

Using Google Trends Data to Learn More About Survey Participation

Gummer, Tobias; Oehrlein, Anne-Sophie

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gummer, T., & Oehrlein, A.-S. (2022). Using Google Trends Data to Learn More About Survey Participation. *Social Science Computer Review*, OnlineFirst, 1-18. <https://doi.org/10.1177/08944393221129179>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see:
<https://creativecommons.org/licenses/by-nc/4.0>

Using Google Trends Data to Learn More About Survey Participation

Social Science Computer Review
2022, Vol. 0(0) 1–18

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/08944393221129179

journals.sagepub.com/home/ssc



Tobias Gummer^{1,2}  and Anne-Sophie Oehrlein¹ 

Abstract

As response rates continue to decline, the need to learn more about the survey participation process remains an important task for survey researchers. Search engine data may be one possible source for learning about what information some potential respondents are looking up about a survey when they are making a participation decision. In the present study, we explored the potential of search engine data for learning about survey participation and how it can inform survey design decisions. We drew on freely available Google Trends (GT) data to learn about the use of Google Search with respect to our case study: participation in the Family Research and Demographic Analysis (FReDA) panel survey. Our results showed that some potential respondents were using Google Search to gather information on the FReDA survey. We also showed that the additional data obtained via GT can help survey researchers to discover topics of interest to respondents and geographically stratified search patterns. Moreover, we introduced different approaches for obtaining data via GT, discussed the challenges that come with these data, and closed with practical recommendations on how survey researchers might utilize GT data to learn about survey participation.

Keywords

survey participation, nonresponse, google trends data, auxiliary data, digital trace data, search engine data

Introduction

The trend of decreasing response rates continues (Brick & Williams, 2013; de Leeuw & de Heer, 2002; de Leeuw et al., 2018), and in many countries, response rates have reached dramatic lows. For instance, in a recent comparison between General Social Surveys, Wolf et al. (2021) reported that the response rate in the U.S. General Social Survey (US-GGS) declined from 70.1% in 2002 to 59.5% in 2018, whereas in the German General Social Survey (ALLBUS), it dropped from 47.3%

¹GESIS Leibniz Institute for the Social Sciences, Mannheim, Germany

²University of Mannheim, School of Social Sciences, Mannheim, Germany

Corresponding Author:

Tobias Gummer, GESIS Leibniz Institute for the Social Sciences, B6, 4-5, Mannheim, 68159, Germany.

Email: tobias.gummer@gesis.org

in 2002 to a low of 32.3% in 2018. With respect to the data collection of the European Social Survey (ESS) in Germany, these authors also reported a decline in the response rate from 51.7% in 2002 to 27.6% in 2018. Even though approaches like nonresponse follow-up surveys (e.g., [Olson et al., 2011](#); [Vandeplas et al., 2015](#)) and the analysis of auxiliary data such as geospatial data (e.g., [Comesse, 2020](#); [West et al., 2015](#)) have aimed at better understanding survey participation, the continuing decline in response rates illustrates that our knowledge of the causes of nonresponse is still too limited to counteract it. Developing the means to stabilize or even increase participation in surveys requires a proper understanding of the participation process. [Tourangeau \(2017\)](#) has suggested that "... it is much harder to fix a problem when we don't really know what the problem is" (p. 812). Thus, while advances in survey design have been made, especially the shift towards self-administered and/or mixed-modes (e.g., [de Leeuw, 2005](#); [de Leeuw, 2018](#); [Luijckx et al., 2021](#)) and an increased use of adaptive and responsive survey design ([Groves & Heeringa, 2006](#); [Schouten et al., 2018](#); [Tourangeau et al., 2017](#); [Wagner, 2008](#)), these developments have not been enough to stop the decline in survey participation. Hence, more knowledge about survey participation and nonresponse is required to develop meaningful tools to enhance participation in surveys.

With increasing access to the internet and to mobile devices in countries around the world (e.g., [Poushter, 2016](#); [Taylor & Silver, 2019](#)), using the internet to gather information became common practice. When receiving information about a survey and an invitation to participate, it is possible that potential respondents search the internet with respect to this contact attempt. Reasonable motives for searches might be to gather additional information about the survey, verify the reputation and credibility of the fieldwork institute and survey sponsors, or to clarify questions regarding survey participation. Thus, researchers may be able to tap into information about the participation process that is stored by search engines such as Google Search.

To the best of our knowledge, only the study by [Fang et al. \(2021\)](#) explored the use of search engine data to learn about survey participation. These authors used search engine data to predict participation in the Dutch Health Surveys. In their study, they used Google Trends (GT) to capture the salience of societal issues—which they considered relevant for survey participation—during the fieldwork phase of the survey (i.e., disease outbreaks, privacy concerns, public outdoor engagement, and terrorism salience). While [Fang et al. \(2021\)](#) promoted a novel approach to using GT data to capture societal trends that might impact survey participation, to our knowledge, no prior research has explored the use of search engine data to learn about what information potential respondents search about a survey. This gap in research is unfortunate since we recognize several areas of application and use cases in which these data might provide valuable insights. First, knowing what information potential respondents might be searching for after being invited to participate in a survey can help researchers to design better information materials, websites, and invitation letters. If respondents receive survey invitations and information materials that do not contain information, they perceive as necessary, it can be reasonably assumed that some might search the internet. This knowledge can be used to design websites that have the most important information readily accessible (e.g., in dedicated FAQs on the landing page). Second, search engine data can not only help to improve the information provided for respondents, but also the survey design itself. The geographical stratification, timing, or sentiments of searches also might provide hints about issues with a survey. For instance, search patterns might reveal whether concerns about the credibility of the survey sponsor or fieldwork institute exist, whether data privacy concerns are prevalent, or whether design characteristics such as prepaid incentives confused potential respondents. Finally, in addition to these areas of the application of search engine data, the unfortunate lack of research is evident in a central question whose answer currently is not backed by empirical evidence: Do potential respondents actually use

the internet to search for information about a survey? In other words, how important is the provision of accessible and comprehensive information about a survey on dedicated survey websites?

Two different approaches are available for using search engine data to improve surveys: during the fieldwork or after the fieldwork is concluded. As we describe below, with respect to Google Search—depending on which approach is used—certain limitations may apply to the data (e.g., the granularity and detail of the data, availability of supplemental information). Due to the lack of research, the usability of search engine data for survey research is unknown. The goal of the first approach is to gain insights into the participation process and potential issues so to adjust the survey design and information material already being used during data collection. These adjustments could be made ad hoc or between waves of reminders (e.g., to provide important information more prominently in reminder letters) to provide improvements for those who have not participated yet. The goal of the second approach is to gain insights into the data collection process for quality assessment or to improve future data collection projects (in panel surveys, these could be future panel waves).

In the present study, we explored the potential of search engine data for learning about survey participation and for informing survey design decisions. Our focus was on Google Search, for which [Timoneda and Wibbels \(2021\)](#) estimated a market share of 92% in Europe, Asia, and North and South America. Data on this search engine is publicly available via the free online tool GT. As described in more detail below, the information provided by GT is drawn from different samples of Google Search data that differ in their reliability to detect rare events ([Rovetta, 2021](#)). We explored the use of GT data for increasing our knowledge about survey participation with respect to the practical implications discussed above. Thus, our research had three major goals: First, we wanted to establish whether some potential respondents used search engines to find information about a survey on the internet. Second, we investigated whether GT data provided a hint as to which information some potential respondents looked up with respect to a survey. We expected that this information might provide some insights into the reasons for nonresponse or help to identify topics that are important to potential respondents who use Google Search. Third, our study also is intended to serve as a first introduction for survey researchers on how to use GT in survey research. To this end, we have devoted special attention to how to use the data obtained from different samples of GT data.

The present study is structured as follows. In the next section we introduce GT data as a form of search engine data that is publicly available to researchers, and we describe how it can be collected to learn about survey participation. Then, we introduce the Family Research and Demographic Analysis (FReDA) panel survey as our case study, which we used to explore the use of search engine data for survey design purposes. Next, we describe the methods we employed in our case study. After presenting our results, we offer concluding remarks, practical recommendations for applying the proposed approach, and an outlook for future research opportunities.

Google Trends Data

Google Trends is a freely accessible service based on Google's search engine data. It enables users to determine how search queries for a specific term (e.g., *freda survey*) developed during a specific timeframe (e.g., *January to June 2021*) and for a specific geographic region (e.g., *Germany*).

The data provided by GT are relative search volumes (RSVs), which are a share computed as the number of search queries for a term divided by the total number of Google Search queries—both are conditional on a selected timeframe and geographic region ([Choi & Varian, 2012](#); [Stephens-Davidowitz & Varian, 2015](#); [Timoneda & Wibbels, 2021](#)). Then, this share is

normalized to values between 0 and 100, where 100 is the maximum value of the search term in the selected timeframe. The RSV can be interpreted accordingly: for example, a value of 50 indicates that the search volume of the specified term at a certain point in time was 50% less than the search volume of this term at the point of maximum search queries in the selected timeframe.

In addition to trend data for a timeframe specified in the request, GT provides geographical information on the distribution of RSVs across lower-level geographical units. For instance, when requesting search data for Germany, GT also provides RSVs for the federal states of Germany. Also, the geographical and trend data are further supplemented with additional search data: GT lists related searches including the terms most frequently queried in combination with the specified search. For instance, related searches to the term “*freda*” could be “*freda survey*” or “*freda invitation*.” We argue that data on related searches can provide additional information for researchers to discover the particular topics in which some of their potential respondents were showing an interest.

The specified timeframe and the timing of the query have substantial implications for the quality of data that can be obtained. Searching GT data for a timeframe within 7 days or less from the day of the query will draw on a sample of real-time data, which means data are available on an hourly granularity. *Real-time data* are the data survey researchers likely will collect if they *ex-ante* design GT data collection to complement their field period. Gathering insights via GT during the field period enables survey researchers to make adjustments during their fieldwork (e.g., change the wording in the next reminder or add information to the survey website).

Searching GT data for a timeframe that is specified to cover a time longer than 7 days from the day of a query will draw on a sample of so called *non-realtime data*. These long-term data are drawn from a sample of search queries going back to 2004. For this sample, [Rovetta \(2021\)](#) has highlighted that the reliability of the data for detecting rare events may vary between daily samples, which is likely the result of the low inclusion probabilities for rare search queries in a daily sample drawn from the historic Google data set. Accordingly, [Rovetta \(2021\)](#) has proposed—similar to [Stephens-Davidowitz and Varian’s \(2015\)](#) suggestion—drawing on repeated samples of GT data to increase the chance of detecting rare events and to average out outliers in the data (i.e., combing data from several GT queries). In contrast to real-time data, non-realtime data retrieved for periods that are more than 7 days in the past do not provide hourly data, instead only daily, weekly, or even monthly data, depending on how far the search goes back in time ([Timoneda & Wibbels, 2021](#)).

Non-realtime data are the data survey researchers collect if they *ex-post* design GT data collection to gain a better understanding of what happened during their field period. Collecting data after fieldwork is complete enables survey researchers to make adjustments to the survey protocols of future surveys or future waves of panels (e.g., change the initial invitation letters to convey specific information).

A study by [Zhu et al. \(2012\)](#) examined the validity of GT data by comparing GT data on aspects of personal life in the city to survey data that featured questions with similar topics. Their results indicated a considerable variation in how well search engine data and survey data correlated. With respect to 10 aspects of personal life, they reported correlations for four aspects, ranging from 0.22 to 0.60. Thus, the [Zhu et al. \(2012\)](#) study showed that the validity of GT data varied dependent on the topic of interest (i.e., the search term).

For readers interested in more extensive and/or general introductions to GT data, we recommend the work of [Timoneda and Wibbels \(2021\)](#), [Stephens-Davidowitz and Varian \(2015\)](#), and [Choi and Varian \(2012\)](#). [Rovetta \(2021\)](#) complements these introductions to GT data by focusing on its reliability and by highlighting that non-realtime data are dependent on the sampling process. In addition, [Zhu et al. \(2012\)](#) have provided some rare insights into the validity of GT data. For

those interested in studies that have used GT data in substantive research projects, we suggest a review by [Jun et al. \(2018\)](#).

Data and Methods

To investigate our research questions and illustrate the use of GT data for learning about survey participation, we relied on a case study—the FReDA survey—which we briefly introduce in this section. Following this description, we detail the supplemental collection of real-time and non-realtime data via GT.

Family Research and Demographic Analysis

Family Research and Demographic Analysis (FReDA) is a survey data infrastructure in Germany established in 2020 that is operated in close cooperation by the Federal Institute for Population Research (BiB), GESIS – Leibniz Institute for the Social Sciences, and the pairfam consortium. Family Research and Demographic Analysis was designed as a panel survey covering a wide range of interdisciplinary family research topics, and since the panel became fully operational in 2022, respondents and their partners have been interviewed via self-administered mixed-mode (mail and web) interviews on a bi-annual basis.

In 2020, a sample was drawn randomly for the FReDA panel survey from German municipalities' registers. This gross sample covered persons living in Germany aged between 18 and 49 years. Recruitment for FReDA started in early 2021 with a dedicated recruitment interview that respondents spent approximately 10 minutes to complete. On April 7, 2021 a total of 54,845 invitations were sent to a tranche of the gross sample.¹ These deliveries included an invitation letter, respondent information (e.g., data privacy information), and a pre-paid incentive of 5€. A randomly selected subsample of respondents also received a printed paper-based questionnaire. For more information on FReDA and its methodology, see [Schneider et al. \(2021\)](#) and [Gummer et al. \(2020\)](#).

We selected FReDA as case study because it was newly recruited and established what enabled us to monitor a situation in which potential respondents learned about a survey for the first time. The delivery of the invitation letters marked the first day that a broader audience outside the scientific community (i.e., the general population) became aware of FReDA. Previous to that delivery, FReDA had been mentioned mainly on the communication channels of the project partners. Moreover, as recruitment was still pending prior to our work on the present study, we were able to complement the field period of the recruitment survey with our GT data collection. In addition, FReDA resembles a conservative test case, since the term *FReDA* not only represents a scientific acronym, but also is well-known as a common name of females, families, or places. Thus, we investigated whether sending invitation letters had a measurable impact on search patterns for a term already searched. This particularity became obvious when comparing Google Search results for FReDA (21,600,000 results) with those of other large German studies such as the Socio-Economic Panel (SOEP) (7,840,000 results) or the German General Social Survey (ALLBUS) (2,520,000 results).

Collecting Real-Time GT Data

To investigate whether Google Search was used by some potential respondents to gather information on FReDA, we issued GT queries for the terms “*freda*” and “*freda survey*” (in German, we searched for this term using “*freda umfrage*”). The search was based on the default settings *All*

Categories and *Web Search* with *Germany* selected as the search region. We retrieved real-time GT data on April 12, 2021 during the field period of the survey. We customized the timeframe from April 5, 2021 to April 12, 2021. This period included the two days before the actual mailing of the invitation letters, which enabled us to explore our first research question about whether the mailing had led to an increase in Google Search queries for “*freda*” and “*freda survey*.”

Retrieving GT data already on April 12 enabled us to obtain real-time data with RSVs of hourly granularity. Our GT query also returned a list of related queries and geographical information. We obtained RSVs on a federal state level that provided us with an overview of how Google Search queries for “*freda*” conducted in Germany between April 5 and April 12, 2021 were distributed among the 16 federal states. In our analyses, the federal state of Saxony-Anhalt represented the reference RSV of 100.

Collecting Non-Realtime GT Data

In contrast to real-time data, search queries for non-realtime data can be issued any day that is more than 7 days away from the selected timeframe. Due to Google caching data only once a day, it is not possible to draw more than one different sample with the same specifications per day (Stephens-Davidowitz & Varian, 2015). Thus, the outcome of a search query depends on which day it is issued. To investigate this dependency, we decided to issue search queries on different days and draw multiple samples of non-realtime GT data—one sample per day. We call this approach *resampling*. In total, we drew 30 samples of the search terms “*freda*” and “*freda survey*,” respectively, between Monday, May 31, 2021 and Tuesday, June 30, 2021. For each query, we used the same settings as for the real-time sample drawn on April 12, 2021. Since the timeframe between the resampling and the queried period was less than 9 months, the 30 samples resulted in daily data, and thus entailed eight datapoints each—one for every day between April 5, 2021 and April 12, 2021.

Analytical Method

To explore our first research question regarding whether potential respondents used Google Search to find information about FReDA, we relied on the real-time sample to detect whether search queries, with respect to our survey, increased after we sent our invitation letters on April 7, 2021.

With respect to our second research question, we again drew on the real-time sample to investigate whether additional insights concerning what potential respondents are interested in prior to participation could be gained by relying on GT data. Thus, we explored the supplemental information our search queries produced.

To investigate our third research question concerning the usability of the GT data from different samples, we compared different approaches of collecting GT data: (1) drawing on the real-time sample, (2) drawing on one non-realtime sample, (3) stepwise combining data from 30 non-realtime samples (i.e., one month of resampling), and (4) combining data from seven consecutive non-realtime samples (i.e., one week of resampling). We decided to include approach (4) as a scenario in which researchers acknowledged the importance of resampling but were not willing to invest too much effort in data collection. We considered this a more practical scenario in which a problem is known but only addressed with some naïve rule of thumb solution (in this case selecting a week as a plausible period for sampling GT data).

Due to its high granularity, we defined the data from the real-time sample as the reference standard and analyzed whether the search query distribution was similar for approaches (2), (3), and (4). For this purpose, we made three sets of comparisons: (a) real-time sample versus each

single one of the 30 samples drawn between May 31, 2021 and June 30, 2021, (b) real-time sample versus stepwise cumulated data from the 30 samples, and (c) real-time sample versus cumulated data from a week of resampling.

The cumulated data for (b) were generated by stepwise calculating an average across all the non-realtime samples drawn (i.e., first across one sample, then across two samples, etc.). Accordingly, with this strategy, we generated 30 averaged samples, each containing one more sample than the last (e.g., the second sample included the daily samples of the first two days, etc.). Similarly, we obtained the average of one week of resampling (c) by using data from seven consecutive days, again cumulating the data accordingly. For the comparison, we computed all possible 7-days timeframes (e.g., day 1 to day 7, day 2 to day 8, day 3 to day 9) within our observation period (i.e., rolling weeks), which limited the number of these averaged samples to 24.

Before cumulating the data, we harmonized the data to ensure comparability between the non-realtime and real-time samples and to ensure easy interpretability. First, the real-time data were available in hourly granularity, whereas the non-realtime data were available only in daily granularity. Thus, we had to transform the real-time data from hourly to daily granularity. Therefore, we calculated the mean of every day and identified the day with the maximum average RSV. Following Google's approach, this RSV was then set to 100 and the RSVs of all the other days were rescaled accordingly. Second, since RSVs can be hard to interpret and the comparability of distributions of RSVs between different samples can be an issue, we decided to convert the RSVs of all samples to the share of search queries in an observation period (SQOP). We computed the $SQOP_i$ of day i in an observation period from $t = 0$ to $t = T$ as follows:

$$SQOP_i = \frac{RSV_i}{\sum_{t=0}^{t=T} RSV_t}$$

SQOPs can be interpreted as the daily share of search queries in an observation period and can take values between 0 and 1. In addition, this transformation eases the comparison of the distributions of search queries between different samples.

We used the nonparametric Kolmogorov–Smirnov test to investigate differences in the distributions of SQOPs between samples. The two-sided Kolmogorov–Smirnov test is generally used to test whether two samples stem from the same distribution. Thus, the cumulative distribution functions (CDF) of the two samples for a variable x are compared to identify the Kolmogorov–Smirnov test statistic D , which represents the CDFs' maximum difference in value:

$$D = \max|CDF_1(x_1) - CDF_2(x_2)|$$

In our analyses, we drew on D as a measure of the differences between two samples' distribution in SQOPs. D can take values in the range of 0–1. Whereas a value of 0 indicates that the distributions are similar for the two samples, the higher the value of D , the more dissimilar the distributions are. We obtained the highest D statistic of 0.875 for the search term “freda” on day 7 and 13 of the daily resampling strategy. For these days, to match the distribution of the real-time sample, one would have to re-distribute the two samples by 65.3 percentage points each.

We used R and the `ks.test` function to calculate the test statistic D when comparing the real-time sample with each of the samples outlined above.

Results

In this section, we discuss the results of using GT data to learn about FReDA participation by referring to two different kinds of samples: real-time data and non-realtime data. We used real-time

data to answer our first research question about whether potential respondents used Google Search to gather information about the FReDA survey. We also used these data to explore our second research question concerning whether GT data can provide insights into the information that potential respondents perceive as necessary to make a participation decision. Our third research goal, providing survey researchers with insights into how to use GT data, is addressed in both the following subsections—“Real-time data” and “Non-realtime data.”

Real-Time Data

Figure 1 shows how the search terms “freda” and “freda survey” were used between April 5, 2021 and April 12, 2021. The vertical line represents the date when all the invitation letters were sent. Prior to sending the invitation letters, we found a comparatively low frequency of searches for both terms. Moreover, “freda survey” was an especially unsearched term. Roughly one day after sending the invitation letters, our data showed a spike in searches, and from then on a higher degree of use of both search terms. This delay was in line with the delivery times of the German postal service used for FReDA. Regarding our first research question, we interpret this finding as evidence that some potential respondents used Google Search to gather information about FReDA. Moreover, with respect to learning about survey participation, we found that some potential respondents were searching for information on FReDA within 2–4 days of the date we sent our invitation letters, which we interpret as an indication of their participation decision process occurring during the same period. After this time, the RSV declined again.

Figure 2 shows additional geographical information available via GT. The data we collected included data for all German federal states, with Saxony-Anhalt representing the maximum. Again, the data obtained refer to the period between April 5, 2021 and April 12, 2021, and were transformed into SQOPs. In Figure 2, we plotted this information on a geographical map of Germany—for illustrative purposes, we plotted the AAPOR Response Rate 2 (RR2; AAPOR, 2016) for the same timeframe (April 5, 2021 to April 12, 2021)² as well as for the entire field

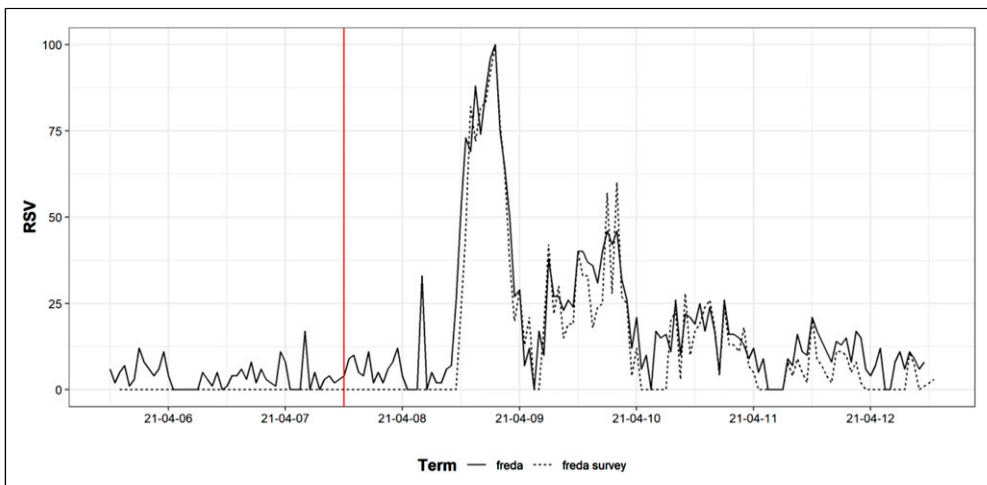


Figure 1. Hourly relative search volume (RSV) of Google searches for the term “freda” and “freda survey”, respectively. The vertical line approximately indicates the dispatch of the first invitation letters. Dates in YY-MM-DD format.

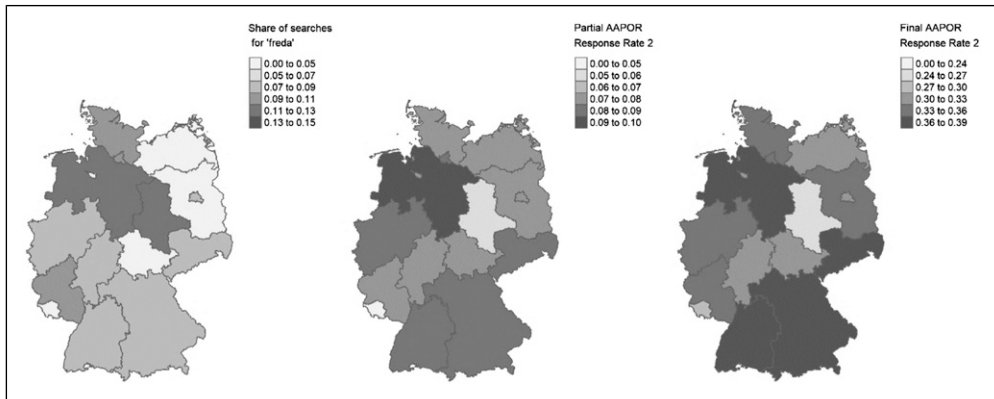


Figure 2. Share of Google searches for the term “freda” in Germany between April 5 and April 12, 2021 (left) compared to partial AAPOR Response Rate 2 measured on April 12, 2021, only 5 days after mailing the initial invitation letter (center), and final AAPOR Response Rate 2 measured at the end of the field period on June 29, 2021.

period (April 5, 2021 to June 29, 2021). We decided to use RR2 because it included partial interviews.

Relating geographically stratified SQOPs with information about survey performance may provide valuable insights, for example, regarding learning about locations in which an especially high demand for information exists, which is reflected by more frequent search behavior. In the present example, we found geographical patterns for “freda” searches and response rates. For instance, the share of searches and response rates were higher in the southern and northwestern federal states than in the northeast. Most noticeable, however, were the response rates of the two states with the highest share of searches: Saxony-Anhalt (11.36% of all searches) and Lower Saxony (11.25% of all searches). Five days after mailing the initial invitation letter, Lower Saxony showed the highest response rate of 9.06%. In contrast, at the same point in time, Saxony-Anhalt represented one of the lowest response rates with only 5.16%. As shown in graph three of Figure 2, these patterns were still valid at the end of the field period on June 29, 2021. The correlation between the SQOPs of “freda” and the RR2 observed in the 16 German federal states was 0.33 for the partial RR2 (during fieldwork) and 0.29 for the final RR2 (fieldwork completed). However, we would like to caution that due to the small sample size of 16 federal states, the correlation coefficients are not statistically significant.

Google Trends data include other information that is potentially useful for survey researchers, for example, a list of related searches that can be used to learn about what additional information was searched with respect to a survey. In comparison to using Google Search’s autocomplete function (i.e., entering “freda” and looking for related terms), GT can provide data on what people actually searched for. Table 1 illustrates this difference by comparing the outcomes of searches by five different “personas” and the information that was acquired using GT. Our personas were different people in different places in Germany with different Google accounts. They had been instructed to search for “freda” on Google Search and document the section “Related searches” usually delivered at the bottom of a Google Search result page. The last row of Table 1 shows the data obtained via GT. Focusing on the data obtained via GT, we found that the 5€ prepaid incentive (sent with the invitation letter as described previously) and the credibility of the survey were the most important topics for people going beyond a more general search for “freda survey.” These findings are in line with previous research highlighting the importance of prepaid incentives for

Table 1. Related queries of Google searches for the term “freda” from four different locations in Germany on July 2, 2021, compared to related search terms obtained from a Google Trends query for “freda” on April 12, 2021.

Google Search	Related search terms
Persona 1	FReDA infas 5 euro; FReDA survey obligatory; Infas FReDA; FReDA project; FReDA study experience; FReDA reliable; FReDA panel; FReDA questionnaire
Persona 2	Freda prename; FReDA study reliable; Freda name, FReDA GESIS; FReDA 5 euro; FReDA study experience; Infas FReDA; FReDA project
Persona 3	FReDA infas 5 euro; FReDA project; Infas FReDA; FReDA survey obligatory; FReDA study experience; FReDA panel; FReDA BiB
Persona 4	Infas FReDA; FReDA study experience; FReDA project; FReDA survey obligatory; FReDA infas 5 euro; FReDA questionnaire; FReDA panel; FReDA BiB
Google Trends	freda survey; freda survey 5 euro; freda survey reliable

Note. All related search terms were translated to English; GESIS and BiB are the names of two of the organizations responsible for FReDA; infas is the name of the institute responsible for the field work of FReDA.

surveys in the German context (e.g., [Gummer et al., 2022](#); [Pforr et al., 2015](#)) and the credibility of a survey (e.g., [Groves et al., 2012](#)). When comparing the findings of the GT data and the different personas, we found the topics identified by each persona and GT to be relatively similar. However, the order of their importance differed, and we found some variation in the search terms the different personas obtained (e.g., whether the fieldwork institute (infas) was included in the search terms). The differences in the data of the personas and GT are likely the result of personalized searches per account.

Non-Realtime Data

We now turn to the case in which GT data were obtained after more than 7 days had passed from the selected timeframe: non-realtime data. [Figure 3](#) shows that the distribution of queries within our observation period of the real-time data differed from the mean of the 30 non-realtime data samples that we obtained in a month of data collection. We still detected an increase in search activity after the invitation letters reached potential respondents (replicating our prior findings), but compared to the analysis presented in [Figure 1](#) based on hourly real-time data, the non-realtime data lacked the more fine-grained information needed to differentiate between low, intermediate, or high frequency search times. This finding is in line with previous research by [Rovetta \(2021\)](#) who highlighted that due to the sampling procedure, results might differ between GT samples.

With respect to all the non-realtime data samples we obtained, GT did not provide us with additional information on geographical or related search terms due to low number of data points in the non-realtime sample (as reported by GT when we requested the data). Since the number of letters sent for the FReDA survey was particularly high compared to other surveys in Germany, we expect a similar lack of information for other large-scale population surveys. In our view, this caveat combined with lower granularity and the deviation from real-time data strongly impairs the usability of non-realtime data regarding the goal of learning about survey participation.

If non-realtime data are to be used to detect patterns in search queries for surveys, previous research ([Rovetta, 2021](#); [Stephens-Davidowitz & Varian, 2015](#)) has proposed combining data

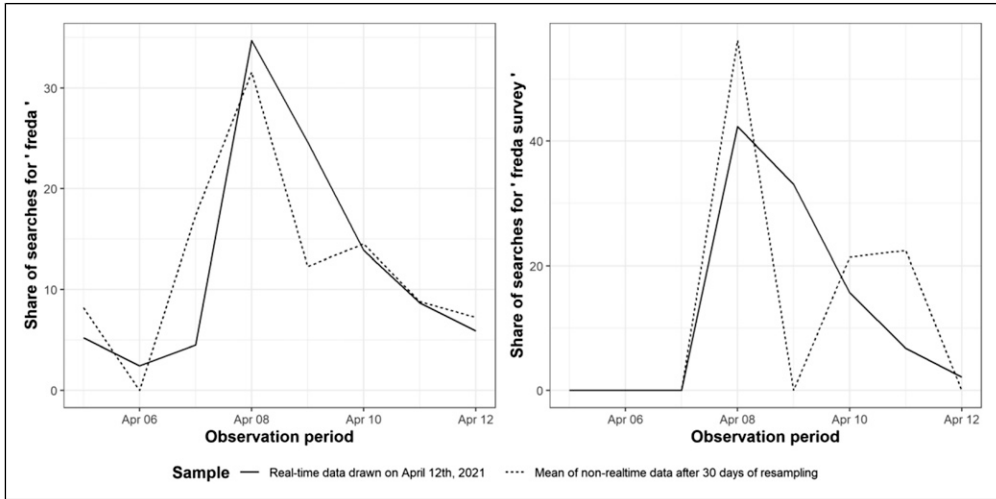


Figure 3. Share of searches for “freda” and “freda survey”; real-time data compared with mean of non-realtime data after 30 days of resampling.

from multiple non-realtime data samples (i.e., combining data from multiple samples, each drawn on a different day). Accordingly, we compared different strategies of combining non-realtime data and how well they resembled the real-time data that we used as a reference. Figure 4 shows how resampling strategies differ from real-time data as measured by the Kolmogorov–Smirnov test statistic D . The box plots in the top panels of Figure 4 show the average test result of each resampling strategy for the search terms “freda” and “freda survey” as well as how the outcomes vary within each strategy. For both search terms, we found the average D of weekly resampling (i.e., combining data from 7 days) and the cumulative 30-days resampling to be lower when compared to using single-day samples. Moreover, the variation in D within the weekly or cumulative 30-day sampling strategies was considerably lower compared to single-day samples. This finding suggests that both resampling strategies yield more consistent and reliable results compared to single-day samples. Overall, our findings highlight that combining data from multiple samples more closely resembles the real-time data when compared to using data from a single-day sample. For the search term “freda,” the cumulative 30-days resampling resembled the distribution of real-time data more closely than weekly resampling. For the search term “freda survey” we detected almost no differences in average D for the cumulative 30-days resampling and weekly resampling. A possible explanation for this finding might be that the respective search terms differed in the amount of noise they captured and how well represented they were in the daily samples provided by GT.

Regarding the cumulative 30-days resampling strategy, the bottom panels of Figure 4 show how D changed with each added sample. As a comparison, we added the average D s of the 24 weekly and the 30 single-day samples. Our results indicate that particularly in the beginning of the 30-day period, increasing the number of cumulated samples improved the D value for both search terms. Whereas for “freda survey” the D value remained the same from day two on, the D values for “freda” still varied and even worsened when the later samples around day 28–30 of our data collection period were added. Most notably, we found that even after 30 days, adding more samples did not guarantee a decrease in D . This surprising finding might be an

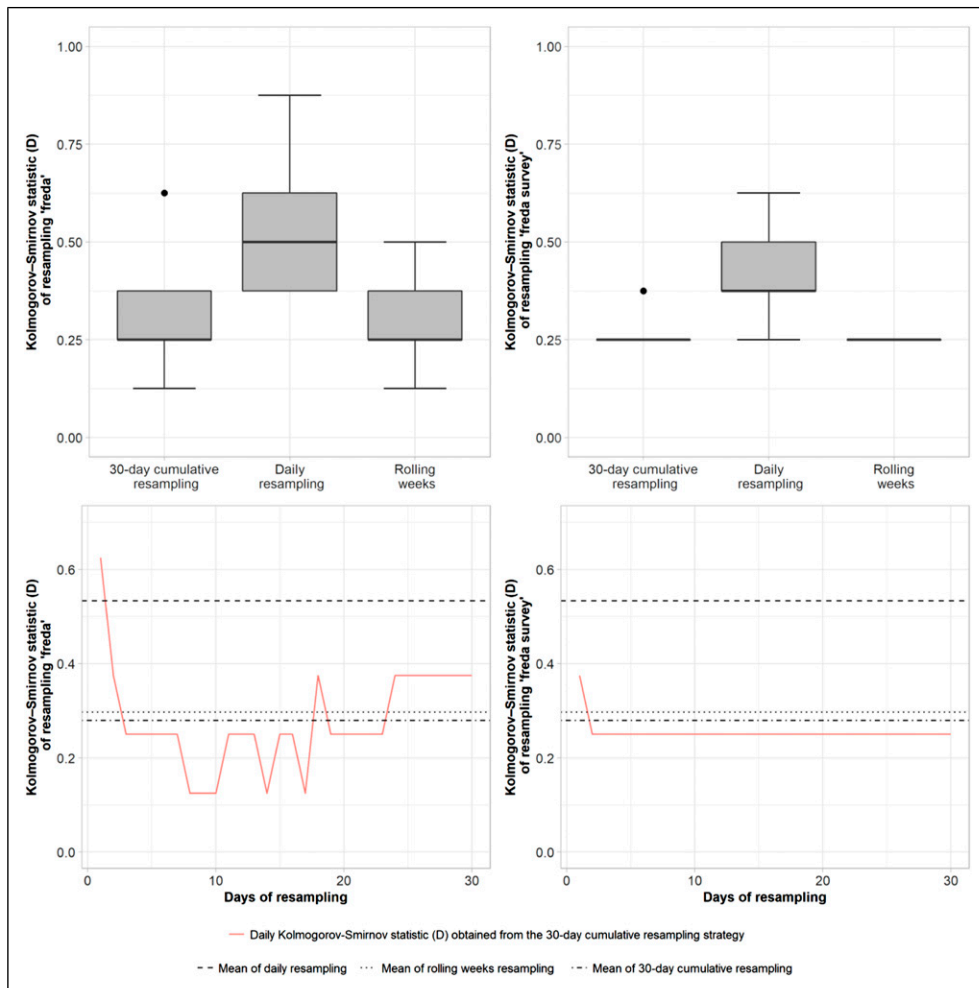


Figure 4. Density plot of the Kolmogorov–Smirnov statistic D obtained for differences between real-time data and different approaches of using non-realtime data.

indication of the need to aggregate even more samples so to stabilize the estimate for certain search terms (as illustrated by the difference in variation for “freda” and “freda survey”).

Conclusion

In the present study, to gather additional information about the participation process in surveys, we investigated the use of search engine data obtained via the GT service. Most importantly, we found that for our case study FReDA, some potential respondents searched for information about this survey via Google Search. This finding highlights the need to provide accessible and findable information for respondents on a survey website. The data that we were able to obtain also enabled us to track at which point of time some potential respondents were processing their participation decision and looking for information. Moreover, we found indications that for those who searched for specific information, the most important topics were the 5€ prepaid incentive and the

credibility of the survey. Thus, when revising respondent information in invitation letters, brochures, and the website, this information can be addressed in more detail and prominence.

To the best of our knowledge, no prior study has addressed the use of GT data for learning about what survey-related information potential survey respondents search for online. Thus, we paid special attention to describing the different ways of obtaining GT data. We found that GT's real-time data yielded the highest potentials for learning about participation in surveys. These data offered the highest granularity combined with additional geographical information and related search terms, which can offer additional insights. This information was not available in the non-realtime data when searching for "*freda*" and "*freda survey*" due to the low number of data points for searches for these terms. We used FReDA as a case study because it is a rather large survey that collected roughly 30,000 interviews in its recruitment wave. The invitation letter tranche we sent out included 54,845 letters. We assume that not many other social science surveys will achieve much higher numbers, and thus leave enough data points in GT's non-realtime data to obtain the required supplemental information.

Recommendations and Implications for Survey Research

Our findings have implications for researchers who want to use GT data to complement their survey fieldwork process and to potentially gain insights about survey participation. First, we recommend designing GT data collection so that real-time data can be obtained. This approach will enable researchers to draw on more information, compared to relying on non-realtime data. The sparsity of information in non-realtime data is especially challenging for surveys with small samples. Thus, when relying on GT data to complement the data collection of small surveys, we strongly recommend a simultaneous collection of real-time data, since non-realtime data will most likely not provide any additional information on related search terms or the geographical stratification of searches that we found useful in our study.

Second, when obtaining real-time data is not feasible—for instance, when fieldwork already has ended—and only non-realtime data can be obtained we recommend a resampling approach that combines multiple daily samples. On average, the 30-days cumulative strategy performed better than the alternatives we tested. However, we found that even after 30 days of resampling, dependent on the search term, variation still existed in how well the aggregated data reflected the real-time data we used as a reference. Consequently, we will refrain from giving a naïve rule of thumb such as "combine samples of one week/month," and instead suggest that researchers start combining multiple samples and continue this approach until adding more samples no longer changes the distribution of search queries for their search term(s) of interest. Alternatively, [Fang et al. \(2021\)](#) have suggested a resampling approach using overlapping shifting timeframes that each cover the period of interest. These researchers also cautioned that GT servers would only allow the retrieval of a maximum of 244 daily samples per inquiry and search term. Future studies might consider the implications of this limitation.

Third, we recommend a careful selection of search terms that relate to a survey and are not used too frequently for other purposes. For our study, we tracked the search terms "*freda*" and "*freda survey*" before we invited potential respondents. This decision enabled us to establish a baseline of how these search terms are commonly used and in which context. Researchers will face the challenge of selecting a variety of simple and general search terms that potential respondents might use while also not selecting terms that are too general and do not allow to draw conclusions with respect to a survey (e.g., "*family survey*" instead of "*freda survey*"). If tracking search behavior is important for a survey, we recommend checking Google Search for how a search term is used and in which context when deciding about a project acronym, name, or title. Moreover, researchers could frame the use of specific text strings in invitation letters (e.g., *the FReDA survey*)

by highlighting them for respondents. This framing might increase the likelihood of respondents searching for these text strings. More research on the topic of search term selection certainly is required.

As we argued previously, information on nonparticipation is sparse but necessary for learning more about what drives respondents to (not) participate so to develop effective techniques to increase response rates and cater to respondents' needs. While GT data are not a panacea for solving the nonresponse problem, they may offer new insights, are freely available, and easy to obtain. Thus, these data can provide insights at a very low cost. In our view, it is reasonable to assume that those who search for information about a survey online are a mix of respondents and nonrespondents. What we found especially helpful was to know at what point in time after mailing the invitation letter some potential participants were looking for information about the survey. We also were able to draw on the information from "Related searches" that helped us to identify topics that we needed to investigate in more detail. This strategy might be especially useful for surveys that are faced with unknown issues during fieldwork and need to discover first insights about where to look for design characteristics that might be flawed. If such analyses are performed during fieldwork, these insights can be used to change the information provided in reminder letters sent to those potential respondents who have not yet made a decision to participate.

Limitation and Research Opportunities

We aimed to make GT data accessible to survey researchers as a way to tackle the nonresponse challenge. Since our study is, as far as we know, the first to focus on the information that potential survey respondents search about with respect to a survey, our study lays a foundation for doing more work on this topic. Accordingly, the limitations of our research also should be addressed by future research. First, to illustrate the use of GT data for learning about survey participation, we drew on FReDA, which is a rather large survey, but the non-realtime data available for smaller projects might be sparse. Thus, we suggest that more research is required to investigate the use of GT data when rare search terms are used. Specifically, it would be helpful to determine at which sample sizes non-realtime data might not be feasible anymore, and so real-time data would be necessary. Thus, we recommend that future research investigates methods of collecting and using GT data for terms that are searched with low compared to high frequency.

Second, as we have argued previously, GT data are likely to include respondents and nonrespondents who searched for information about a survey as well as noise. We were not able to determine the number of respondents and nonrespondents who searched for information, since we could not link the survey and GT data on a person-level. However, we find it reasonable to assume that respondents are less likely to search for information about a survey if they already have decided to participate. In addition, we argue that learning about respondents who were reluctant at first, then searched for sensible information, and then participated would be helpful for understanding the participation decision process.

Third, since we intended to break ground on the link between GT data and survey participation, we focused on the practical challenges of obtaining and using these data. In our understanding, GT can provide first insights (i.e., an explorative study) into which follow-up research to conduct and which design characteristics to adjust to enhance survey participation. We did not analyze ways of using this information to change a design, and thus see merit in future research that combines GT with follow-up studies. For instance, GT data could be used to learn about topics that respondents search information about, and then cognitive pretesting could be applied to informational material for respondents with respect to these topics. This approach could be used to develop improved information materials for respondents. Moreover, we only focused on the insights gained by using GT. As argued previously, we view GT data as a complement to other methods such as cognitive

pretesting, field pretests, and focus groups. Each method yields different insights into the participation process of a survey, and we assume that each one will be valuable in its own way, given the particulars of a survey. However, compared to the other methods mentioned, GT data comes at very low cost and is an unobtrusive data collection approach, which makes it an interesting addition to more conventional methods. Future research could try to systematically compare the insights gathered by each method and assess the respective costs. In addition, findings from other areas outside survey research (e.g., usability research, search engine optimization, human–computer interaction) may help to better utilize search engine data to provide relevant information to potential respondents. For instance, search engine optimization could be utilized to ensure that respondents can quickly and easily find information about a survey (i.e., ensure that the required information of project websites is shown at the top of the search page). The findings from human–computer interaction may help researchers to better understand how potential respondents use search engines, and which search terms they rely on to learn about a survey. In the present study, we relied on “*freda*” and “*freda survey*,” since we reasoned that while Freda is a well-known name for other things, in the German context, it would be specific enough so it could be used to detect valuable information about our survey. Our analyses showed that the related search terms were related to the survey, which is a test for the validity of search terms suggested by [Zhu et al. \(2012\)](#) and also applied by [Fang et al. \(2021\)](#). Nevertheless, more elaborate strategies for selecting search terms are required, especially with respect to covering the universe of all relevant terms for a survey.

Fourth, in our analyses, we used real-time data as a reference distribution. In other words, we investigated how well different strategies of aggregating non-realtime data yielded the same distributions of searches as the real-time data. This approach rests on the assumption that the larger, finer granular sample that GT offers for real-time data is reliable. Testing this assumption in future research is an important task for better understanding the potential of GT data. We suggest that such studies could apply the approach used by [Fang et al. \(2021\)](#)—overlapping samples of real-time data, each covering a shorter timeframe that then is compared between samples. Since different periods likely will have comparability issues due to the different reference values on which the RSVs are scaled, we suggest also considering whether the calibration approach suggested by [Fang et al. \(2021\)](#) could be applied.

Finally, GT provides relative values (i.e., RSVs), and does not offer absolute frequencies of search queries for a specific search term in a timeframe. In the present study, we were not interested in uncovering absolute frequency or the salience of searches about FReDA compared to other search terms. However, our study rests on the assumption that a relevant number of searches were conducted by potential respondents, and thus, GT data would yield valuable insights. For our study, we tested this assumption by collecting data for several days prior to the mailing of our invitation letters, so we could establish baseline data. Strictly speaking, however, absolute frequencies would be needed to test this assumption and to assess how many potential respondents searched for information about the survey. Absolute frequencies are even more important for studies that follow the approach of [Fang et al. \(2021\)](#) who propose collecting data about the salience of societal trends that are relevant for survey participation. Potential solutions for this challenge might be to query additional search terms (i.e., an external baseline) for which GT then adjusts the RSVs. However, this approach would still not yield absolute search frequencies, rather just add a baseline for comparison (for which absolute search frequencies also are unknown). Considering the challenge of obtaining absolute frequencies, we encourage future research to consider the possibility to utilize the Google Keyword Planner offered via the Google Ads service. Given that we did not consider this service during our data collection and study planning, we recommend future research to systematically investigate its use when planning their analyses.

Overall, we found GT data to be easily accessible and to provide potential additional insights into survey participation. In our view, the most important points regarding these data are that they are free and available. Since declining response rates have troubled survey researchers for years, they have to draw on whatever data can help them learn more about how to make better surveys and motivate potential respondents to participate.

Acknowledgments

The authors would like to thank Bernhard Miller for his excellent comments on an earlier draft of this study that encouraged us to follow this line of research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the German Federal Ministry of Education and Research (BMBF) as part of FReDA (grant number 01UW2001B).

Data Availability

Replication code and data are available at the GESIS SowiDataNet|datorium repository: <https://doi.org/10.7802/2453>.

ORCID iDs

Tobias Gummer  <https://orcid.org/0000-0001-6469-7802>

Anne-Sophie Oehrlein  <https://orcid.org/0000-0003-2617-7296>

Notes

1. The recruitment process was split into two tranches. On April 7, 2021, 54,845 invitation letters were sent to potential respondents of the first tranche. One week later, on April 14, another 53,411 identical invitation letters were sent to potential respondents of the second tranche. Thus, a total of 108,256 potential respondents were invited to participate in the FReDA panel. For the purpose of the present study, we focused on the first tranche.
2. The calculation of RR2 for Figure 2 includes all web questionnaires submitted between April 8 and April 12, 2021. With respect to the mail questionnaires, we considered participation until April 15, 2021 so to account for any postal service delivery delays.

References

- AAPOR. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. The American Association for Public Opinion Research.
- Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *Annals of the American Academy of Political and Social Science*, 645(1), 36–59. <https://doi.org/10.1177/0002716212456834>
- Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, 88(S1), 2–9. <https://doi.org/10.2139/ssrn.1659302>

- Cornesse, C. (2020). The utility of auxiliary data for survey response modeling: Evidence from the German Internet Panel. *Survey methods: Insights from the field*. <https://doi.org/10.13094/SMIF-2020-00008>
- de Leeuw, E. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.
- de Leeuw, E. (2018). Mixed-Mode: Past, present, and future. *Survey Research Methods*, 12(2), 75–89. <https://doi.org/10.18148/srm/2018.v12i2.7402>
- de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41–54). John Wiley & Sons.
- de Leeuw, E., Hox, J., & Luiten, A. (2018). International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data. *Survey methods: Insights from the field*. <https://surveyinsights.org/?p=10452>.
- Fang, Q., Burger, J., Meijers, R., & van Berkel, K. (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>
- Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society. Series A*, 169(3), 439–457. <https://doi.org/10.1111/j.1467-985X.2006.00423.x>
- Groves, R. M., Presser, S., Tourangeau, R., West, B. T., Couper, M. P., Singer, E., & Toppe, C. (2012). Support for the survey sponsor and nonresponse bias. *Public Opinion Quarterly*, 76(3), 512–524. <https://doi.org/10.1093/poq/nfs034>
- Gummer, T., Christmann, P., Verhoeven, S., & Wolf, C. (2022). Using a responsive survey design to innovate self-administered mixed-mode surveys. *Journal of the Royal Statistical Society A*, 185(3), 916–932. <https://doi.org/10.1111/rssa.12835>
- Gummer, T., Schmiedeborg, C., Bujard, M., Christmann, P., Hank, K., Kunz, T., Lück, D., & Neyer, F. J. (2020). The impact of Covid-19 on fieldwork efforts and planning in pairfam and FReDA-GGS. *Survey Research Methods*, 14(2), 223–227. <https://doi.org/10.18148/srm/2020.v14i2.7740>
- Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting & Social Change*, 130, 69–87. <https://doi.org/10.1016/j.techfore.2017.11.009>
- Luijckx, 0052., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Fredriksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Þór Gunnarsson, S., Hjaltason, A. B., Joye, D., Lomazzi, V., Mainieri, A. M., Milbert, P., Ochsner, M., Pollien, A., Pollien, A., Sapin, S., & Wolf, C. (2021). The European values study 2017: On the way to the future using mixed-modes. *European Sociological Review*, 37(2), 330–346. <https://doi.org/10.1093/esr/jcaa049>
- Olson, K., Lepkowski, J. M., & Garabrant, D. H. (2011). An experimental examination of the content of persuasion letters on nonresponse rates and survey estimates in a nonresponse follow-up study. *Survey Research Methods*, 5(1), 21–26. <https://doi.org/10.18148/srm/2011.v5i1.4324>
- Pförr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, M., Hajek, K., Helmschrott, S., Kleinert, C., Koch, A., Martin, S., Saßenroth, D., Schmiedeborg, C., Trüdinger, E.-M., & Rammstedt, B. (2015). Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. *Public Opinion Quarterly*, 79(3), 740–768. <https://doi.org/10.1093/poq/nfv014>
- Poushter, J. (2016). *Smartphone ownership and internet usage continues to climb in emerging economies*. Pew Research Center.
- Rovetta, A. (2021). Reliability of Google trends: Analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. *Frontiers in Research Metrics and Analytics*, 6, 670226. <https://doi.org/10.3389/fma.2021.670226>

- Schneider, N. F., Bujard, M., Wolf, C., Gummer, T., Hank, K., & Neyer, F. J. (2021). Family research and demographic analysis (FReDA): Evolution, framework, objectives, and design of "the German family-demographic panel study". *Comparative Population Studies*, 46, 149-186. <https://doi.org/10.12765/CPoS-2021-06>.
- Schouten, B., Peytchev, A., & Wagner, J. (2018). *Adaptive survey design*. CRC Press.
- Stephens-Davidowitz, S., & Varian, H. (2015). A hands-on guide to Google data. In.
- Taylor, K., & Silver, L. (2019). *Smartphone ownership is growing rapidly around the world, but not always equally*. Pew Research Center.
- Timoneda, J. C., & Wibbels, E. (2021). Spikes and variance: Using Google trends to detect and forecast protests. *Political Analysis. online first*. <https://doi.org/10.1017/pan.2021.7>
- Tourangeau, R. (2017). Presidential address: Paradoxes of nonresponse. *Public Opinion Quarterly*, 81(3), 803–814. <https://doi.org/10.1093/poq/nfx031>
- Tourangeau, R., Brick, J. M., Lohr, S., & Li, J. (2017). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society A*, 180(1), 203–223. <https://doi.org/10.1111/rssa.12186>
- Vandeplas, C., Joye, D., Staehli, M., & Pollien, A. (2015). Identifying pertinent variables for nonresponse follow-up surveys. Lessons learned from 4 cases in Switzerland. *Survey Research Methods*, 9(3), 141–158. <https://doi.org/10.18148/srm/2015.v9i3.6138>
- Wagner, J. (2008). *Adaptive survey Design to reduce nonresponse bias*. PhD thesis. University of Michigan.
- West, B. T., Wagner, J., Hubbard, J., & Gu, H. (2015). The utility of alternative commercial data sources for survey operations and estimation: Evidence from the national survey of family growth. *Journal of Survey Statistics and Methodology*, 3(2), 240–264. <https://doi.org/10.1093/jssam/smv004>
- Wolf, C., Christmann, P., Gummer, T., Verhoeven, S., & Schnaudt, C. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly*, 85(2), 623–648. <https://doi.org/10.1093/poq/nfab039>
- Zhu, J. J. H., Wang, X., Qin, J., & Wu, L. (2012). Assessing public opinion trends based on user search queries: Validity, reliability, and practicality. In Paper presented at the Annual Conference of the World Association for Public Opinion Research, Hong Kong, 2012.

Author Biographies

Tobias Gummer is a senior researcher and team leader at GESIS – Leibniz Institute for the Social Sciences, Mannheim, and private lecturer at the University of Mannheim, Germany.

Anne-Sophie Oehrlein is a PhD student at GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany.