches may not be representative of the target population. While Google is by far the most popular search engine, making up 94% of all worldwide online search queries in 2020 (25), those who use Google may be systematically different from those who do not. Additionally, search phrases chosen for inclusion may further affect generalizability (e.g., due to language or dialect choices). With respect to our case study, we likely captured searches from children with higher rates of internet access (which varies by income and rurality), who may have English as a first language (since we restricted our search strategy to the English language), and who are older.

Secondly, Google search volume is an indirect measure of the epidemiological quantity of interest - the prevalence or incidence of the condition or event under study. In our case study, we were interested in child abuse and neglect incidence, but our measure captured *searches* for phrases related to these constructs. We assumed that changes in search volume are indicative of changes in abuse and neglect incidence because the included search phrases were chosen based on existing literature; however, the plausibility of this assumption should be examined for the construct of interest, though this may be difficult when Google search volume because the construct is difficult to measure using traditional methods. Searches may not reflect the true level or patterning of the construct of interest if, for example, not all relevant search terms/phrases have been included or if the total number of Google searches (the denominator) is changing over time. Koutaniemi et al. found that in Finland, seasonal patterns in Google searches related to domestic abuse paralleled domestic abuse-related calls to police, suggesting that Google searches may be a reasonable proxy for measures of household violence (5). Additionally, Google Trends data for searches related to loss of smell correlated strongly (Spearman's rank coefficients between 0.633 and 0.952) with COVID-19 cases and deaths across 8 countries during December 2019 to March 2020 (26). On the other hand, an earlier version of the GHT-API called Google Flu Trends API was criticized for its inability to accurately estimate influenza-like illnesses (27–29), and studies suggested its utility lies in being combined with additional surveillance data or as part of a machine learning algorithm (30-32), rather than on its own. Similarly, GT results have been found to be more reflective of media buzz than underlying epidemiological burden for six diseases (33). Thus, studies examining the relationship between measures based on Google searches and epidemiologic measures for specific outcomes of interest will be useful to assess the validity of Google searches across contexts.

Third, missing data may be an issue. While averaging over multiple non-zero samples may help remedy missingness, the resulting estimate may be biased upwards since non-missing data points will over-represent samples that have relative volume falling above the GHT-API suppression threshold for the geo-time period. This is complicated by the fact that the suppression threshold is not released by Google and may vary by geographic and/or temporal scale. Note that because of this suppression threshold, values cannot be compared *across* geographic units. In our child abuse case study, missingness was an issue we overcame by forgoing examination of separate abuse domains, averaging across multiple samples, and excluding small states with high rates of missingness.

Lastly, returned values from the GHT-API are a probability so it is difficult to disentangle trends in the numerator (usually the quantity of interest) from trends in the denominator (total search volume). Therefore, it is important that findings are interpreted with respect to proportionate search volume and not to the underlying construct of interest.

Despite its limitations, the GHT-API holds promise for the field of public health and could provide valuable and timely insights in many scenarios. GHT-API data has the potential to provide extremely early signals for certain phenomena for which other surveillance or reporting data are delayed, and for health outcomes that are difficult to measure by traditional mechanisms. The GHT-API has been used to gauge public concern around the Flint, Michigan water crisis (7), complement dengue surveillance activities in Brazil (30), and examine child abuse during an economic downturn (4). For our case study [under review; citation to be added], the GHT-API served as a timely resource for measuring changes in child abuse searches during a period when measurements based on traditional surveillance systems likely underestimated child abuse cases.

Conclusion

We provided a methodological framework alongside best practices for utilizing the GHT-API to study epidemiologically relevant constructs of interest. This framework systematically crafts a successful search strategy to accurately measure the construct of interest, optimizes sensitivity and specificity, improves estimate stability through multiple samples, and overcomes missingness resulting from insufficient volumes. Due to limitations of the GHT-API data, careful interpretation of results is advised.

References

1. Google Trends API Request form. Accessed April 1, 2020. https://docs.google.com/forms/d/e/1FAIpQLSenHdGiGl1YF-7rVDDmmulN8R-ra9MnGLLs7gIIaAX9VHPdPg/viewform

2. Google Trends. Google. https://trends.google.com/trends/?geo=US. Accessed July 17, 2021.

3. Catillon M, Mandl KD. Scrutiny for Child Abuse and Neglect During the COVID-19 Pandemic [preprint]. *medRxiv*. 2020. (doi:10.1101/2020.10.12.20210997). Accessed March 28, 2021.

4. Stephens-Davidowitz S. Unreported Victims of an Economic Downturn. Seth Stephens-Davidowitz website. http://sethsd.com/research. Published online July 26, 2014. Accessed August 24, 2020.

5. Koutaniemi EM, Einiö E. Seasonal variation in seeking help for domestic violence based on Google search data and Finnish police calls in 2017. *Scand J Public Health*. 2021;49(3):254-259.

6. Stephens-Davidowitz S. Essays Using Google Data [dissertation]. Cambridge, MA: Harvard University; 2013.

7. Matsa, KE, Mitchell, A, Stocking, G. Searching for News: The Flint water crisis. Pew Research Center. https://www.journalism.org/essay/searching-for-news/. Published April 27, 2017. Accessed May 27, 2020.

8. Prata N, Weidert K, Zepecki A, et al. Using Application Programming Interfaces (APIs) to Access Google Data and Gain Insights Into Searches on Birth Control in Louisiana and

Mississippi, 2014-2018: Infoveillance Study. *J Med Internet Res*. 2021;23(7):E25923.

9. Helft M. Google Uses Searches to Track Flu's Spread. *The New York Times*. https://www.nytimes.com/2008/11/12/technology/internet/12flu.html. Published November 11, 2008. Accessed May 27, 2021.
10. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012-1014.
11. Nsoesie EO, Oladeji O, Abah ASA, Ndeffo-Mbah, ML. Forecasting influenza-like illness trends in Cameroon using Google Search Data. *Sci Rep*. 2021;11(1):E6713.
12. Stocking G, Matsa KE. Using Google Trends data for research? Here are 6 questions to ask. Medium - Pew Research Center. Published April 27, 2017. https://medium.com/@pewresearch/using-google-trends-data-for-research-here-are-6-questions-to-ask-a7097f5fb526. Accessed May 25, 2021.
13. FAQ about Google Trends data - Trends Help. Google. https://support.google.com/trends/answer/4365533?hl=en. Updated 2021. Accessed May 25, 2021.
14. Matsa KE, Mitchell A, Stocking G. Methodology. Pew Research Center's Journalism Project. https://www.journalism.org/2017/04/27/google-flint-methodology/. Published April 27, 2017. Accessed August 24, 2020.
15. Zepecki A, Guendelman S, DeNero J, Prata N. Using Application Programming Interfaces to Access Google Data for Health Research: Protocol for a Methodological Framework. *JMIR Res Protoc*. 2020;9(7):E16543.

16. Raubenheimer JE. Google Trends Extraction Tool for Google Trends Extended for Health data. *Software Impacts*. 2021;8: E100060.

17. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill*. 2019;5(2):E13439.

18. Riddell CA, Farkas K, Neumann K, Santaularia NJ, Ahern J, Mason SM. US shelter in place policies and child abuse Google search volume during the COVID-19 pandemic. *Preventive Medicine*. 2022; https://doi.org/10.1016/j.ypmed.2022.107215.

19. About the CDC-Kaiser ACE Study. Centers for Disease Control and Prevention. https://www.cdc.gov/violenceprevention/aces/about.html. Updated April 6, 2021. Accessed September 11, 2021.

20. Pennebaker JW, Susam JR. Disclosure of traumas and psychosomatic processes. *Soc Sci Med*. 1988;26(3):327-32

21. Search tips for Trends. Google Trends Help. https://support.google.com/trends/answer/4359582?hl=en. Updated 2021. Accessed June 25, 2020.

22. DMA Regions. Nielsen. https://www.nielsen.com/us/en/contact-us/intl-campaigns/dma-maps. Accessed July 18, 2021.

23. COVID-19: How Cable's Internet Networks Are Performing. ncta. https://www.ncta.com/COVIDdashboard. Accessed May 27, 2021.

24. Koeze E, Popper N. The Virus Changed the Way We Internet. *The New York Times*. https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html. Published April 7, 2020. Accessed May 27, 2021.

25. Statista. Online Search Usage. https://www.statista.com/study/15884/search-engine-usage-statista-dossier/. Published 2021. Accessed May 25, 2021.

26. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell–related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol*. 2020;10(7):839-847.

27. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: Traps in Big Fata Analysis. *Science*. 2014;343(6176):1203-1205.

28. Pollett S, Boscardin WJ, Azziz-Baumgartner E, et al. Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection. *Clin Infect Dis*. 2017;64(1):34-41.

29. Lohr S. Google Flu Trends: The Limits of Big Data. Bits Blog. https://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/. Published March 28, 2014. Accessed August 25, 2021.

30. Romero-Alvarez D, Parikh N, Osthus D, et al. Google Health Trends performance reflecting dengue incidence for the Brazilian states. *BMC Infect Dis*. 2020;20(1):252.

31. Klembczyk JJ, Jalalpour M, Levin S, et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. *J Med Internet Res*. 2016;18(6):E175.

32. Dugas AF, Jalalpour M, Gel Y, et al. Influenza Forecasting with Google Flu Trends. *PLoS ONE*. 2013;8(2):E56176.

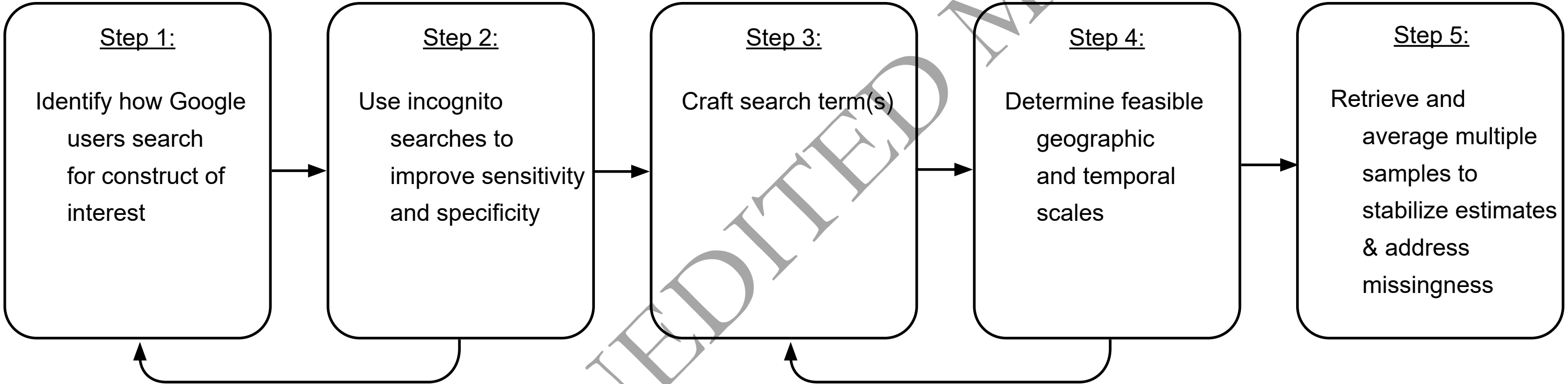
33. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health*. 2017;7(3):185-189.

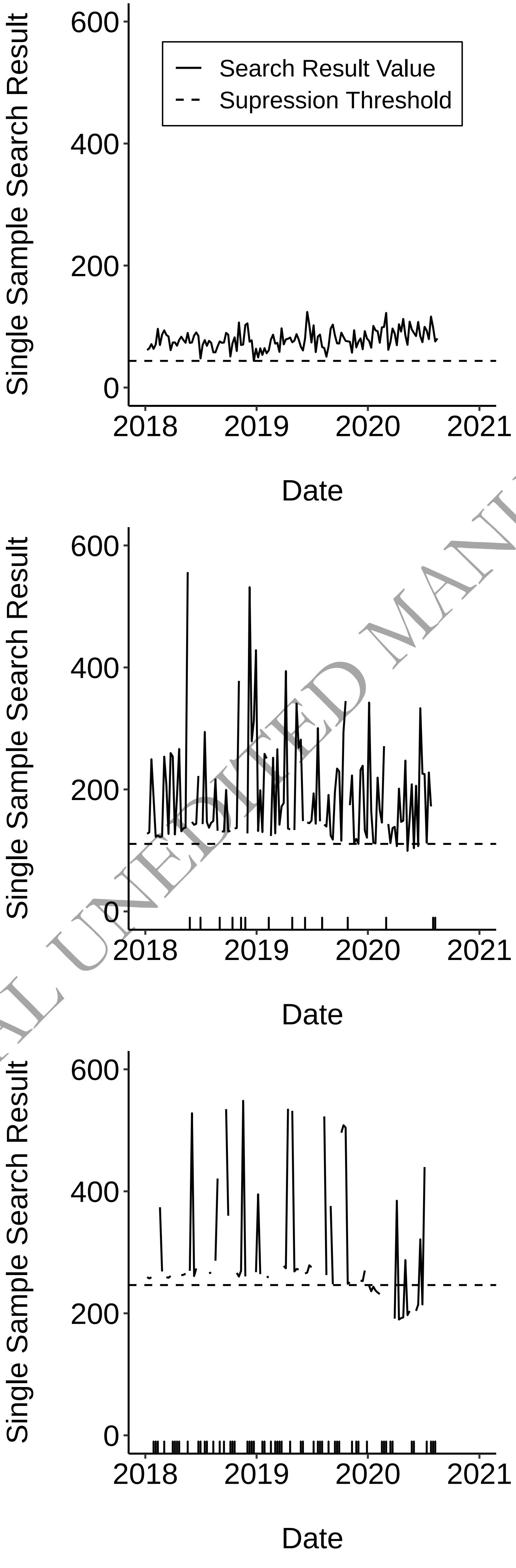
Figures

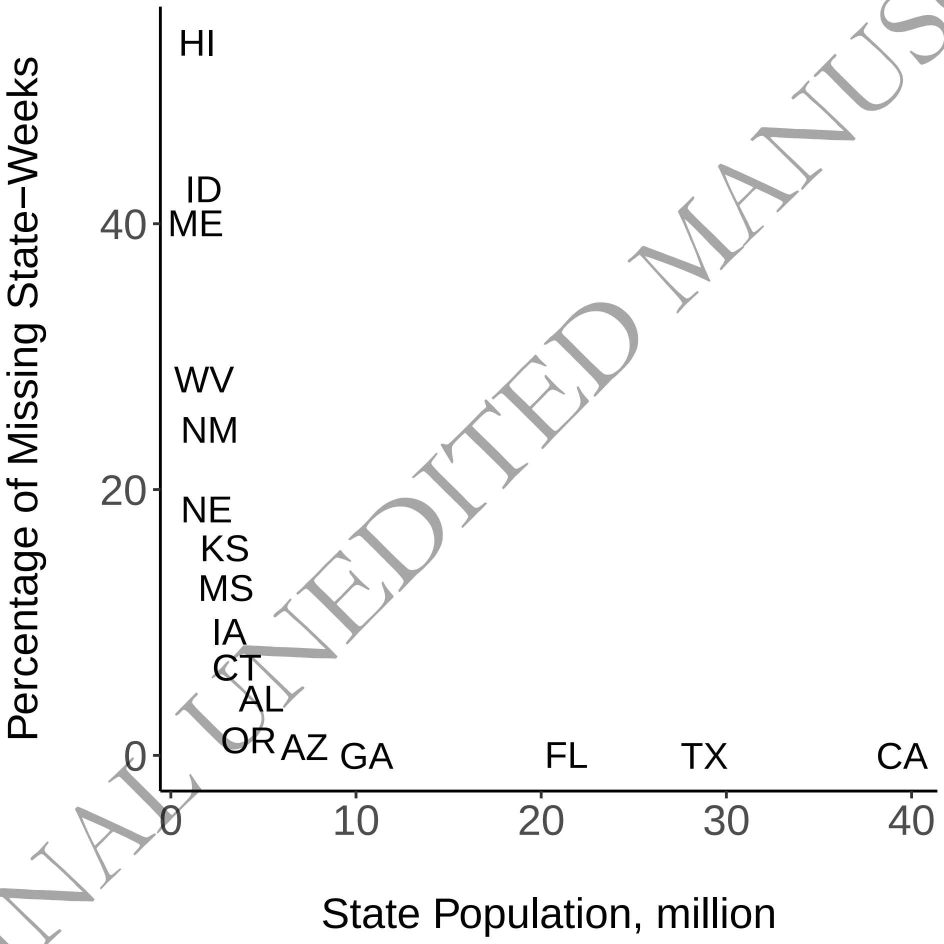
Figure 1: Google Health Trends API search strategy flowchart

Figure 2: Potential suppression thresholds for the child abuse case study in California, Iowa, and New Hampshire, January 2018 – August 2020. Potential suppression thresholds, denoted by dashed lines, represent a threshold below which zero-values were always returned. These appear to vary by geographic unit and population size with large states like California (first panel) returning no zero-values, but smaller states such as Iowa and New Hampshire (second and third panel respectively) possessing higher thresholds and more missing data. Weekly search volumes are based on a single sample of the combination term; dashed lines representing the potential suppression threshold were calculated as the minimum non-zero value returned for that state pre-2020; missing data points (equivalent to zero-value data points) are indicated by the tick marks directly above the x axis. It is clear the number of zero-values increases as the potential suppression threshold increases: Out of a total of 137 possible state-weeks, 0% (0/137) of weeks are missing for California whose minimum returned value is 25.4, 10.2% (14/137) of weeks are missing for Iowa whose minimum returned value is 110.7, and 40.9% (56/137) of weeks are missing for New Hampshire whose minimum returned value is 243.1.

Figure 3: Percentage of state-week samples suppressed by state population, across all ten samples. Percentage of suppressed state-weeks was calculated for each state as the quotient of the number of state-weeks missing over all ten samples divided by the total number of distinct state-weeks in the study (1370); State populations were obtained from the American Community Survey for the years 2018 and 2019 and averaged. States with more than 55% of their data missing are excluded from this figure.


Step 1:
Identify how Google users search for construct of interest
Step 2:
Use incognito searches to improve sensitivity and specificity
Step 3:
Craft search term(s)
Step 4:
Determine feasible geographic and temporal scales
Step 5:
Retrieve and average multiple samples to stabilize estimates & address missingness


Search Result Value
Supression Threshold
Single Sample Search Result
0
200
400
600
2018
2019
2020
2021
Date
Single Sample Search Result
0
200
400
600
2018
2019
2020
2021
Date
Single Sample Search Result
0
200
400
600
2018
2019
2020
2021
Date


Percentage of Missing State–Weeks
State Population, million
0
20
40
0
10
20
30
40
HI
ID
ME
WV
NM
NE
KS
MS
IA
CT
AL
OR
AZ
GA
FL
TX
CA