



Data Observer

Susanne Bartig, Herbert Brücker, Hans Butschalowsky, Christian Danne, Antje Gößwald, Laura Goßner, Markus M. Grabka*, Sebastian Haller, Doris Hess, Isabell Hey, Jens Hoebel, Susanne Jordan, Ulrike Kubisch, Wenke Niehues, Christina Poethko-Mueller, Maximilian Priem, Nina Rother, Lars Schaade, Angelika Schaffrath Rosario, Martin Schlaud, Manuel Siegert, Silke Stahlberg, Hans W. Steinhauer, Kerstin Tanis, Sabrina Torregroza, Parvati Trübswetter, Jörg Wernitz, Lothar H. Wieler, Hendrik Wilking and Sabine Zinn

Corona Monitoring Nationwide (RKI-SOEP-2): Seroepidemiological Study on the Spread of SARS-CoV-2 Across Germany

<https://doi.org/10.1515/jbnst-2022-0047>

Received August 9, 2022; accepted August 9, 2022

*Corresponding author: **Markus M. Grabka**, Socio-Economic Panel, German Institute for Economic Research, Berlin, Germany, E-mail: mgrabka@diw.de

Susanne Bartig, Hans Butschalowsky, Antje Gößwald, Isabell Hey, Jens Hoebel, Susanne Jordan, Ulrike Kubisch, Christina Poethko-Mueller, Angelika Schaffrath Rosario, Martin Schlaud, Silke Stahlberg and Jörg Wernitz, Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany

Herbert Brücker, Laura Goßner and Parvati Trübswetter, Institute for Employment Research (IAB) of the Federal Employment Agency (BA), Nuremberg, Germany

Christian Danne and Maximilian Priem, DIW Econ, Berlin, Germany

Sebastian Haller and Hendrik Wilking, Department of Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany

Doris Hess and Sabrina Torregroza, infas Institute for Applied Social Sciences, Bonn, Germany

Wenke Niehues, Nina Rother, Manuel Siegert and Kerstin Tanis, Research Center of the Federal Office for Migration and Refugees, Nuremberg, Germany

Lars Schaade, Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin, Germany; and Institute Leadership, Robert Koch Institute, Berlin, Germany

Hans W. Steinhauer and Sabine Zinn, Socio-Economic Panel, German Institute for Economic Research, Berlin, Germany. <https://orcid.org/0000-0003-0542-9933> (H.W. Steinhauer)

Lothar H. Wieler, Institute Leadership, Robert Koch Institute, Berlin, Germany

Abstract: SARS-CoV-2, the coronavirus, spread across Germany within just a short period of time. Seroepidemiological studies are able to estimate the proportion of the population with antibodies against SARS-CoV-2 infection (seroprevalence) as well as the level of undetected infections, which are not captured in official figures. In the seroepidemiological study Corona Monitoring Nationwide (RKI-SOEP-2), biospecimens and interview data were collected in a nationwide population-based subsample of the Socio-Economic Panel (SOEP). By using laboratory-analyzed blood samples to detect antibodies to the SARS-CoV-2 virus, we were able to identify a history of vaccination or infection in study participants. By combining these results with survey data, we were able to identify groups within the population that are at increased risk of infection. By linking the RKI-SOEP-2 survey data with data from other waves of the SOEP survey, we will be able to examine the medium- to long-term impacts of the COVID-19 pandemic, including effects of long COVID, in diverse areas of life. Furthermore, the data provide insight into the population's willingness to be vaccinated as well as related attitudes and conditions. In sum, the RKI-SOEP-2 survey data offer a better understanding of the scope of the epidemic in Germany and can help in identifying target groups for infection control in the present and future pandemics.

Keywords: SARS-CoV-2, COVID-19, seroepidemiological study, seroprevalence, vaccine status, willingness for vaccination, SOEP

JEL Classification: I12, I19, C89

1 Introduction

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), the coronavirus, was first detected in Germany in late January 2020. Since then, the country has experienced several waves of infection. As of March 15, 2022, a total of 17,432,617 infections had been officially recorded by the Robert Koch Institute (RKI 2022a). This figure is the number of laboratory-confirmed infections reported to health authorities nationwide within the legal framework of the Infection Protection Act (IfSG) and compiled by the Robert Koch Institute (RKI). As a significant proportion of those infected with SARS-CoV-2 have an asymptomatic course of infection, it is highly likely that the official figures are below the actual number of infections (Neuhauser et al. 2021). The magnitude of underreporting may be substantial, even with expanded testing strategies. In addition, the prioritization of PCR testing in early 2022 (BMG 2022) due to expected bottlenecks in laboratory capacity may have resulted in additional undetected infections in non-priority populations. Population-based cohort studies with serological determination of antibodies against SARS-CoV-2 enable scientists to detect previously unknown past infections and thus estimate the number of unreported infections. Moreover, they allow for

observation of antibody decay after a period of time according to relevant criteria such as symptoms and age. In addition, if the individual's vaccination status is recorded in a population-based cohort study, it is possible to examine the association between vaccination and individual antibody concentration as well as overall antibody prevalence.

A large number of seroepidemiological studies have been conducted in Germany to date, but most of them focus on a specific region or city and often on specific population groups (e.g., health care workers, blood donors). Another limitation relates to the study population, as the majority of target subjects are adults. Besides this, most studies are cross-sectional; only a small minority is panel studies.¹ Only two seroepidemiological studies conducted in Germany up to now have aimed to provide nationwide coverage and target the general population. One was the Corona-Bund Study conducted by a consortium led by the Ifo Institute in cooperation with Forsa (IFO Institute & Forsa 2022). However, the sample frame was an access panel (Dülmer et al. 2005), which can entail distortions in extrapolating the results.² The second was the predecessor to the present study, RKI-SOEP-1 (Hoebel et al. 2021), which was conducted by the RKI together with the SOEP and the IAB.

The second wave of the study Corona Monitoring Nationwide (RKI-SOEP-2) is a nationwide population-based cohort study. It is a cooperative project of the following institutions: the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin), the Robert Koch Institute (RKI), the Institute for Employment Research (IAB), and the Research Center of the Federal Office for Migration and Refugees (BAMF-FZ).³ Participants in the study are SOEP respondents aged 14 years and older. In addition to respondents from the regular SOEP, this study also includes individuals with an immigration background from the IAB-SOEP Migration Sample as well as refugees from the IAB-BAMF-SOEP Survey of Refugees. All SOEP respondents were invited to participate in the special survey (RKI-SOEP-2) after the regular 2021 SOEP wave.

In addition to a short questionnaire, respondents were given a test kit for the collection of capillary blood to detect immunoglobulin G (IgG) antibodies against the spike protein and the nucleocapsid protein of SARS-CoV-2. Antibodies against the spike protein may result either from an infection or from vaccination. Antibodies against the nucleocapsid protein, in contrast, result only from an infection.

¹ For an overview of the various seroepidemiological studies, see RKI (2022b).

² According to the RatSWD, microdata from this study are not available to the scientific community (<https://www.konsortswd.de/ratswd/themen/krisen/corona/301/>).

³ The predecessor study (SOEP-RKI) was conducted from October 2020 to March 2021 and is described in detail in Hoebel et al. (2021).

By examining both types of antibodies, antibody responses resulting from vaccination can theoretically be distinguished from those resulting from infection with SARS-CoV-2. Because SOEP is a multifaceted household panel study, longitudinal information on respondents is also available on a wide range of individual and household characteristics. These allow, among other things, in-depth analysis of intra- and inter-individual associations with infection and vaccination status.

The primary objectives of the Corona Monitoring Nationwide (RKI-SOEP-2) study were to investigate:

- 1) the seroprevalence, that is, the proportion of the population in Germany in which IgG antibodies against SARS-CoV-2 were detectable,
- 2) the extent of undetected SARS-CoV-2 infections,
- 3) risk and protective factors for SARS-CoV-2 infection, considering demographic, socioeconomic, and health-related factors,
- 4) symptomatology in the presence of long COVID disease, and
- 5) vaccination status or willingness to be vaccinated

in a nationwide sample of the general population aged 14 years and older in Winter 2021/22.

Through the inclusion of the IAB-SOEP and the IAB-BAMF-SOEP samples, the Corona Monitoring Nationwide (RKI-SOEP-2) study oversamples refugees and other individuals with an immigration background. This means that the aforementioned research objectives can be pursued not only in studies focusing on specific immigrant groups, but also in studies comparing immigrant groups and German nationals. Corona Monitoring Nationwide (RKI-SOEP-2) also lays the groundwork for future research on medium- to long-term consequences of SARS-CoV-2 infection by making the microdata collected as part of the study available to the broader scientific community.

This paper is structured as follows: The study design and the study process, including a description of the questionnaire and data protection procedures, are presented in Section 2. Section 3 describes the gross sample of selected households and target persons as well as the net sample of surveyed respondents. Section 4 offers a discussion of the research potentials and limitations of the study. Section 5 describes data access, provides a summary and presents an outlook for future research.

2 Method

2.1 Study Design

Corona Monitoring Nationwide (RKI-SOEP-2) is a population-based, epidemiological cohort study. Its units of observation are participants in the 2021 annual

SOEP wave and their household members aged 14 and older. The SOEP is a longitudinal survey of private households in Germany and all individuals living in them, conducted annually since 1984 (Göbel et al. 2019). The survey covers a wide range of topics, from demographics, income, the labor market, education, and health to general attitudes, concerns, and life satisfaction.

The SOEP consists of different sub-samples, which are either random samples of the total population or subsamples of selected population groups, including families, high-income earners, and refugees and other immigrants. The IAB-SOEP sample is one of the SOEP's migration samples. It consists of immigrants who entered Germany between 1995 and 2013 as well as second-generation immigrants (Kroh et al. 2015). The IAB-BAMF-SOEP study is comprised of refugees who entered Germany between 2013 and 2020 and applied for asylum (Kühne et al. 2019; Steinhauer et al. 2022), most of whom are from Syria, Iraq, and Afghanistan. The exact sample compositions, participation propensities, and (non-)response rates for the various SOEP samples since it began in 1984 are described in detail in Siegers et al. (2022). Overall, the SOEP had a high repeat participation rate of 85% up to 2020, which fell to just under 70% in 2021. The CAPI-by-Phone survey mode, introduced after the start of the pandemic (the standard is CAPI), and the change of survey institute from Kantar Public to infas were probably the two main factors behind this drop.⁴ The advantage of a panel study like the SOEP is that rich information is available from previous waves to compensate for selective dropouts in statistical analyses of the entire population. Based on the extensive information available in previous waves, the SOEP therefore provides methodologically high-quality weighting factors for each annual survey wave (for more details, see Siegers et al. 2022). In addition, SOEP counters panel attrition through regular, large-scale refreshers (the next one in 2022).

In the Corona Monitoring Nationwide (RKI-SOEP-2) study, individuals were invited who:

- lived in a SOEP household that participated in the 2021 annual SOEP survey
- were 14 years or older (at the time of the survey)
- gave their written consent to participate in the study
- were able to take a capillary blood sample on their own (self-sampling)
- were able to read the study instructions and information in German, Arabic, Farsi, English, Polish, Bulgarian, or Romanian.

⁴ More in-depth analyses of the change in willingness to participate in the survey are underway.

2.2 Study Process

2.2.1 Recruitment of Participants

All SOEP households in the gross sample were invited to participate in the study in a letter sent by mail. The fieldwork was conducted by the infas Institute for Applied Social Sciences.⁵

First, a letter was mailed to each target household informing them about the study and letting them know that they would be receiving a personal invitation to take part. Shortly thereafter, an invitation packet was sent to each target respondent. The packets contained both the personal invitation and the study materials (flyer about the study, motivation letter from the president of the Robert Koch Institute, data protection declaration, consent form, participation plan, questionnaire, blood self-sampling kit including instructions and packaging materials for safe shipment, as well as two return envelopes: one for the consent form and the questionnaire, and one for the blood sample). From previous waves of the SOEP survey, we knew the main languages used by the individuals who were invited to participate. Based on this information, the aforementioned documents were sent out in seven different languages: German, Arabic, Farsi, English, Polish, Bulgarian, and Romanian.

To increase participation in the study, the target respondents were informed that after participating in the study, they would receive written notification of their laboratory results, thus providing them with information about their own antibody status. In addition, they were informed that they would receive a monetary incentive (10 euros for adults, 5 euros for adolescents) after returning the study documents. If they did not respond, a reminder was sent two to three weeks after the first invitation. In addition, the interviewer responsible for the household contacted the target respondents to answer any open questions and urge them to participate. The fieldwork phase began with the mailing of the first informational letters in the 37th calendar week of 2021. The first invitation packets, containing the invitations to each person in the household and the study materials, were mailed out in the 45th calendar week of 2021. Fieldwork extended through March 6, 2022.

For logistical reasons, the gross sample was divided into three tranches that were invited sequentially. Since the main SOEP survey did not begin until the second quarter of 2021 and since not all households had been interviewed by the start of the Corona Monitoring Nationwide (RKI-SOEP-2) study, tranching followed the schedule for processing the SOEP households in the main survey. The aim was to achieve a certain time gap between participation in the main survey and

⁵ For more information see <https://www.infas.eu/>.

participation in the Corona Monitoring Nationwide (RKI-SOEP-2) study. As a result, the gross sample of the Corona Monitoring Nationwide (RKI-SOEP-2) study was divided into three tranches in a ratio of approximately 60%/10%/30%. The gross sample consisted of 21,456 individuals.

2.2.2 Data Collection

The study consisted of a questionnaire part and a blood test part. In the questionnaire part, respondents completed a survey asking for information about previous infections, vaccination against the virus, attitudes towards vaccination, and health issues. The questionnaire could either be completed in written form or using the computer-assisted web interviewing (CAWI) survey technique. In the blood test part, respondents collected a capillary blood sample by pricking a fingertip and allowing the blood to dry on a collection card. We used the CE-certified sample collection and submission kits produced by the manufacturer Euroimmun AG (Lübeck, Germany) (Figure 1).

The blood collection kit contained detailed illustrated instructions with written explanations under each picture, a blood collection card, a compress, two adhesive bandages, two alcohol swabs, two sterile lancets, and a sealable plastic bag with desiccant. There were five circles on the blood collection card as markers for the blood samples. The illustrated instructions included with the sample collection kit also provided a web link and QR code leading to videos showing how sample collection was to be done. In addition, respondents were directed to a study-specific webpage with frequently asked questions (FAQ) and answers. Respondents were cautioned not to collect a blood sample if there were any issues or acute health conditions that might skew the results or make it painful to collect a sample.

Respondents were asked to send their blood collection cards by mail to the RKI, if possible on the day of sample collection, and to send the completed questionnaire and signed consent form to infas. The materials that respondents returned to the RKI and infas (blood collection cards, questionnaires, and consent

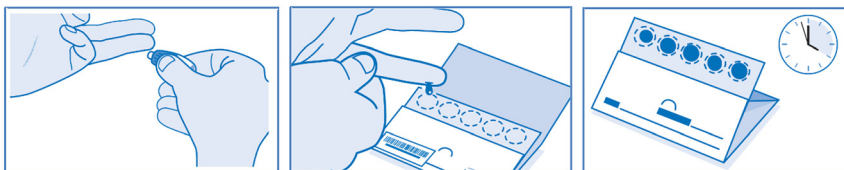


Figure 1: Self-collection of capillary blood from a Fingertip. Source: Robert Koch Institute.

forms) had a machine-readable bar code printed on them to enable linkage of laboratory and survey results at a later point in time.

2.2.3 Notifications of Results

Written reports of individual laboratory results were sent by mail to the participants. For questions regarding the results, respondents were directed to the study-specific website providing FAQ, a form to contact medical study staff, and the number of a telephone hotline.

2.2.4 Quality Assurance

Quality assurance measures were integral to the entire study process and were carried out by all project partners (Infas, RKI, SOEP, IAB, and BAMF-FZ). Thanks to the close monitoring of fieldwork during the study and weekly meetings of all project partners, it was possible to identify unexpected developments quickly and respond appropriately (e.g., by extending fieldwork to increase response rates). Minutes were taken at all project meetings and made available to the entire study staff.⁶ The RKI's Central Epidemiological Laboratory is accredited according to the international standards ISO/IEC 17025 for testing and calibration laboratories and ISO 15189 for medical laboratories and is therefore subject to the highest quality standards. Compliance with these standards was ensured through continuous internal and external comparative measurements (round robin tests) as well as independent audits. In addition, extensive data quality controls were carried out. The focus here was on ensuring the protection of study participant data. Study and quality assurance processes were conducted on a random basis, and study documents were examined.

2.3 Laboratory Analytics

The analysis of the dried blood samples was carried out in the Central Epidemiological Laboratory at the RKI. Dried blood spots were punched out⁷ and extracted from the filter paper of the blood collection card according to the manufacturer's

⁶ Infas is certified according to the international standard ISO 20252 for market, opinion, and social research and ensured the quality of fieldwork according to the specifications of this standard.

⁷ A "DBS Puncher 1296-071" from PerkinElmer (Waltham, MA, USA) was used.

instructions. Analyses were performed automatically on a high-throughput analyzer “EUROLab Workstation ELISA” from the manufacturer Euroimmun. Blood samples were tested for immunoglobulin G (IgG) antibodies against two different antigenic structures of SARS-CoV-2 using two commercial enzyme-linked immunosorbent assays (ELISA) from Euroimmun: The “Anti-SARS-CoV-2-NCP ELISA” was used to test for the presence of antibodies against the nucleocapsid protein (NCP) (manufacturer’s specifications from serum samples, dated November 9, 2021: sensitivity 94.6%: > 10 days after symptom onset or direct pathogen detection by PCR analysis; specificity 99.8%). Measurement results were semi-quantitative, given as ratios, and divided into the three categories “positive” (≥ 1.1 ratio), “borderline” (≥ 0.8 to < 1.1 ratio) or “negative” (< 0.8 ratio) according to the manufacturer’s specifications for serum samples.

The “Anti-SARS-CoV-2-QuantiVac ELISA” was used for the quantitative detection of antibodies against the S1 domain of the virus’s spike protein. This test was used for the indirect detection of either a previous infection with SARS-CoV-2 or a vaccination with an S1/RBD-based vaccine (manufacturer’s specification from serum samples, dated November 11, 2021: sensitivity 90.3%: > 10 days after symptom onset or positive direct pathogen detection; specificity 99.8%). Measurement results were given in relative units (RU/mL) and classified according to the manufacturer’s specifications for serum samples into the three categories: “positive” (≥ 11 RU/mL), “borderline” (≥ 25.6 to < 35.2 RU/mL), or “negative” (< 25.6 RU/mL). For better comparability with the results of test systems from other manufacturers, the results were converted into “Binding Antibody Units per mL” (BAU/mL) according to the First WHO International Standard for anti-SARS-CoV-2 immunoglobulin, NIBSC code 20/136. For this purpose, the relative units were multiplied by 3.2 according to the manufacturer’s instructions. Blood samples that tested above the upper detection limit of the test (120 RE/mL or 384 BAU/mL) were diluted tenfold and quantified with the “Anti-SARS-CoV-2-QuantiVac-ELISA 1010.”

For both the “Anti-SARS-CoV-2-NCP ELISA” and the “Anti-SARS-CoV-2-QuantiVac ELISA,” the manufacturer reports very high levels of agreement between results from capillary dry blood samples and venous serum blood samples: 99.5%; (positive percent agreement: 98.4%; negative percent agreement: 100%; $n = 215$), and 100% (positive percent agreement: 100%; negative percent agreement: 100%; $n = 26$ and $n = 85$ in a second study), respectively.⁸ Borderline results were not considered for either test.

⁸ Additional internal validation procedures were carried out by the RKI Laboratory for quality assurance.

2.4 Questionnaire Content

In the questionnaire part of the study, respondents completed a 12-page questionnaire. Table 1 summarizes the topics covered:

Table 1: Overview of questionnaire topics.

SARS-CoV-2 infection

- Previous infection detected by PCR testing
- Date of positive test result
- Severity of illness

Vaccination against COVID-19

- Vaccination status
- Date and site of vaccination
- Vaccine type
- Intention to be vaccinated and barriers to vaccination
- Attitude towards COVID-19 vaccination

Health

- Subjective health status
- Physical functioning
- Physical and mental limitations
- Change of health status since pandemic
- Symptoms of illness in the last six months and their impact on different areas of life

Use of health services

- Doctors' visits in the last three months
- Hospitalization in the current year

Health behavior

- Weekly amount of walking/cycling (active transport)
- Changes in active transport since pandemic
- Motives for active transport
- Use of protection measures against COVID-19

Pets

- Pets in household

Health information-seeking regarding the pandemic

- Perceived informedness
- Information sources used
- Information sources that influenced the vaccination decision
- Language of information sources used

Impact of the pandemic on daily life

- Perceived risk of infection
- Personal burden in different areas of life

Assessment of government's response to the pandemic

- Satisfaction with governmental pandemic management
 - Exaggeration of risks in the media
-

Wherever possible, survey items were designed such that comparability with other studies and commonly used items, including those from regular SOEP waves, was ensured. Study participants whose blood sample was not valid were asked to repeat their blood test. In this case, they also received another short two-page questionnaire so that the most time-sensitive survey information was current at the date of their newly taken blood sample. The survey items in this short version covered the topics of SARS-CoV-2 infection and vaccination.

2.5 Data Protection and Ethics

Similar to other studies conducted by RKI, IAB, BAMF, and SOEP, the Corona Monitoring Nationwide (RKI-SOEP-2) study was subject to strict data protection provisions set out in the EU General Data Protection Regulation (GDPR) and the Federal Data Protection Act (BDSG). The ethics committee of the Berlin Chamber of Physicians assessed the ethics of Corona Monitoring Nationwide (RKI-SOEP-2) study and provided approval (Eth-33/20). Participation in the study was voluntary. The participants were informed about the aims and contents of the study, how their data would be handled, and when the data would be deleted. Respondents provided written informed consent.

3 Sample Description

3.1 Gross Sample

The gross sample of the Corona Monitoring Nationwide (RKI-SOEP-2) study consisted of 12,101 households (containing 21,456 target respondents) in which at least one household member had completed a household or personal questionnaire in the regular 2021 wave of the main SOEP study (SOEP-Core). Of the 12,101 invitation letters sent out for RKI-SOEP-2, 11,785 reached households in the gross sample.⁹ Within these households that were invited, a total of 20,774 persons aged 14 and older were asked to participate in the study. Of these, 1737 were under 18 years of age. The maximum age of a person invited to participate was 102. Of those invited, 12,407 individuals were assigned to the first tranche, 1984 to the second tranche, and the remaining 6383 to the third tranche. 1950 persons (including 217 minors) had a refugee background and 4203 were born abroad (with 280 minors). 51.4% of

⁹ The remaining households could not be reached because they had moved abroad or to an unknown address or all household members were deceased.

all individuals in the gross sample were female and 48.6% male, and 23.0% lived in East Germany (including Berlin). Overall, the (invited) sample covers 400 counties and county-level cities (NUTS-3), ranging from NUTS-3 areas with 1–944 target respondents. On average, a NUTS-3 area contains 52 target persons with a standard deviation of 67 persons. A household in the RKI-SOEP-2 gross sample includes between one and nine individuals, with an average of 1.8 individuals aged 14 and older and a standard deviation of 0.9.

The RKI-SOEP-2 sample originated from a complex study design aimed at sampling households within many regional units distributed across Germany and at oversampling certain population groups (e.g., immigrants, refugees, families with many children). The willingness of these households to participate in subsequent SOEP waves, and thus to be included in the RKI-SOEP-2 sample, varied across population groups. For example, people with an immigrant background and those in full-time employment have been shown to have a lower willingness to participate in subsequent waves than people who do not have an immigrant background or are not in full-time employment (Siegers et al. 2022). This study therefore has a higher sampling error than a pure random sample in which households or individuals are drawn from a nationwide list and all participate. Adjustment factors (sampling weights) are an appropriate means of compensating for the resulting deviations in population statistics. The design effect quantifies the increase in sampling error caused by doing so (Kish 1965). For the SOEP weights, the estimated design effect is between 3 and 3.5. This means that the gross sample size of 20,774 persons corresponds to an effective sample size of 5935–6925 persons. That is, in terms of sampling error, this is the sample size we would have gotten from a pure random sample without special design features such as over-sampling or attrition and non-response over the course of the panel.

3.2 Response Rates and Weighted Sample Statistics

A total of 11,162 persons aged 14 years or older participated in the Corona Monitoring Nationwide (RKI-SOEP-2) study. Using the standards of the American Association for Public Opinion Research (AAPOR 2016), the calculated response rate was 53.7% for the total sample and 51.9% for the 14–69 age group. Table 2 shows the detailed response rates and characteristics of the RKI-SOEP-2 net sample. The response rate of female participants was generally slightly higher (56.3%) than that of male participants (51.1%). With the exception of those over 80 years of age, the response rate increased with age from 41.0% in the youngest age group (14–17 years) to 68.9% in the 65–79 age group. Using the CASMIN (Comparative

Table 2: Response rates and characteristics of the RKI-SOEP-2 net sample.

	RR	<i>n</i> ^a	% ^b
Total (14+ years)	53.73%	11,162	—/—
Total (14–69 years)	51.92%	9462	—/—
Sex			
Women	56.28%	5999	50.78
Men	51.05%	5158	49.20
Non-binary (<i>divers</i>)	(55.56%)	5	0.02
Age group ^c			
14–17 years	40.97%	712	8.47
18–34 years	41.84%	1935	22.17
35–49 years	50.28%	2478	20.52
50–64 years	61.55%	3488	25.54
65–79 years	68.87%	2091	18.16
80+ years	58.34%	458	5.14
Education ^d			
Low (CASMIN 1)	44.42%	1820	20.77
Middle (CASMIN 2)	59.11%	3936	33.89
High (CASMIN 3)	68.51%	3375	26.21
Missing/not classifiable ^e	39.89%	2031	19.13
Household size			
1 person	50.78%	2734	35.66
2–4 persons	55.48%	7989	62.15
5+ persons	44.25%	439	2.19
Migration status ^d			
German-born	60.52%	9920	84.41
Foreign-born, no refugee background	37.25%	840	13.17
Refugee background	17.09%	333	2.41
Date of participation ^f			
November–December 2021	—/—	7769	78.49
January–February 2022	—/—	3393	21.51
Federal state			
Schleswig–Holstein	50.17%	437	3.67
Hamburg	52.88%	248	2.25
Lower Saxony	52.44%	1042	9.44
Bremen	47.97%	71	0.78
North Rhine–Westphalia	50.26%	2248	21.28
Hesse	52.25%	754	7.59
Rhineland–Palatinate	50.37%	410	4.34
Baden–Württemberg	58.24%	1386	13.85
Bavaria	55.46%	1762	15.84
Saarland	45.49%	111	1.49
Berlin	56.83%	537	5.07
Brandenburg	52.94%	423	2.42
Mecklenburg Western-Pomerania	55.06%	234	1.60

Table 2: (continued)

	RR	<i>n</i> ^a	% ^b
Saxony	60.90%	757	5.20
Saxony–Anhalt	53.15%	354	2.58
Thuringia	56.23%	388	2.62

RR = Response Rate 6 according to AAPOR; *n* = crude number of participants; ^aDue to a very small number of missing values, the case numbers for some variables do not always add up to the total case number, ^bweighted percentage, ^cmeasured at date of first contact (letter introducing the study), ^daccording to data from SOEP wave 2020 (v37), ^enot classifiable: e.g., refugees with foreign qualifications not included in CASMIN, ^fmeasured by the date the dry blood spot sample was received by the laboratory, if no test was sent to the laboratory, the date of questionnaire completion.

Analyses of Social Mobility in Industrial Nations) educational classification, the highest response rate was observed among highly educated respondents (68.5%) followed by the middle education group (59.1%) and the low education group (44.4%). Willingness to participate was greater in households with two to four persons (55.5%) than in one-person households (50.8%) or in households with more than five persons (44.3%). Furthermore, response rates differed by migration status. The response rate among German-born respondents was significantly higher (60.5%) than among foreign-born respondents without a refugee background (37.3%) and also significantly higher than among foreign-born respondents with a refugee background (17.1%). Response rates varied by federal state between 45.5% (Saarland) and 60.9% (Saxony).

In the weighted sample (see Section 3.3, Weighting), 50.8% of the respondents were female and 49.2% were male. About quarter (25.5%) of respondents were between 50 and 64 years old, and a slightly lower percentage were in the 18–34 age group (22.2%) and in the 35–49 age group (20.5%). Significantly fewer respondents were in the youngest (14–17) and oldest (80 years and older) age groups. The majority belonged to the middle education group (33.9%); about a quarter to the high education group (26.2%); and slightly fewer to the low education group (20.8%). Around 63% of respondents lived in a household with two to four persons and over one third (35.7%) in one-person households. Only a very small minority of respondents (2.4%) were foreign-born with a refugee background. The vast majority (84.4%) was German-born, and 13.2% were foreign-born without a refugee background. The regional coverage of the sample largely corresponded to the expected distribution across the different states. Of the 11,162 respondents, more than three quarters (78.5%) took part in the study in the period November to December 2021.

3.3 Weighting

For the Corona Monitoring Nationwide (RKI-SOEP-2) study, appropriate weights were created. The weighting strategy considered the different selection processes (contactability and participation) from the gross sample to the net sample of respondents at both the household and the individual level separately. For each selection process at each level (e.g., the process of determining who qualified for selection), a model was estimated to derive the corresponding probabilities used for the adjustment of weights. The variables that may enter these models covered topics such as socio-demographics, household composition, regional characteristics, health status, worries and attitudes regarding political and societal issues, and others. Variables best describing the selection processes were identified using lasso regression (Tibshirani 1996). On the household as well as the individual level, weights were adjusted to conform to known population distributions provided by the Federal Statistical Office. On the household level, these included household typology, size, home ownership, and federal state, and on the individual level, they included age, sex, and citizenship (German vs. non-German). A more detailed version of the weighting strategy is provided for the Corona Monitoring Nationwide (RKI-SOEP-2) study by Danne et al. (2022).

4 Discussion

4.1 Research Potential and Strengths of the Study

The second wave of the Corona Monitoring Nationwide (RKI-SOEP-2) study provides data on topics relevant to public health after two years of the SARS-CoV-2 pandemic. The main aim of the study was to estimate the percentage of the population that had been either infected by SARS-CoV-2 or vaccinated by February 2022, differentiated by social groups (e.g., age, sex, educational level). The study data thus allow for identification of groups at risk of SARS-CoV-2 infection. The prevalence of vaccination, attitudes towards vaccination, health information-seeking behavior, and conditions reported that influenced decision-making will also provide a better understanding of the epidemiological status quo. Appraising the expected percentage of long COVID cases and analyzing the medium- to long-term effects of the pandemic on health conditions, health behavior, and individual socio-economic situations will ultimately make it possible to identify areas and groups requiring special policy measures in the future.

Because of the extensive measures undertaken to include SOEP respondents from across Germany as well as a wide variety of socio-economic living conditions, the study data provide the basis for differentiated analysis and identification of specific subgroups that are more severely affected by the pandemic. The fact that all of the information from this study can be linked to respondent data from earlier or later waves of the SOEP offers extensive potential for analysis of the medium- to long-term consequences of the COVID-19 pandemic.

In sum, the results of this study will provide important information to government bodies and other stakeholders that will be useful in evaluating the management of the pandemic thus far and in shaping future responses to this and other pandemics.

4.2 Limitations

Besides its strengths, the Corona Monitoring Nationwide (RKI-SOEP-2) study also has some limitations. First, individuals were only invited to participate if they had taken part in the regular 2021 wave of the SOEP-Core survey. Since fieldwork was still ongoing for the SOEP and particularly for the IAB-BAMF-SOEP Refugee Study during the fieldwork for this study, only a limited number of refugees could be included. This might to some extent result in selection bias, particularly for participants from refugee subsample. Second, response rates among immigrants and refugees were lower than among non-immigrants. These lower response rates were probably driven regular SOEP respondents' lack of experience with written surveys that they have to return by mail: The regular waves of the SOEP are mainly carried out as face-to-face interviews. Another limitation results from the fact that panel studies are usually subject to healthy volunteer bias, meaning that people who are suffering from a severe disease are less likely to participate (e.g., Schnell and Heller 2000; Zheng et al. 2020), potentially leading to an underestimation of the severely diseased population. Although separate weighting was conducted for this study, taking into account a wide variety of selection processes, it is possible that some amount of bias due to unobserved variables such as aversion to the study topic may not have been fully corrected. Moreover, data from the current 2021 SOEP wave were not yet available for the weighting process.

Limitations are also to be expected on the laboratory side. An inherent limitation of IgG antibody studies as a basis for prevalence estimates of SARS-CoV-2 immunity is that the presence of IgG antibodies does not mean full immunity, and a lack of IgG antibodies does not mean a complete lack of protection. However, serological studies may provide an estimate of the proportion of the population that has been in contact with the virus antigen(s) and may have developed at least

some degree of immunity or be less likely to become severely diseased after infection. In this regard, a few factors need consideration. One of these is that study participation may depend on infection and vaccination status. This is addressed through careful non-response analyses and appropriate weighting algorithms.

Antibody development (seroconversion) depends on the severity of the disease or the type of vaccine, on age, and on preexisting health conditions (e.g., Wei et al. 2021). Antibody tests and questionnaire data do not have perfect sensitivity and specificity (Scheinlauer et al. 2022), and there is some antibody decay over time (Achiron et al. 2021). This is also of relevance to this study, because the fieldwork period lasted for about four months, and during this time, the infection process across Germany was highly dynamic. This impairs accurate determination of seroprevalence. Furthermore, the previously used estimation methods (e.g., Hoebel et al. 2021) may require further adjustment since they apply to pre-Omicron variants. Data on the kinetics of antibody response after SARS-CoV-2 Omicron infection is still scarce. First results of neutralization profiles against the Omicron BA.1 variant in comparison to former SARS-CoV-2 variants indicate different antibody profiles/kinetics depending on preexisting SARS-CoV-2 (vaccine- or infection-induced) immunity after infection with Omicron (Rössler et al. 2022).

5 Data Access and Summary

The microdata from the Corona Monitoring Nationwide (RKI-SOEP-2) study are provided exclusively for scientific research and are free of charge. Use of the data for commercial purposes is prohibited. Due to the different levels of sensitivity of the various data under data protection law (survey data versus health-related laboratory data), different forms of data access are available. The survey data can be ordered from the SOEP Research Data Center and are available with v38 and subsequent SOEP data releases. After signing a data distribution contract, users can download the data using an individual download link. The link is time-limited, encrypted, and can only be used in conjunction with a personal password. For data protection reasons, the health-related laboratory data can only be used at the SOEP guest workstations at DIW Berlin. The microdata from the Corona Monitoring Nationwide (RKI-SOEP-2) study include a unique person identifier (PID) that makes it possible to link the information from this study with survey data from previous or subsequent SOEP waves.

The information from the Corona Monitoring Nationwide (RKI-SOEP-2) study provides a unique data set for analyzing infection with or vaccination against the SARS-CoV-2 virus in the population 14 years of age and older through early 2022. These new data also provide a basis for identifying groups at higher risk of

infection. The linkage of the data with other SOEP survey waves also makes it possible to examine the medium- to long-term consequences of COVID-19, including long COVID, in diverse areas of life.

Acknowledgments: We thank our colleagues from the SOEP, the Robert Koch Institute Departments 2 and 3, the Centre for Biological Threats and Special Pathogens, the Federal Office for Migration and Refugees, and the Institute for Employment Research of the Federal Employment Agency for their support and cooperation. Special thanks go to the staff of DIW Econ GmbH for carrying out the weighting. We also thank the employees of the infas Institute for Applied Social Sciences who contributed to the planning and implementation of fieldwork and data collection, as well as Alexander Krönke, Annett Klingner, and Jörg Schaarschmidt for creating illustrated instructions for self-collection of capillary blood from the fingertip (Figure 1). Finally, we are grateful to all of the study respondents for their willingness to participate in and support this study.

Research funding: The Corona Monitoring Nationwide (RKI-SOEP-2) study was funded by the German Federal Ministry of Health.

Conflicts of interest: The authors declared no conflicts of interest.

References

- Achiron, A., Gurevich, M., Falb, R., Dreyer-Alster, S., Sonis, P., and Mandel, M. (2021). SARS-CoV-2 antibody dynamics and B-cell memory response over time in COVID-19 convalescent subjects. *Clin. Microbiol. Infect. Suppl.* 27: 1349.e1–1349.e6.
- American Association for Public Opinion Research (AAPOR) (2016). *Standard definitions: Final disposition codes of case codes and outcome rates for surveys*, 9th ed. AAPOR, Alexandria, VA. Available at: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.
- Bundesministerium für Gesundheit (BMG) (2022). Fragen und Antworten zu COVID-19 Tests. Available at: <https://www.bundesgesundheitsministerium.de/coronavirus/nationale-teststrategie/faq-covid-19-tests.html>.
- Danne, C., Priem, M., and Steinhauer, H.W. (2022). SOEP-Core – 2021: Sampling, Nonresponse, and Weighting in Wave 2 of Living in Germany – Nationwide Corona-Monitoring (RKI-SOEP2). SOEP Survey Papers Series C, No. 1175.
- Dülmer, H., Klein, M., Ohr, D., Quandt, M., and Rosar, U. (2005). Die Kölner Wahlstudie 2002: Einordnung, Konzeption und Datenbasis. In: Güllner, M., Dümer, H., Klein, M., Ohr, D., Quandt, M., Rosar, U., and Klingemann, H.D. (Eds.), *Die Bundestagswahl 2002: Eine Untersuchung im Zeichen hoher politischer Dynamik*. VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 7–14.
- Göbel, J., Grabka, M.M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *J. Econ. Stat.* 239: 345–360.

- Hoebel, J., Busch, M.A., Grabka, M.M., Zinn, S., Allen, J., Gößwald, A., Wernitz, J., Goebel, J., Steinhauer, H.W., Siegers, R., et al. (2021). Seroepidemiological study on the spread of SARS-CoV-2 in Germany: study protocol of the “CORONA-MONITORING Bundesweit” study (RKI-SOEP study). *J. Health Monit.* 6: 1–17.
- IFO Institut and Forsa (2022). Die Deutschen und Corona. Schlussbericht der BMG-“Corona-BUND-Studie”, <https://www.ifo.de/DocDL/studie-2020-corona-bund-endbericht.pdf> (accessed 27 November 2020).
- Kish, L. (1965). *Survey sampling*. New York: Wiley.
- Kroh, M., Kühne, S., Goebel, J., and Preu, F. (2015). The 2013 IAB-SOEP Migration Sample (M1): sampling design and weighting adjustment. SOEP Survey Papers No. 271.
- Kühne, S., Jacobsen, J., and Kroh, M. (2019). Sampling in times of high immigration: the survey process of the IAB-BAMF-SOEP survey of refugees. *Survey Methods: Insights from the Field*.
- Neuhauser, H., Schaffrath Rosario, A., Butschalowsky, H., Haller, S., Hoebel, J., Michel, J., Nitsche, A., Poethko-Müller, C., Prütz, F., Schlaud, M., et al. (2021). Germany’s low SARS-CoV-2 seroprevalence confirms effective containment in 2020: Results of the nationwide RKI-SOEP study. medRxiv. Available at: <https://www.medrxiv.org/content/10.1101/2021.11.22.21266711v1.full>.
- Robert Koch-Institut (RKI) (2022a). COVID-19: Fallzahlen in Deutschland und weltweit. Fallzahlen in Deutschland. Data from Tuesday March 15, 2022, 00:00 a.m. (updated online at 7:40 a.m.). Available at: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html.
- Robert Koch-Institut (RKI) (2022b). Forschung zu Coronavirus SARS-CoV-2: Seroepidemiologische Studien in Deutschland (Data from: June 20, 2022). www.rki.de/covid-19-ak-studien (accessed 5 August 2022).
- Rössler, A., Knabl, L., von Laer, D., and Kimpel, J. (2022). Neutralization profile after recovery from SARS-CoV-2 Omicron Infection. *N. Engl. J. Med.* 386: 1764–1766.
- Schnell, R. and Heller, G. (2000). The Choir Invisible: zur Analyse der gesundheitsbezogenen Panelmortalität im Sozio-Ökonomischen Panel SOEP. In: Helmert, U., Bammann, K., Voges, W., and Müller, R. (Hrsg.), *Müssen Arme früher sterben? Soziale Ungleichheit und Gesundheit in Deutschland*. Weinheim: Juventa Verl, pp. 115–134.
- Siegers, R., Steinhauer, H.W., and Schütt, J. (2022). SOEP-Core v37: Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP) (1984 until 2020). SOEP Survey Papers No. 1106.
- Steinhauer, H.W., Siegers, R., Siebert, M., Jaccobsen, J., and Zinn, S. (2022). Sampling, Nonresponse, and Weighting of the 2020 Refreshment Sample (M6) of the IAB-BAMF-SOEP Refugee Panel. SOEP Survey Papers No. 1104.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* 58: 267–288.
- Wei, J., Matthews, P.C., Stoesser, N., Maddox, T., Lorenzi, L., Studley, R., Bell, J.J., Newton, J.N., Farar, J., Diamond, I., et al. (2021). Anti-spike antibody response to natural SARS-CoV-2 infection in the general population. *Nat. Commun.* 12: 6250.
- Zheng, Z., Rebholz, C.M., Matsushita, K., Hoffman-Bolton, J., Blaha, M.J., Selvin, E., Wruck, L., Sharrett, A.R., and Coresh, J. (2020). Survival advantage of cohort participation attenuates over time: results from three long-standing community-based studies. *Ann. Epidemiol.* 45: 40–46.e4.