1600 2026

SOEP Survey Papers
Series C - Data Documentations (Datendokumentationen)

SOEP-Core – 2024: Sampling, Nonresponse, and Weighting in Sample S

Hans Walter Steinhauer, Rainer Siegers, Felix Süttmann, Reiner Gilberg, Folkert Aust, Martin Kleudgen, and Christian Hunkler

DIW BERLIN

Running since 1984, the German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at **http://www.diw.de/soepsurveypapers**

Please cite this paper as follows:

Hans Walter Steinhauer, Rainer Siegers, Felix Süttmann, Reiner Gilberg, Folkert Aust, Martin Kleudgen, and Christian Hunkler. 2026. SOEP-Core – 2024: Sampling, Nonresponse, and Weighting in Sample S. SOEP Survey Papers 1600. Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Anton-Wilhelm-Amo-Str. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core – 2024: Sampling, Nonresponse, and Weighting in Sample S

Hans Walter Steinhauer[1], Rainer Siegers[1], Felix Süttmann[1], Reiner Gilberg[2], Folkert Aust[2], Martin Kleudgen[2], and Christian Hunkler[1]

[1]German Institute for Economic Research (DIW Berlin)
[2]infas Institute for Applied Social Science

March 17, 2026

Abstract
Sample S is the 2024 refreshment sample of the SOEP-Core. We applied a stratified two-stage cluster sampling design to draw a sample of anchor persons from the German population register. The sample adds 5,314 new households to the panel. Potential selectivity in participation is corrected by using a two-step adjustment for contact and participation. We find attributes of interviewers, the anchor person as well as regional characteristics to have an impact on the success of contacting households. In contrast, participation is mainly impacted by characteristics of the anchor person and the direct neighborhood.

1 Introduction

Panel studies provide valuable insights for social and behavioral research, allowing for the analysis of dynamic patterns and long-term trends for a given population. However, they rely on the participation of individuals who are willing to contribute their time and experiences over an extended period. Maintaining panel participation and ensuring "representativeness" for the general population pose significant challenges. Panel studies need periodic refreshments to provide a sufficiently large sample. Refreshment samples introduce new households into an ongoing panel to compensate for attrition and capture changes in the target population over time. This process plays a vital role in ensuring the continued "representativeness" of the panel and safeguarding the generalizability of the findings.

The Socio-Economic Panel (SOEP), established in 1984, is one of the longest-running panel studies in the field of economic and social sciences. It provides valuable data on various socio-economic aspects of individuals and households over an extended period. We design the panel to include participants from all relevant subgroups of the population living in Germany and draw refreshment samples periodically to account for attrition and changes in the population.

Panel attrition is a common challenge panel studies often face. It can have significant effects on the reliability of study findings. Attrition refers to the loss of participating households or individuals over time in a panel study, resulting from factors such as participant dropout, non-contact, or non-cooperation. Attrition can have several implications. If panel attrition does not occur randomly, certain factors may be associated with participants dropping out of the study. This so-called selectivity can introduce bias if the attrition is related to variables being studied. Moreover, attrition reduces the sample size over time, leading to a loss of statistical power. Smaller sample sizes decrease the precision of estimates and may limit the ability to detect significant effects or associations, especially for (small) subgroups. Addressing attrition is crucial in panel studies to minimize bias and ensure the robustness of findings. However, the loss in sample size cannot always be internally compensated for, and thus refreshment samples are necessary.

We use a stratified two-stage cluster sampling design. At the first stage, we stratify the municipalities responsible for the population register according to federal states. At the second stage, we select individuals from the register born before 2007. Further, contact and participation of households in the panel might be selective. We analyze this possible selectivity using a two-step procedure. First, we analyze contact. Second, given a successful contact, we analyze participation in the panel. Finally, we use raking procedures to make sample distributions conform to those of the population and integrate the refreshment sample into the SOEP.

This paper details the approach applied to refresh the SOEP-Core covering the general population. In Section 2, we provide information on the target population and the sampling frame and the design for the refreshment sample. The fieldwork process and its results are described in Section 3. Section 4 details the different steps of weighting. Characteristics of the final weights are displayed in Section 5 and Section 6 summarizes.

2 Population and Sampling Design

The target population for the refreshment sample S covers the general population in Germany aged 18 years and older living in private households. We sample from this population using the population register as a sampling frame. The German population register ("Einwohnermelderegister" or "Melderegister") is the administrative database that records demographic information of residents living in Germany. It serves as the primary source of official population data and plays a crucial role in selecting samples from the general population in Germany. The population register is maintained by local registration offices ("Einwohnermeldeämter") in each municipality. It is a decentralized system, with each registration office responsible for keeping accurate and up-to-date records of individuals residing within the respective municipality. Residents are required by law to report any changes, such as change of address or significant life events like births, deaths, marriages, and divorces, to their local registration office. The information contained in the population register typically includes the following details:

- personal information including full name, date of birth, gender, and nationality;
- the current residential address;
- information on marital status, such as whether a person is single, married, divorced, or widowed; and
- details about family relationships, including spouses, children, and other dependents residing at the same address.

Unfortunately, the population register does not contain any information on household compositions. Moreover, registration offices do not necessarily have to provide all the information listed above. Especially information on family relationships that would be useful in constructing households is not provided (see §46 BMG). Thus, sampling households directly from this register cannot be done without further steps. We account for this in the weighting procedure described in Section 4.

When drawing samples from the population register, the information listed above can be used to stratify the sample or focus on specific parts of the population. Further, when using the German population register as a sampling frame, a two-stage cluster sampling design is appropriate. Here, we select the municipalities hosting the population register as primary sampling units (PSU) at the first stage and at the second stage sample individuals listed in the registers within each municipality as secondary sampling units (SSU). Sampling individuals provides us with an address necessary to contact the household via an anchor person ("Ankerperson").

The sampling approach, applied by the infas Institute for Applied Social Science (hereafter "infas"), is a stratified two-stage cluster sampling design. The municipalities selected at the first stage are stratified by federal states and municipality size. Municipalities are selected with a probability proportional to size (PPS), where the measure of size is defined as the number of individuals living in the municipality. At the second stage, within the selected municipalities, a sample of individuals born before 2007 is randomly selected from the registers. These individuals serve as anchor persons to obtain access to the households. Table 1 displays the distribution of the sample by federal states.

Table 1: Number of target persons in the population by federal state.

	Population		Sample	
Federal state	Number	Proportion	Number	Proportion
Schleswig-Holstein	2,542,271	0.037	1616	0.038
Hamburg	1,482,992	0.021	828	0.019
Lower Saxony	6,779,390	0.098	4184	0.098
Bremen	423,712	0.006	239	0.006
North Rhine-Westphalia	14,829,917	0.213	8962	0.211
Hesse	5,508,255	0.079	3393	0.080
Rhineland-Palatinate	3,813,407	0.055	2284	0.054
Baden-Württemberg	9,533,518	0.137	5922	0.139
Bavaria	10,380,942	0.149	6562	0.154
Saarland	847,424	0.012	544	0.013
Berlin	3,177,839	0.046	1775	0.042
Brandenburg	1,906,704	0.027	1096	0.026
Mecklenburg-Western Pomerania	1,271,136	0.018	799	0.019
Saxony	3,177,839	0.046	1995	0.047
Saxony-Anhalt	2,118,560	0.030	1275	0.030
Thuringia	1,694,848	0.024	1035	0.024
Total	69,488,754	[a)]0.998	42,509	1.000

Note: [a)]Proportions do not sum up to one due to rounding.

3 Fieldwork Results and Response Rates

The 42,509 address information provided by the registration offices was validated by infas. Prior to field work, sampled anchor persons were sent an invitation letter by mail outlining the data protection regulations and explicitly emphasizing that participation was voluntary. Before the start of the main field phase, the anchor persons were invited to a short pre-recruitment survey. In this survey we asked for their consent to participate or to refuse any further contact. Interviewers contacted the anchor persons between March 2024 and December 2024 for the main interview. Table 2 details the results of the fieldwork on the household-level. In total, 5,314 households completed the interview in full or partial, resulting in a response rate on the household-level, calculated according to American Association for Public Opinion Research (2023), of $RR2 = \frac{5314}{39{,}613} = 0.134$. The refusal rate is $REF1 = \frac{8992}{39{,}613} = 0.227$. Both rates are similar to previous refreshment samples in the SOEP. Some addresses were not deployed in the field because either interviewers did not contact them before the end of the field period or the number of desired interviews was already reached; see AAPOR code 3.11 in table 2. Some addresses were categorized as non-eligible because the household has moved abroad, was out of sample or screened out; see AAPOR codes 4.1 to 4.5 in Table 2.

Table 2: Fieldwork results on the household-level according to American Association for Public Opinion Research (2023).

Final Disposition Code	Number	Proportion
1. Interview		
(1.10) Complete	3021	0.071
(1.20) Partial	2293	0.054
2. Eligible, Non-Interview		
(2.10) Refusal	8430	0.198
(2.12) Break-off	562	0.013
(2.20) Non-contact	4735	0.111
(2.31) Dead	286	0.007
(2.32) Physically / mentally unable	1378	0.032
(2.33) Language problems	545	0.013
(2.36) Miscellaneous	14,183	0.334
3. Unknown eligibility, non-interview		
(3.11) Not attempted or worked	4180	0.098
4. Not Eligible		
(4.10) Out of sample / screened out	18	0.000
(4.20) Moved abroad	130	0.003
(4.40) Untraceable	2191	0.052
(4.50) Non-residential building	557	0.013
Total	42,509	1.000

Note: Proportions might not sum up to one because of errors due to rounding.

4 Cross-Sectional Weighting

The derivation of weights typically involves three main steps (Brick & Kalton, 1996). First, design weights are computed as the inverse of the inclusion probability (see Section 2). In the second step they are adjusted to account for unit nonresponse, a process referred to as sample weighting adjustment by Kalton and Kasprzyk (1986). In the final step, weights are calibrated to ensure that the estimates align with known population parameters, such as totals, ratios, or specific distributions. This step is referred to as population weighting adjustment. Kroh, Siegers, and Kühne (2015) comprehensively outline the general weighting strategy employed by the SOEP and the incorporation of new samples.

Please note that the population register lists individuals without providing information about their household context. However, the SOEP is a household panel survey in which all adults are interviewed. Therefore, a household with two or more individuals, for example, has a higher probability of selection compared to a single-person household. To determine a household's sampling probability, we assign sampling probabilities to all members of the existing households, even though these individuals were not initially sampled as anchor persons. This requires accounting for various characteristics used in stratifying the sample. Once the sampling probabilities for each household member are identified, we can calculate the household sampling probabilities within each household. The household weights are then derived as the inverse of these household sampling probabilities.

Kroh, Kühne, Jacobsen, Siegert, and Siegers (2017) provide a formal documentation of this procedure.

To address potential selectivity resulting from the field work and nonresponse, we employ models that first estimate the success of contacting a household. Given a successful contact, we model the decision-making process of each household and estimate the probability to participate. The models are based on:

1. contact: all households that have been deployed to the field and have been approached,
2. participation: all households that have been successfully contacted.

Given the limited availability of data on households sampled, we use area-level information regarding the residential environment provided by infas360 (see https://datenkatalog.infas360.de/). In addition, we use design information such as stratification variables and data on processes of the field work including interviewer information and attributes of the first contact attempt enter the equation. Finally, the information on the anchor persons provided by the population register, such as for example sex or date of birth were used, too.

4.1 Sample Weighting Adjustments

In the second step of adjusting the design weights, it is crucial to identify strong predictors of nonresponse. To achieve this, we conduct an iterative process, examining all variables included in the previously described data. We select variables that significantly influence the participation decision using bivariate regression analyses. Next, we remove variables from the set of significant predictors if their absolute correlation with one another is 0.95 or higher. This step prevents the inclusion of highly correlated variables in the analysis. The remaining variables form the basis of a preliminary nonresponse model. To obtain a final model, we apply a variable selection procedure in both forward and backward directions, using the Bayesian Information Criterion (BIC) as selection criterion. This approach ensures a more parsimonious model by retaining only the most relevant variables. Table 3 presents the resulting model estimates of the probability to be successfully contacted as well as the response propensities used for deriving weighting adjustments.

Looking at characteristics predicting a successful contact, we find that households who participated in the pre-recruitment survey have a higher probability to participate. In terms of timing a first contact attempt, those contacted at weekend days as well as those contacted in the morning and evening turned out to be more successful. Looking at interviewer attributes, we find interviewers born between 1957 and 1990 to be more successful at contacting their assigned households, compared to their younger or older colleagues. Moreover, interviewers having another occupation than working or pensioners (compared to student, unoccupied, houseman or -wife), are more likely to successfully contact a household. Interviewers with basic (CASMIN 1a, 1b, 1c) and intermediate (CASMIN 2a, 2b, 2c_gen, 2c_voc) education and training degree more frequently contact households successfully. Regarding language, we find native-German speaking interviewers and interviewers speaking English as foreign language to be less successful at contacting households (compared to interviewers speaking Polish, Russian, Turkish, or other languages). When it comes to the attributes of the anchor person, we see that successful contact is more

likely with households drawn via an older anchor person (66-75 and 76 years and older), than via a younger anchor person (18-25 and 26-35 years of age).

Table 3: Models estimating the probability to be successfully contacted as well as the response propensities used for deriving weighting adjustments.

Variable Value	Contact Estimate (Std. Error)	Participation
(Intercept)	0.583***	−2.155***
	(0.028)	(0.045)
1. Attributes of first contact attempt		
Pre-recruitment survey	0.590***	
participated	(0.019)	
Contact day	0.201***	
weekend	(0.018)	
Contact time	0.363***	
evening	(0.015)	
Contact time	0.219***	
morning	(0.024)	
2. Attributes of the interviewer		
CASMIN	0.367***	
1a, 1b, 1c	(0.036)	
CASMIN	0.332***	
2a, 2b, 2c_gen, 2c_voc	(0.016)	
Occupation	0.449***	0.283***
other[1]	(0.017)	(0.030)
Year of birth	0.394***	
1957-1990	(0.016)	
Foreign language	−0.219***	
English	(0.043)	
Native Language	−0.490***	0.114***
German	(0.018)	(0.033)
3. Attributes of the anchor person		
Age (categorized)	−0.083***	−0.287***
18-25	(0.025)	(0.055)
Age (categorized)	−0.136***	
26-35	(0.021)	
Age (categorized)	0.179***	0.129***
66-75	(0.021)	(0.038)
Age (categorized)	0.243***	−0.313***
≥76	(0.022)	(0.048)
Nationality		−0.445***
other than German		(0.050)
Sex		0.183***

Table 3 continued.

Variable Value	Contact Estimate (Std. Error)	Participation Estimate (Std. Error)
male		(0.028)
4. Attributes of the residential area (block of buildings)		
Density of constructions	−0.109***	
high	(0.017)	
Density of constructions	−0.103***	
very high	(0.023)	
Population density	−0.080***	0.218***
high	(0.019)	(0.034)
Population density	−0.099***	0.207***
middle	(0.021)	(0.038)
Type of settlement block	−0.201***	
commercial area	(0.060)	
Share of owners	−0.140***	
lowest quartile	(0.021)	
Purchase index (€/m^2)		−0.133***
lowest quartile		(0.036)
Rent index (€/m^2)	−0.188***	0.219***
highest quartile	(0.020)	(0.035)
Share of inhabitants aged 18-30		0.177***
highest quartile		(0.033)
Share of inhabitants aged 75 and older		0.134***
highest quartile		(0.032)
Share of graduates	0.100***	
more than 2.5%	(0.024)	
Share of Catholics	−0.218***	
high	(0.023)	
Share of Catholics	−0.125***	
middle	(0.021)	
Share of Catholics	−0.101***	
low	(0.021)	
Share of Protestants	−0.130***	
high	(0.020)	
Share of Protestants	−0.070***	
middle	(0.019)	
5. Attributes of the residential area (neighborhood)		
Unemployment rate		0.164***
lowest quartile		(0.033)
Index on availability of public transport	−0.081***	
high	(0.020)	
Index on availability of public transport	−0.104**	
low	(0.032)	

Table 3 continued.

Variable Value	Contact Estimate (Std. Error)	Participation Estimate (Std. Error)
Number of school graduates: secondary school highest quartile	0.152*** (0.032)	
Number of school graduates: Fachhochschulreife highest quartile	0.110*** (0.025)	
Number of school graduates: Abitur highest quartile	−0.099*** (0.029)	
Number of persons aged 15+ with lower secondary school qualification highest quartile	−0.128*** (0.025)	
Number of persons aged 15+ without school-leaving qualification lowest quartile	0.237*** (0.031)	
Number of persons aged 15+ without school-leaving qualification highest quartile	−0.178*** (0.027)	
Number of apprentices highest quartile		0.214*** (0.038)
Number of persons aged 15+ with doctorate lowest quartile	−0.210*** (0.031)	
Number of persons aged 15+ with doctorate highest quartile	0.123*** (0.027)	
Area of district (in m^2) highest quartile	−0.096*** (0.018)	
Share of persons without a migration background lowest quartile	−0.115*** (0.025)	
Share of persons without a migration background highest quartile	0.140*** (0.026)	
Share of persons with an Italian migration background highest quartile	−0.196*** (0.019)	
Share of persons with a Yugoslavian migration background highest quartile	0.099*** (0.021)	
Share of persons with a Polish migration background lowest quartile	−0.114*** (0.020)	
Share of persons with a Polish migration background highest quartile	−0.098*** (0.018)	
Share of persons with another migration background lowest quartile	−0.075*** (0.021)	
Share of persons with a Soviet migration background highest quartile	−0.098*** (0.019)	
Share of persons with a Turkish migration background lowest quartile	0.077*** (0.023)	
Index of Socio-economic Deprivation low (2^{nd} decile)	0.138*** (0.029)	

Table 3 continued.

Variable Value	Contact Estimate (Std. Error)	Participation Estimate (Std. Error)
Index of Socio-economic Deprivation	-0.093^{***}	0.220^{***}
middle (6^{th} decile)	(0.025)	(0.046)
N	39327	29984

Notes: Dependent variable: Success in contacting the household (1 = yes, 0 = no), participation of the household (1 = yes, 0 = no). [1] other than working or pensioner (e.g. unoccupied, student, or housewife). Significance indicated by $*** \equiv p < 0.001$, $** \equiv p < 0.01$, and $* \equiv p < 0.05$. The model is estimated using the function `glm()` with a cloglog linkfunction in R (R Core Team, 2023).

Looking at attributes of the residential area, only households located in areas where 2.5 or more percent of the people graduated are more likely to be contacted successfully. All other attributes have a negative impact. This holds for the high and very high density of constructions (compared to middle, low, very low) as well as a middle and high population density (compared to very low, low, very high). Interviewers are less succesful in contacting households located in settlement blocks in commercial areas. The same is true for households in areas with a low share of owners and a high rent index. For the share of Catholics (low, middle, high, compared to very low, or very high) and Protestants (middle, high) we find a negative impact on the likelihood to be contacted. Moving on to the attributes of the neighborhood, we find a low as well as a high index of the availability of public transport to have a negative effect on successful contacts with a household. Households in an area with a higher number of school graduates from secondary schools as well as secondary school graduates with an advanced technical college entrance qualification ("Fachhochschulreife") are more likely to be contacted successfully. In contrast to this, a high number of students graduating with the highest secondary-school diploma ("Abitur"), lowers the probability for successful contact. In neighborhoods with a high number of persons aged 15 years and older, having lower secondary school qualification, households are more likely to be contacted successfully. The same holds for households in neighborhoods with a high number of persons aged 15 years and older having no school-leaving qualification, whereas in areas with a low number the likelihood for a successful contact is higher. In areas with a low number of persons aged 15 years and older the likelihood for successful contact is low and in areas with a higher number it is higher. Households residing in large districts (in terms of m^2) have a lower probability for successful contact. For households in neighborhoods wit a low (high) share of persons without a migration background, the likelihood for successful contact is lower (higher), indicating a linear relationship. In neighborhoods with a high share of Italian (Yugoslavian) migrant households have a lower (higher) rate of successful contacts. Households located in a neighborhood with a low or high share of Polish migrants have a lower success rate when it comes to contact. A lower likelihood for successful contacts is also found for households in neighborhoods with a low share of other migrants as well as a high share of Soviet migrants. A higher likelihood for successful contacts is found for households in neighborhoods with a low share of Turkish migrants. Households in a neighborhood with a value in the 2^{nd} decile of the Index of socio-economic deprivation (GISD) have a higher probability for successful contact, whereas households in neighborhoods with a value in the 6^{th} decile

have a lower one.

When it comes to participation, we find fewer characteristics that significantly influenced the household's decision. Interviewers with an intermediate level of education, indicated by CASMIN (2a, 2b, 2c_gen, 2c_voc) were more likely to convince a household to particpate in the survey. The same holds for interviewers whose native language is German. Looking at the anchor persons' characteristics, we see that households sampled via an anchor person aged 18 to 25 years of age or 76 years and older were less likely to participate in the survey. In contrast, households sampled via an anchor person aged 66 to 75 are more likely to take part. An anchor person with a nationality other than German lowers the propensity to participate in the panel. Surprisingly, households sampled via a male anchor person have a higher probability to participate. Looking at attributes of the residential area we see that households in highly (high and very high) populated areas are more likely to take part. Households in areas with a low purchase index are less likely to participate in the survey, whereas households in areas with a high rent index are more likely. Both, households in areas with a high share of younger (18 to 30 years) and a high share of older (75 years and older) persons have a higher propensity to participate in the panel. Coming to attributes of the neighborhood, we see that a low unemployment rate has a positive effect on the decision to participate in the panel. The same holds for households in neighborhoods with a high number of apprentices. Finally, households in neighborhoods with a value of the index of socio-economic deprivation in the 6^{th} decile impacts the decision to participate positively.

4.2 Population Weighting Adjustments

In the final step of the weighting process, we use post-stratification and raking techniques to adjust the weights obtained previously. This adjustment aligns the weights with known population totals, as well as joint and marginal distributions. The specific method used depends on the available population data, with a comprehensive overview provided by Kalton and Flores-Cervantes (2003). The resulting weights from this step form the basis for deriving both cross-sectional and longitudinal weights for subsequent survey waves, beginning with wave 2. The population parameters and distributions used in these adjustments were provided by the Federal Statistical Office, based on data from the German Microcensus. Margins used in the post-stratification process are:

number of households with at least one person of the population by

- household typology,
- household size,
- size of municipality,
- federal states,
- rural-urban-classification,
- house owner, as well as
- migration background, year of immigration, and nationality; and

number of persons of the population by

- sex, nationality, and age groups,

- migration background, as well as
- year of immigration.

For information on the categories of the different variables, see Siegers, Steinhauer, and Schütt (2022).

5 Characteristics of Weights

Due to stratification and disproportional allocation of households, there is some variance in the design weights. Adjusting the initial design weights for unit nonresponse by multiplying them with the inverse of estimated participation probabilities increases variation in the second weighting step resulting in the adjusted design weights. The population weighting adjustments to derive calibrated weights further add to the variation and magnitude of weights, see Table 4. The resulting calibrated weights are provided in the variable `hhrf0` included in the data set `hpathl` as well as in the variable `bohhrfs` in the data set `hhrf`.

After the integration step, a further post-stratification step was carried out. Here, weights of all households in the SOEP are adjusted with respect to the standard marginal distributions. Using the resulting standard SOEP weighting factors (`hhrf` included in `hpathl` and `phrf` included in `ppathl`), the sample S cases can then be analyzed jointly and comparatively in combination with all other SOEP observations.

Table 4: Characteristics of weights after the steps of the weighting process (rounded to integer values).

		Quantiles							
Step	Min.	10%	25%	50%	75%	90%	Max.	Mean	SD
DW	169	555	756	885	1208	1803	2695	1052	485
SWA	866	3036	3999	5725	8621	12,942	38,619	6983	4389
PWA	615	2893	3908	5767	9297	14,517	69,653	7537	5635

Abbreviations: SD = standard deviation, DW = design weighting, SWA = sample weighting adjustment, PWA = population weighting adjustment.

6 Summary

The new Sample S is a refresher sample adding additional 5,314 households to the SOEP. Like previous refreshment samples, it was drawn from the population register using a stratified two-stage sampling design. Concerning field work, we find a few potential selectivities. Using the information available from field work, interviewers, anchor persons, and regional attributes of the household, we account for possible selectivities due to contact and participation in our weighting strategy. We find attributes of interviewers, the anchor person as well as regional characteristics to have an impact on the success of contacting households. In contrast, participation is mainly impacted by characteristics of the

anchor person and the direct neighborhood. We adjust design weights by the inverse of the estimated probabilities for contact and participation to correct for unit nonresponse. In the final step we use post-stratification to make the distributions conform to those provided by the microcensus. Weights for this sample are provided on the household as well as on the person-level. Moreover, households have been seamlessly integrated into SOEP-Core.

References

American Association for Public Opinion Research. (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (10th ed.). Retrieved from https://aapor.org/wp-content/uploads/2024/03/Standards-Definitions-10th-edition.pdf

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical methods in medical research*, *5*(3), 215–238. doi: 10.1177/096228029600500302

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. *Journal of official statistics*, *19*(2), 81–97.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey methodology*, *12*(1), 1–16.

Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). *Sampling, nonresponse, and integrated weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4)—revised version* (SOEP Survey Papers No. 477). Berlin: DIW/SOEP. Retrieved from https://www.diw.de/documents/publikationen/73/diw_01.c.572346.de/diw_ssp0477.pdf

Kroh, M., Siegers, R., & Kühne, S. (2015). Gewichtung und integration von auffrischungsstichproben am beispiel des sozio-oekonomischen panels (soep). In J. Schupp & C. Wolf (Eds.), *Nonresponse bias: Qualitätssicherung sozialwissenschaftlicher umfragen* (pp. 409–444). Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from https://doi.org/10.1007/978-3-658-10459-7_13 doi: 10.1007/978-3-658-10459-7_13

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Siegers, R., Steinhauer, H. W., & Schütt, J. (2022). *SOEP-Core v37 Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2018)* (SOEP Survey Papers No. 1106). Berlin: DIW/SOEP.