

Kühne, Simon; Kroh, Martin; Liebig, Stefan; Zinn, Sabine

**Article — Published Version**

## The Need for Household Panel Surveys in Times of Crisis: The Case of SOEP-CoV

Survey Research Methods

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Kühne, Simon; Kroh, Martin; Liebig, Stefan; Zinn, Sabine (2020) : The Need for Household Panel Surveys in Times of Crisis: The Case of SOEP-CoV, Survey Research Methods, ISSN 1864-3361, European Survey Research Association, Southampton, Vol. 14, Iss. 2, pp. 195-203, <https://doi.org/10.18148/srm/2020.v14i2.7748> , <https://ojs.ub.uni-konstanz.de/srm/article/view/7748>

This Version is available at:

<https://hdl.handle.net/10419/222577>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc/4.0>

# The Need for Household Panel Surveys in Times of Crisis: The Case of SOEP-CoV

Simon Kühne  
Faculty of Sociology  
Bielefeld University, Germany

Martin Kroh  
Faculty of Sociology  
Bielefeld University, Germany, and  
Socio-Economic Panel (SOEP)  
DIW Berlin, German Institute for Economic Research

Stefan Liebig  
Socio-Economic Panel (SOEP)  
DIW Berlin, German Institute for Economic Research

Sabine Zinn  
Socio-Economic Panel (SOEP)  
DIW Berlin, German Institute for Economic Research

The spread of the novel coronavirus, SARS-CoV-2, poses major challenges for individuals and societies at large. The question now is how individuals and society are dealing with these challenges, and what health, psychological, social, and economic effects they will have to bear. Meaningful answers can only be provided using a generalizable database that contains contextual information such as family characteristics and offers a life-course perspective. We argue that these criteria are best met by household panel survey data. In this paper, we illustrate the importance of such data and show how they can aid in explaining the current and future effects of the pandemic on individuals, households, and society. Furthermore, we describe the survey design of SOEP-CoV, a supplementary telephone survey to the German Socio-Economic Panel (SOEP) study.

*Keywords:* COVID-19; panel surveys; longitudinal data; household surveys; soep

## 1 Societies in Crisis

The novel coronavirus (SARS-CoV-2) and the respiratory disease it causes, COVID-19, have spread worldwide within a matter of weeks. On March 11, 2020, the World Health Organization (WHO) classified COVID-19 as a pandemic. In many countries, experts forecast a sharp increase in infections with the virus ultimately expected to infect up to 70 percent of the population if vaccines are unavailable in the foreseeable future (German Federal Press Conference on March 11, 2020). To prevent the medical and health system from becoming overwhelmed by patients in need of treatment, the rate of new infections must be minimized and the spread of the disease must be slowed over as long a period as possible. Keeping the number of simultaneous infections to a minimum will also protect the social groups that are most at risk: the elderly and people with pre-existing health conditions. Due to the expected increase in the number of new infections, current and future measures and guidelines such as contact bans and school closures affect virtually all aspects

of life in many countries—including Germany, the country under investigation here—and will most likely continue to do so in the near future.

With the measures currently being undertaken and those probable in the near future—entailing the avoidance of almost all social contact, even within families, and leading to the cancellation of events, the closure of educational and care facilities, the restriction of individual freedoms of movement and assembly, and even the closing of borders—globally, societies around the world facing a challenge that threatens to cut deep into social, political and economic structures and processes and to change the way people go about daily life for some time to come. Beyond short-term period effects, the resulting medium- and long-term consequences for society and the individual cannot yet be assessed. This is in part due to the lack of knowledge about the individual significance, impacts, and above all subjective experience of this event and how it will affect people in the long term. We know from past experiences that collectively experienced crises can have formative effects over the lifespans on entire generations—as seen with the Great Depression, the Second World War, and the fall of the Berlin Wall—and that impacts at the individual and societal level only become evident with a time lag (Mannheim, 1928).

Urgent questions arise as to the medical and health im-

---

Simon Kühne, Bielefeld University, Faculty of Sociology, Universitätsstraße 25, 33615 Bielefeld, Germany (E-mail: simon.kuehne@uni-bielefeld.de)

pacts of COVID-19; the social, psychological, economic, and political factors that play a role in its spread; and the consequences thereof. These questions cannot currently be investigated due to the lack of comprehensive databases allowing for reliable inferences about societies at large.

## 2 COVID-19 and the Social Sciences

Since the enormous global impacts of COVID-19 began to become evident, scientists in the natural and life sciences have been working under great pressure to understand the virus, the disease, and its spread. Until recently, the media have focused primarily on researchers in the fields of virology, biology, and epidemiology when reporting on the virus. In many countries, virologists such as Anthony Fauci in the United States, Johan Giesecke in Sweden, and Christian Drosten in Germany have become famous in their respective countries for presenting scientific findings and recommendations in the media.<sup>1</sup> Policy makers (in Germany) have also based their decisions largely on the opinions of experts from the natural and life sciences. The committees created in the early weeks of the pandemic to advise the government and other decision makers included relatively few experts or scientists from other disciplines such as the social sciences and humanities. This is problematic, as medical or natural science perspectives alone do not provide a sufficient basis for political and policy decisions on comprehensive responses to the crises. Virologists, for instance, may identify the mechanisms by which the virus spreads between humans and conclude that people need to maintain a certain physical distance in order to minimize infections. However, their data and conclusions cannot provide insights into how physical distancing (unfortunately often termed "social distancing") is impacting the economy, how well households are able to cope with the responsibility for home schooling, or how the crisis is affecting people's mental health and well-being. Another frequently neglected aspect in medical or natural science perspectives relates to the public measures undertaken to minimize the infections. Politicians and decision makers need to know how willing people are to change their everyday behavior, under what conditions they are willing to do so, and to what extent incentives or monitoring may be needed. Public measures must also be evaluated with respect to unintended consequences—both short-term and long-term—on other, non-health related sectors of society such as the economy and the education system. To answer these urgent questions, social science data and research are needed.

As the virus and public measures to contain its impact affect more and more areas of life, the call for non-medical perspectives on the crisis and its impact has grown louder. Policy makers and the media are now taking different perspectives into account, including research by social scientists in the fields of sociology, political science, and economics, as well as research in branches of the humanities such as philoso-

phy.<sup>2</sup> At present, social scientific studies collecting online survey data from self-recruited participants have the highest media visibility. A comprehensive list of current ongoing corona-related survey projects in Germany has been released by the German Data Forum, RatSWD.<sup>3</sup> These surveys make an important contribution to better understanding the current impact of the virus and public measures to contain it on the lives of individuals. However, these studies face limits when it comes to providing results that can be generalized to a country's entire population. This is because people who select into a given study themselves, rather than being selected through a random statistical procedure, generally do not offer a representative picture of the larger population. Compared to randomly selected respondents, they tend to come from particular groups, such as people with a special interest in the subject of the study, or people with a desire to contribute to scientific studies because of their professional or educational background (Marcus, Bosnjak, Lindner, Pilischenko, & Schütz, 2007, 3). Furthermore, by definition, online studies underrepresent population groups without Internet access or people who rarely or never use the Internet. Although such groups have been declining over recent years, they still make up a significant proportion of Germany's population: about 16% as of 2018 (Statista, 2020).

Another common challenge of online surveys is that of the difficulty of collecting contextual information. For instance, conducting multiple personal interviews in households is difficult to administer in online surveys. When such information (e.g., data about other household members) is collected, it is usually only from the person interviewed. Two problems arise here: First, it is only possible to collect proxy information about third parties inflating the risk of misreporting. Secondly, problems of missing data emerge due to the natural inability of one person to report fully on another person. The time taken to collect data (on others) also usually poses an obstacle to comprehensive data collection (Rolstad, Adler, & Rydén, 2011, 80). Missing data are generally a serious problem for one-off cross-sectional surveys, as missing information can neither be collected in follow-up surveys nor be derived from past rounds of data collection.

Moreover, cross-sectional (online) surveys can not be used to analyze long-term consequences or impacts of past experiences. Whereas cross-sectional surveys can collect some information on past experiences based on biographical infor-

<sup>1</sup>For instance, a Google search for "Christian Drosten", the director of the Institute of Virology at Charité Hospital Berlin produced 2.6 million search results (on May 6, 2020).

<sup>2</sup>One example of policy makers incorporating social science perspectives into their decision making is the "Expert Council Corona" in Northrhine Westphalia, Germany.

<sup>3</sup>See <https://www.ratswd.de/en/studies>. As of May 8, 2020, the list comprises 44 quantitative survey projects in the social sciences with 27 based on open self-recruitment.

mation provided by respondents, information about past attitudes or feelings is usually highly problematic due to measurement error (Tourangeau, Rips, & Rasinski, 2000).

Longitudinal surveys with a multi-person perspective such as household panel studies are a solution to these problems. Such data make it possible to model the individual's embedding within social contexts, individual reactions to the crisis, and the effects of the crisis on both individuals and their households. In the best-case scenario, these kind of surveys already provide rich information about individuals and households collected over the course of many waves *prior* to the crisis.

The question of survey mode is another important aspect to consider when thinking about the future research potential of survey data. The majority of large-scale household panel studies rely on interviewer-administered surveying (e.g., the Socio-Economic Panel Survey (SOEP), UK Understanding Society, and the Household, Income and Labour Dynamics in Australia (HILDA) Survey, all of which rely largely on face-to-face interviewing; and the Swiss Household Panel and the Panel Study of Income Dynamics (PSID), both of which rely on telephone interviewing). Clearly, face-to-face interviewing is currently not possible in most countries due to protective measures against COVID-19. However, both telephone interviewing and (future) face-to-face interviewing offer a number of advantages over self-administered survey modes such as web surveys, and researchers are still willing to spend much more on data collection in order to make use of these advantages: First, interviewers can establish a social relationship with respondents, and many respondents prefer a personal, conversational-style interview over completing a questionnaire entirely on their own. Moreover, in longitudinal studies, interviewer continuity itself promotes panel stability (e.g., Campanelli & O'Muricheartaigh, 1999). This is one reason why panel stability rates are usually higher with interviewer-administered survey modes than with self-administered modes such as online surveys. Second, interviewers make it possible to use complex instruments such as household grids in data collection. Third, (future) face-to-face surveying with an interviewer present offers possibilities to integrate medical or psychological testing into an ongoing survey. For instance, in wave 2015 of the SOEP Innovation Sample (SOEP-IS), saliva samples were collected by interviewers who had undergone special training (Gerstorff & Schupp, 2016). Thus, interviewers in ongoing household panel surveys could also collect biological samples for antibody testing in the context of SARS-CoV-2 & COVID-19.

### 3 The Need for Household Panel Survey Data

We argue that household panel surveys are needed in order to address the most fundamental questions about the individual and societal consequences of the corona pandemic.

Individuals are embedded in social contexts, structures,

groups, and networks. For many, the core family or household represents the most important social group, influencing almost all aspects of life and (social) (inter)action. Many people tend to evaluate situations not solely based on their own needs and wants, but also based on the needs of their close-relatives and household members. Moreover, families and households differ in the resources available to them; households with access to more resources and capital have more options available when facing threats and stressors. To analyze how the corona crisis is affecting individuals and households (differently) and how its micro-level consequences translate into complex social phenomena and macro-level structural change, a household perspective is needed.

The individual life course perspective provides further crucial insights into the impacts of crises such as the corona pandemic. As Ryder (1965, p. 856) notes "A person's past affects his present, and his present affects his future." Past experiences and resources accumulated throughout life affect how individuals are able to handle an acute crisis. In addition, the impacts people are experiencing in the crisis will affect their future over years to come. Longitudinal panel data collected both before and after the pandemic are needed to analyze these interrelations over time. Cross-sectional or longitudinal trend survey data are not capable of this. From a statistical perspective, longitudinal panel data also provide a stronger foundation for causal inferences, as they are based on within-variation at the respondent and household level, making it possible to control for unobserved heterogeneity.

## 4 The SOEP-CoV Survey Project

In April 2020, a research team at the Socio-Economic Panel (SOEP) at DIW Berlin and the Universität Bielefeld launched the project "SOEP-CoV - The Spread of the Coronavirus in Germany: Socio-Economic Factors and Consequences" (see [www.soep-cov.de](http://www.soep-cov.de), currently in German only). The project is funded by the German Federal Ministry of Education and Research (BMBF) as part of its "call for proposals for research on COVID-19 in the wake of the Sars-CoV-2 outbreak".<sup>4</sup>

The research project aims to investigate the acute, medium-term, and long-term socio-economic factors in and consequences of the spread of the coronavirus in Germany. To this end, two telephone surveys were planned, the first one conducted throughout the acute phase of the pandemic from April to July 2020, and a second survey scheduled to take place after the acute phase of the pandemic (we hope in 2021).

<sup>4</sup>See <https://www.bmbf.de/foerderungen/bekanntmachung-2865.html>, accessed May 27, 2020.



## Survey Design and Sampling

The basis for the SOEP-CoV project is the Socio-Economic Panel (SOEP, see Giesselmann et al., 2019), a longitudinal household survey that has been running since 1984 and is the largest ongoing panel study in Germany (approximately 20,000 households with more than 30,000 adult household members surveyed annually, see Siegers, Belcheva, and Silbermann, 2020). Recruiting participants from an ongoing household panel survey offers the advantage of being able to draw on a wealth of existing household information from previous interviews, including family structure, the household division of labor, the household's financial and child care situation, as well as individual information on socio-demographic characteristics, health, well-being, personality, and social and political attitudes. First, this meant that in the SOEP-CoV study, we did not need to spend any interview time collecting the kind of background information on respondents and households that is needed for in-depth data analysis. This allowed us to spend more interview time on questions related to the COVID-19 pandemic. Second, the use of established SOEP survey instruments in SOEP-CoV that will also be used in future waves of the SOEP guarantees the possibility of genuine panel analysis in future research on short-term and long-term changes in individuals' attitudes and behaviors as a consequence of the corona pandemic. Such future research will be able to study numerous aspects of life, including changes in employment, income, social contacts, and well-being.

The potential for life-course research on the long-term consequences of the pandemic is strengthened, on the one hand, by the already high longitudinal response rate in the SOEP of around 90 percent from wave 3 onward (Siegers et al., 2020). On the other hand, the SOEP will adapt future replication cycles of survey instruments to be better able to studying consequences of the pandemic on, for instance, mental health, inequalities, but also on critical life events. Moreover, formats of dependent interviewing will be used in particular contexts, such as households experiencing an infection, to better understand behavioral risk factors, the course of the disease, and its long-term health effects.

The SOEP-CoV gross sample consists of all SOEP households that have been part of the SOEP since at least 2018, in sum 12,000 households. The SOEP Innovation Sample and the refugee samples are excluded from this study<sup>5</sup> as are households for which no valid telephone number was available (about 25 percent of all households). In each household, one individual is interviewed via telephone.

Interviewing a single individual in each household via telephone differs from the SOEP-standard of in-person household interviews, however, from our perspective, it is the most practical approach given the legal, time, and cost restrictions we were facing. We chose the telephone interviewing mode as face-to-face interviewing—the predominant sur-

vey mode in the SOEP—is practically impossible at times of physical distancing measures. Questionnaire length is often an issue in telephone surveys, however, due to rich information on households and individuals already available from past SOEP waves, our questionnaire can focus on Corona-related topics. The average interview took about 27 minutes, thus, the questionnaire length suits the telephone survey mode. Given restricted financial and personnel resources, we chose to interview only one individual in many households rather than all individuals in fewer households. As the regular (future) SOEP waves will interview all individuals in each household, maximizing the number of households for which at least some information on their current situation during the pandemic is available increases future research potential—especially since all panel studies face panel attrition. Finally, we expect that the majority of SOEP households will switch back to their standard survey mode (usually CAPI) in wave 2021, thus, offering potential to include more complex survey instruments on the pandemic and collect biological samples.

To be able to analyze the temporal trajectory of the corona pandemic, the SOEP-CoV sample was divided into nine tranches, so-called replicated samples. Random allocation to these tranches was performed in such a way that the full and complex design information of the existing SOEP subsamples was preserved in each of the replicated samples.<sup>6</sup> Allocation, for instance, involved assigning households in each of the regional sampling points to these tranches and thus providing a nationally representative picture of the population of German private households in each tranche.

We implemented a step-wise fieldwork design with the first four tranches being interviewed in consecutive two-week intervals and the last five tranches being interviewed at one-week intervals. Fieldwork in Tranche 1 started on March 30th and Tranche 9 will be completed on June 28th. The (bi)weekly sample size thus decreases continuously over time with a gross sample of 3,000 households in Tranche 1 and 600 households in Tranche 9 (see Figure 1). The design follows our initial estimates of the dynamic and length of the (first) wave of infections in Germany. In fact, the actual decrease in the intensity of the crisis and its effects on respondents' lives during the current acute phase of the pandemic largely matches our initial expectations, as Figure 1 illustrates. Our design therefore follows an a priori expectation on our part. If, contrary to our expectations, the intensity of the crisis develops in a different direction, we are able to ad-

<sup>5</sup>The latter represent a very specific population group for which specific instruments and multilingual contact procedures are required, which cannot be implemented with an acceptable quality in the context of this study and in the short time available.

<sup>6</sup>Replicated samples, sometimes also termed interpenetrating samples or random groups, are often used for resampling techniques of variance estimation (Wolter, 1995).

just the design accordingly, e.g., by extending the field times of individual tranches or by combining tranches.

Interviewing nine temporally consecutive replicated samples in predefined periods instead of interviewing all households at one point in time during the pandemic has three advantages. First, the prevalence of infections, the implementation of public measures that aim at containing its spread, and societal consequences of the pandemic develop dynamically and often sequentially over a longer period of time. Treating the tranches as a combination of repeated cross-sectional studies makes it possible to investigate time trends over the course of the pandemic, for instance, in political attitudes towards government performance. Second, repeated cross-sectional tranches may also be treated as a pseudo-panel that makes it possible to estimate panel effects from a short-term perspective of four months of the pandemic (Deaton, 1985, 1-2). Finally, from a long-term perspective, systematically varying fieldwork time during the pandemic within the SOEP's genuine panel design of annual waves introduces exogenous variation in the experienced intensity of the pandemic. This in turn makes it possible to estimate heterogeneous treatment effects. In combination with regional variation, this exogenous temporal variation in the treatment effect of the pandemic makes it easier to empirically isolate causal processes. For instance, comparing SOEP-CoV interviews in German federal states Bavaria and Baden-Württemberg between the 20th and the 21st week of the year may help in identifying the effects of certain public measures to contain infections.

### Questionnaire

The questionnaire covers five main topics in addition to questions measuring the prevalence of COVID-19 in individual respondents and their households:

1. Health behavior and health inequality: Are different social status groups being affected differently by the virus and its consequences? What health-related behaviors increase/decrease the risk or severity of an infection?
2. Labor market and economic situation: What groups are facing job loss due to the pandemic? What occupations are drastically changing the way they work? How are self-employed people responding to the crisis and loss of income?
3. Social life, networks, and mobility: How are families coping with home schooling and closed schools/daycare? How is social isolation distributed throughout society? Are people making longer-term changes in their daily activities?
4. Mental health and well-being: How are worries and fears about personal health and well-being distributed

throughout the population? Is the crisis showing permanent effects on mental health?

5. Social cohesion: How do respondents feel the various levels of government (national, federal, local) have been performing during the crisis? What effects has the pandemic had on attitudes toward solidarity and trust in other people and in political institutions?

Wherever possible, existing questions from the standard SOEP questionnaires are being used for SOEP-CoV to enable genuine panel analyses.<sup>7</sup> This assures comparability with past and future survey waves. For new survey questions on the virus and related public measures, the project team cooperated with experts from the Robert Koch-Institute (RKI), Charité Hospital Berlin, Max-Planck-Institute for Human Development, the Institute for Interdisciplinary Research on Conflict and Violence (IKG, Bielefeld), and the Berlin Social Science Center (WZB).

### Fieldwork and Weighting Approach

Out of the gross sample of 6,052 households in Tranches 1 and 2, 3,621 participated in the telephone interviews. The response rate in the early tranches of SOEP-CoV was therefore 60 percent (AAPOR RR1). Based on this, we expect an overall net sample size of 7,200 households for the total of all nine tranches.

On average, there were 3.5 contact attempts needed for each successful interview. The maximum number of telephone contact attempts on varying days of the week and times a day was 23.

About 24 percent of all households could not be contacted at all. Of those that could be contacted, 22 percent refused to participate in the interview. In order to correct for potential nonresponse bias in estimates, the survey data are weighted in a number of steps (see Figure 2). To compensate for potential undercoverage bias, the weighted data are poststratified using data from the German Microcensus.<sup>8</sup>

Concretely, four nonresponse models (complementary log-log regressions, *cloglog*) were estimated. Three of the four models determine the criteria that SOEP households had to fulfill to participate in the SOEP-CoV study: (i) Households had to be part of the 2020 sample of SOEP households and have a valid telephone number (landline or mobile), (ii) Households had to be reachable by phone, and (iii) Households that could be reached by phone had to be willing to participate. In all three models, a total of more than 300 socio-demographic, regional, and economic household characteristics that are available for SOEP households are examined for their significance with regard to participation in the

<sup>7</sup>See [https://www.diw.de/en/diw\\_02.c.222729.en/questionnaires.html](https://www.diw.de/en/diw_02.c.222729.en/questionnaires.html)

<sup>8</sup>See <https://www.forschungsdatenzentrum.de/en/household/microcensus>, accessed May 8, 2020.

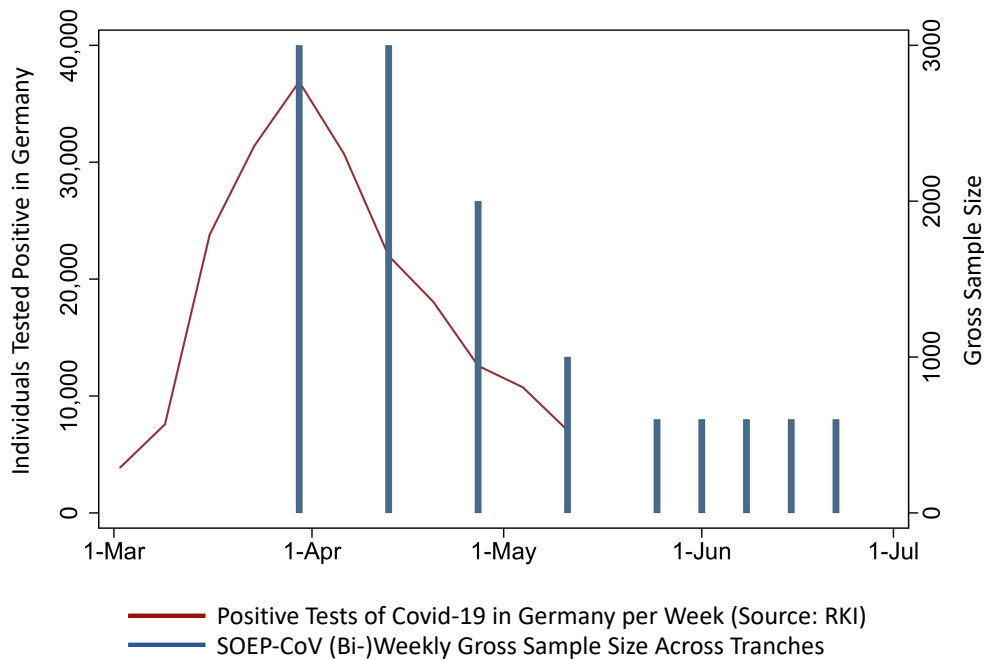


Figure 1. Sample Sizes of SOEP-CoV Tranches over the Fieldwork Period

SOEP-CoV study. This large number of weighting variables is standard in the SOEP weighting, see Siegers et al. (2020). Thus, the SOEP-CoV weighting allows joint analysis with other SOEP samples as well as longitudinal analysis. Most of the weighting variables are taken from the latest published version of the SOEP data, v35 (wave 2018). This results in a gap of about 20 months between data collection for SOEP-CoV and the SOEP wave in 2018, which may lead to a bias regarding the timeliness of the variables used for weighting. However, we consider this bias to be negligible since the weighting variables are either used to map trends or are time-invariant.

Only one person in each participating household is interviewed. The selection of this person is not random, but depends on which household member answers the phone and is willing to participate in the survey. The underlying process is modelled in a further and final nonresponse weighting step. Missing values in the weighting variables are mapped by dummy variables. In this way, the fact that a value is missing is directly incorporated into the weighting model.

Whether a household could be contacted in Tranches 1 and 2 depended on the availability of a valid landline and/or mobile phone number and the time of day at which the attempt was made to reach the household. Also, whether a household member was employed in the public sector had a positive and significant influence on contact. The data showed an increase over the two tranches in the number of interviews with households in which at least one person was an essential or frontline worker (e.g., professions referred

to in Germany as "systemically relevant" such as a doctor, police officer, nurse, etc.) and in households in which at least one member was tested for SARS-CoV-2. In general, more women and persons with at least a secondary school certificate answered the phone and agreed to be interviewed. Household members working full time, persons with a nationality other than German, and adults under 26 years of age participated less frequently in the telephone interview than their counterparts.

Weighting factors for SOEP-CoV are obtained by inverting the product of the three (participation) probabilities predicted on the basis of the household nonresponse models. Multiplied by the cross-sectional SOEP household weight of the SOEP 2020 survey and adjusted to the distribution of official statistics (Microcensus 2018) with respect to the federal state, municipality size, household size, household type, and home ownership, the resulting weights allow an extrapolation to all private households in Germany. Weights at the individual level are obtained by a further poststratification step and by multiplying the resulting (poststratified) weight by the inverted participation probability predicted on the basis of the model that describes the participation propensity of household members. The poststratification step at the individual level is made with regard to gender and age distribution, and whether a person has German citizenship or not (based on the 2018 Microcensus). At the time of the first data release for SOEP-CoV, there were no current figures for 2019 or 2020 on the distributions of German private households relevant for the SOEP available from official statistical

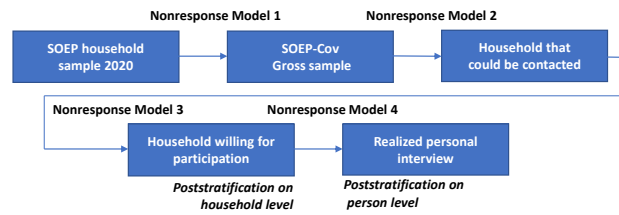


Figure 2. Steps in the Weighting Procedure of SOEP-CoV

sources. However, since the distributions of the variables we use for poststratification have not been subject to any significant changes since 2018, we consider the approach used here to be appropriate.

To sum up, SOEP and SOEP-CoV data are based on probability sampling and the weighted survey data represents a generalizable database for studying the effects of the crisis on private households in Germany.

### Data Access

SOEP-CoV data will be released as part of the standard SOEP data scientific use files version v37 (in late 2021/early 2022). Researchers can apply to the Research Data Center of the SOEP to obtain the data.<sup>9</sup>

### Analysis & Results

The SOEP-CoV project team publishes results and reports via the project website [www.soep-cov.de](http://www.soep-cov.de). Our first set of reports will focus on two main questions: 1) How is the coronavirus pandemic changing the lives of households and individuals in Germany? For this, we will add survey information from prior SOEP waves and compare current and previous survey responses at an individual and household level. And 2) What social groups are most severely affected by the pandemic? Again, rich SOEP data from previous years such as occupational biography information and household composition allow for an in-depth analysis of social consequences across a multitude of social groups.

## 5 Discussion

The coronavirus pandemic affects individuals and households in almost all aspects of life and poses major challenges to societies as a whole. Social science data and research are needed in order to effectively manage the crisis and to minimize its negative effects on individuals and society.

In this paper, we argue that longitudinal household panel data are ideally suited for research on urgent questions about the short-term and long-term effects of the coronavirus. In contrast to cross-sectional (online) surveys of individuals, household panel surveys make it possible to analyze the effects of the pandemic contextually, that is, embedded in social structures and from a life-course perspective.

The Socio-Economic Panel (SOEP) is, to our knowledge, the first of the large-scale household panel surveys to implement a supplementary questionnaire dealing with the current crisis. The ongoing SOEP-CoV study is adding up-to-date information on individuals' and households' current situation in the corona crisis to the wealth of prior information available about these households in the SOEP. This design enables researchers to study the diverse social, economic, and psychological effects of the crisis within different groups in society. The results also provide the basis for research on resilience during crises in relation to economic, human, and social capital (and inequality) accumulated in the past. The survey design of SOEP-CoV and its integration into the SOEP offers an important starting point for the introduction of corona-related studies by other large-scale panel surveys worldwide.

Future SOEP survey waves will allow researchers to study the long-term effects of the corona pandemic. Moreover, the SOEP also offers the potential for large-scale medical testing and add-ons related to public health, including antibody tests.

### Acknowledgments

The authors would like to thank all other members of the SOEP-CoV project team: Theresa Entringer, Jan Goebel, Markus M. Grabka, Daniel Graeber, Hannes Kröger, Johannes Seebauer, Carsten Schröder, and Jürgen Schupp. Special thanks also go to Florian Griesse for managing the questionnaire design, to the weighting team Rainer Siegers and Hans Walter Steinhauer, and to Gert G. Wagner for his valuable comments on our project.

We also thank the reviewers, Cristiano Vezzoni, Giulia M. Dotti Sani, Riccardo Ladini, and Francesco Molteni (University of Milan Statale, Italy), for their detailed comments and time devoted to improving the quality of this paper.

### References

Albaum, G., & Smith, S. (2012). Why people agree to participate in surveys. In L. Gideon (Ed.), *Handbook of survey methodology for the social sciences*. New York: Springer.

<sup>9</sup>See [https://www.diw.de/en/diw\\_02.c.222518.en/research\\_data\\_center\\_of\\_the\\_soep.html](https://www.diw.de/en/diw_02.c.222518.en/research_data_center_of_the_soep.html).



- Campanelli, P., & O’Muricheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality & Quantity*, 33, 59–76.
- Deaton, A. (1985). Panel data from time series of cross sections. *Journal of Econometrics*, 30, 109–126.
- Fowler, F. (2014). *Survey research methods* (5th ed.). Thousand Oaks: Sage.
- Gerstorff, S., & Schupp, J. (2016). SOEP wave report 2015. Retrieved from [https://www.diw.de/sixcms/detail.php?id=diw%5C\\_01.c.535712.de](https://www.diw.de/sixcms/detail.php?id=diw%5C_01.c.535712.de)
- Giesselmann, M., Bohmann, S., Goebel, J., Krause, P., Liebau, E., Richter, D., . . . Liebig, S. (2019). The individual in context(s): Research potentials of the Socio-Economic Panel Study (SOEP) in sociology. *European Sociological Review*, 35(5), 738–755.
- Mannheim, K. (1928). Das Problem der Generationen. *Kölner Vierteljahreshefte für Soziologie*, 7, 157–184.
- Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schütz, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in web surveys. *Social Science Computer Review*, 25, 372–383.
- Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? a review and meta-analysis. *Value in Health*, 14, 1101–1108.
- Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review*, 30, 843–861.
- Siegers, R., Belcheva, V., & Silbermann, T. (2020). SOEP-core v35 - documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP) (1984 until 2018). *DIW/SOEP: SOEP Survey Papers*, 826.
- Statista. (2020). Share of internet users in Germany from 2001 to 2019. Retrieved from <https://www.statista.com/statistics/380514/internet-usage-rate-germany/>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Wolter, K. M. (1995). *Introduction to variance estimation*. Springer.

### Commentary

Nothing in a period of emergency like the one we are experiencing is more relevant than time and timing. When sudden and unexpected events—such as the COVID-19 pandemic—occur, highly organized existing research infrastructures can really show their potential by reducing the time-lag between the new societal questions and the data able to address them. With an important addition: assurance of data quality, especially in terms of sample quality.

On these premises, the article provides a clear presentation of the SOEP-CoV survey, a telephone panel survey built

onto the existing German Socio-Economic Panel to investigate the effects of the novel coronavirus on the German population. The article is well-organized, well written and very convincing in showing the advantages of rooting a COVID-19 survey on a previously existing research program. Even more importantly, the article provides a strong argumentation in favour of enhancing existing infrastructures (and creating new ones) and asking for additional and stable funding to pursue these endeavours. As social scientists, it is imperative that we work to set and follow these standards.

There is no discussion on the relevance of the main point made by the authors, which is perfectly supported by the research they describe. For this reason, the authors could have been more daring and frank in confronting with the limitations of their approach. The impression is that they overlooked the possible weaknesses of their design and the sub-optimal choices they were forced to take.

Three points are worth mentioning. First, when discussing the weakness of cross-sectional designs, they say: “Another common challenge of online surveys is that of the difficulty of collecting contextual information. Another common problem with individual centred studies such as online surveys is that of missing contextual information.”

When such information (e.g., data about other household members) is collected, it is usually only from the person interviewed. Two problems arise here: First, it is only possible to collect proxy information about third parties inflating the risk of misreporting. Later, when describing their design: “Only one person in each participating household is interviewed. The selection of this person is not random but depends on which household member answers the phone and is willing to participate in the survey.” This choice has a number of implications. As for the representativeness of their sample, the authors surely produce a complete and convincing account of the weighting approach that allows inference to the population. Nonetheless, this does not solve the problems of having to collect information concerning third parties from a single respondent. Moreover, the difficulties connected with the administration of long questionnaires via telephone are completely neglected, although they are well established in literature, ranging from constraints on questions’ formulation to limitations on the length of the questionnaire (Albaum & Smith, 2012; Fowler, 2014)<sup>10</sup>. To the opposite, the authors mention as an advantage of their choice that “interviewers make it possible to use complex instruments such as household grids or life history calendars in data collection”, that is clearly not the case for telephone interviews.

The second issue relates to the decision of splitting their sample through the fieldwork. The authors correctly decided to follow the trajectory of the pandemic thus opting for a 9 tranches fieldwork design. Nonetheless, their decision is

<sup>10</sup>References are listed among the references of the main article.

justified only by the following statement: “The design follows our expectations of a decrease in the intensity of the crisis and its effects on respondents”. They could be right, they could also be wrong. But in any case, the decision is arbitrary and does not always easily allow for adjustments following unexpected developments. A possibility to consider could have been a more continuous design that would have allowed to follow the day-to-day evolutions (new laws, increases or decreases of restrictions, etc.) with a higher degree of precision. And without any a-priory assumption on the development of the crisis.

A final point concerns the structure of the article, in terms of space devoted to the different parts. The authors devoted almost two pages in explaining why Panel Surveys in general are very powerful tools when studying individual and household changes. There is little to argue about the strengths of panel data, but in the economy of a short commentary, this seems overstated. At the same time, the references to the specific situation in which the SOEP-CoV survey is developed, that is the COVID-19 pandemic, seems somehow instrumental. There is not a mention to specific research questions that could be better addressed with the use of such a research design in the frame of Coronavirus crisis. This would have enriched the article, as well as including an example or anticipation of the results.

To conclude, we acknowledge that such a solid infrastructure like the SOEP has proved to be very flexible to quickly respond to the unprecedented situation that societies are currently facing. This flexibility cannot be taken for granted and represents a considerable strength of the SOEP program. Therefore, this point could have been stressed with even more vigour in the article.

Cristiano Vezzoni, Giulia M Dotti Sani, Riccardo Ladini,  
Francesco Molteni  
Department of Social and Political Sciences  
University of Milan Statale