1602 2026

SOEP Survey Papers
Series C - Data Documentations (Datendokumentationen)

IAB-SOEP-Migration – 2024: Sampling, Nonresponse, and Weighting of the Sample M8d

Hans Walter Steinhauer, Parvati Trübswetter, Tanja Fendel, Felix Süttmann, and Rainer Siegers

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, Parvati Trübswetter, Tanja Fendel, Felix Süttmann, and Rainer Siegers. 2026. IAB-SOEP-Migration – 2024: Sampling, Nonresponse, and Weighting of the Sample M8d. SOEP Survey Papers 1602. 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

IAB-SOEP-Migration – 2024: Sampling, Nonresponse, and Weighting in Sample M8d

Hans Walter Steinhauer[1], Parvati Trübswetter[2], Tanja Fendel[2], Felix Süttmann[1], and Rainer Siegers[1]

[1]German Institute for Economic Research (DIW Berlin)
[2]Institute for Employment Research (IAB)

March 17, 2026

Abstract
This report provides details on the sampling design, the fieldwork, as well as nonresponse and population adjustments for the 2024 sample M8d of the IAB-SOEP-Migration Survey. The survey is conducted in cooperation between the Institute for Employment Research (IAB) and the Socio-Economic Panel (SOEP). It refreshes the IAB-SOEP-Migration samples M8a, M8b and M8c by adding 348 households of foreigners from countries outside the European Union (EU) to allow for evaluating the Skilled Labor Immigration Act (Fachkräfteeinwanderungsgesetz). The act came into effect on March 1, 2020. Its goal is to facilitate the immigration of skilled workers from non-EU countries to Germany.

Acknowledgments
This report is related to and relies on Steinhauer, Trübswetter, and Zinn (2022), Steinhauer et al. (2025a), and Steinhauer et al. (2025b).

1 Introduction

Panel studies are widely recognized as valuable tools in social and behavioral research, offering insights into dynamic patterns and long-term trends within a given population. These studies rely heavily on the participation of individuals who are willing to contribute their time and experiences over an extended period of time. However, maintaining panel participation and ensuring that the sample remains "representative" of its population is challenging. A crucial aspect of panel studies is the refreshment of the sample. Adding new participants into an ongoing panel compensates for attrition while also capturing changes in the population over time. This process is essential for maintaining the "representativeness" of the panel and for preserving the integrity of the findings.

The Socio-Economic Panel (SOEP) is one of the longest-running panel studies in the field of economic and social sciences. The SOEP provides data on various socio-economic aspects of individuals and households over an extended period since 1984. By including participants from different subgroups of the German population, the panel ensures diversity. Specifically, the IAB-SOEP-Migration samples M1, M2, M7, M8a, M8b, and M8c focus on migration to Germany motivated by job opportunities. More details on the IAB-SOEP-Migration samples are provided by Brücker et al. (2014). The SOEP samples M8a and M8b cover a specific subgroup of the population living in Germany, namely skilled labor immigration from non-EU countries to Germany; for details see Steinhauer et al. (2022) and Steinhauer et al. (2025a). The SOEP sample M8c covers skilled labor immigration in general; see Steinhauer et al. (2025b). These samples provide a basis to evaluate the Skilled Labor Immigration Act (Fachkräfteeinwanderungsgesetz), which initially came into effect on March 1, 2020, with the goal of facilitating the immigration of skilled workers from non-EU countries to Germany. To further support this evaluation and compensate for the loss of households in the sample due to panel attrition, we introduce SOEP sample M8d. Moreover, sample M8d augments the previous migration samples M1, M2, M7, M8a, M8b, and M8c.

This paper details the approach applied to refresh the IAB-SOEP-Migration samples covering the population of non-EU nationals migrating to enter the German labor market. In Section 2, we provide information on the target population, the sampling frame, and details on the sampling design. 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, while Section 6 summarizes.

2 Sampling Design

The target population for sample M8d comprises households with persons from non-EU countries who received a work permission from the federal employment agency between May 15, 2023, and September 30, 2023. The Federal Agency for Employment listed 29,797 target persons in the population, see Table 1. From this population we draw a sample of 8,000 individuals with the design also applied in previous samples M8a, M8b, and M8c; see Steinhauer et al. (2022) for details.

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

	Population		Sample	
Federal State	Number	Proportion	Number	Proportion
Baden-Württemberg	4478	0.150	1197	0.150
Bavaria	5895	0.198	1533	0.192
Berlin	1382	0.046	475	0.059
Brandenburg	555	0.019	154	0.019
Bremen	285	0.010	90	0.011
Hamburg	948	0.032	245	0.031
Hesse	1978	0.066	544	0.068
Mecklenburg-Western Pomerania	706	0.024	208	0.026
Lower Saxony	3024	0.101	755	0.094
North Rhine-Westphalia	5541	0.186	1528	0.191
Rhineland-Palatinate	1293	0.043	364	0.046
Saarland	157	0.005	40	0.005
Saxony	1478	0.050	379	0.047
Saxony-Anhalt	618	0.021	143	0.018
Schleswig-Holstein	812	0.027	189	0.024
Thuringia	647	0.022	156	0.020
Total	29,797	1.000	8000	1.000

Note: Proportions might not add up to one because of rounding.

3 Fieldwork results and Response Rates

The addresses sampled from official records at the Federal Agency for Employment were validated by infas (Institute of Applied Social Sciences) and deployed to the field. Some addresses were found not to be valid. That is why the number of households reported here is reduced to 7,430. The interviews for the sample M8d were conducted between April 2024 and December 2024. Well in advance, the selected sample of households received letters via mail emphasizing that participation was voluntary. Table 2 details the results of the fieldwork on the household-level. In total, there were 348 complete or partial interviews, resulting in an overall response rate on the household-level, calculated according to American Association for Public Opinion Research (2023), of $RR2 = \frac{348}{4{,}596} = 0.076$ (with $RR2_{CAPI} = 0.069$ and $RR2_{CAWI} = 0.078$). The number of households explicitly refusing to participate is rather low. The resulting overall refusal rate is $REF1 = \frac{361}{4{,}596} = 0.079$ (with $REF1_{CAPI} = 0.126$ and $REF1_{CAWI} = 0.065$). Both rates are similar to previous samples. Some addresses, although sampled and valid, were not contacted in the field because either interviewers did not reach them by the end of the field period or the numbers of desired interviews was already reached (see AAPOR code 3.11 in Table 2). Other addresses are out of sample, because the household has moved abroad or was 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	CAPI Number	CAPI Proportion	CAWI Number	CAWI Proportion
1. Interview				
(1.10) Complete	57	0.039	163	0.027
(1.20) Partial	14	0.010	114	0.019
2. Eligible, Non-Interview				
(2.10) Refusal	114	0.079	80	0.013
(2.12) Break-off	16	0.011	151	0.025
(2.20) Non-contact	209	0.144	0	0.000
(2.31) Dead	1	0.001	0	0.000
(2.32) Physically / mentally unable	12	0.008	0	0.000
(2.33) Language problems	72	0.050	0	0.000
(2.36) Miscellaneous	480	0.331	25	0.004
3. Unknown eligibility, non-interview				
(3.11) Not attempted or worked	59	0.041	0	0.000
(3.20) Nothing ever returned	0	0.000	3029	0.506
4. Not Eligible				
(4.10) Out of sample / screened out	180	0.124	955	0.160
(4.20) Moved abroad	13	0.009	0	0.000
(4.40) Untraceable	181	0.125	1465	0.245
(4.50) Non-residential building	40	0.028	0	0.000
Total	1448	1.000	5982	1.000

Note: Proportions might not sum up to one because of roundings.

4 Cross-Sectional Weighting

According to Brick and Kalton (1996), the computation of weights typically involves three main steps. First, the design weights are calculated as the inverse of the inclusion probability (see Section 2). These are then adjusted to account for unit nonresponse, a process referred to as sample weighting adjustment by Kalton and Kasprzyk (1986). Finally, in the third step, the weights are calibrated to ensure that the estimates align with known population parameters, such as totals, ratios, or specific distributions. Kalton and Kasprzyk (1986) refer to this step as population weighting adjustment. For a comprehensive understanding of the general weighting strategy employed by the SOEP and the incorporation of new samples, please refer to Kroh, Siegers, and Kühne (2015).

To adjust for unit nonresponse, we make use of the Integrated Employment Biographies (IEB), which is administrative data provided by the IAB. The IEB is spell data based on IAB's employment history (BeH), IAB's benefit recipient history (LeH), the participants-in-measures data (MTG), and job search data originating from the applicants' pool database (BewA). Thus, the IEB include observations of unemployment benefits, job

search, and participation in active labor market programs; see Oberschachtsiek, Scioch, Seysen, and Heining (2009) for details. Beyond that, it includes socio-demographic information on gender, age, and nationality as well as geographic information, including, for example, regional classifications.

Please note that the IEB lists individuals without providing information about their household context. However, the SOEP is a household panel survey in which all adults are interviewed. Consequently, a household with two individuals, for example, has twice the 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 process requires accounting for characteristics used in clustering the sample. Once the sampling probabilities for each household member are identified, we calculate the household sampling probabilities by summing these probabilities within each household. The household weights are then derived as the inverse of these household sampling probabilities. It is important to note that the number of households in the target population is unknown and cannot be determined from the sampling frame, as the register does not include household identifiers.

To address potential selectivity resulting from fieldwork and nonresponse, we employ two models that capture the success of contacting a household as well as the decision-making process of households regarding participation. The models incorporate information on

a) all households that have been deployed to the field and have been approached/ contacted; and

b) both participating and nonparticipating households.

The first model estimates the probability of successfully contacting a household. Given successful contact, the second model estimates the probability of the household participating in the panel. Given the limited availability of data on households in the initial sample, we use area-level information regarding the residential environment provided by infas360 (see https://datenkatalog.infas360.de/). Additionally, we use design information and data on fieldwork processes. The latter include information on the interviewer alongside attributes of the first contact attempt. This information only applies to the CAPI field, because the CAWI does not require interviewers.

4.1 Sample Weighting Adjustments

In the second step of correcting the design weights, it is essential to identify strong predictors of nonresponse. To accomplish this, we conduct an iterative process that involves examining information listed above. We select variables that demonstrate significant influence on the participation decision through bivariate regression analysis. Subsequently, we remove variables from the set of significant variables if their absolute correlation value with each other is greater than or equal to 0.95. This step ensures that highly correlated variables are not duplicated in the analysis. The remaining variables then form the basis for a preparatory nonresponse model. To obtain the final model, we employ a variable selection procedure in both forward and backward directions, using the Bayesian Information Criterion (BIC) as the selection criterion. This approach allows us to arrive at a more parsimonious model, retaining only the most relevant variables. The resulting

models estimating the probability to be successfully contacted as well as the response propensities used for deriving weighting adjustments, are presented in Table 3.

Various factors influence the success of contacting households for interviews. These factors include the timing of the first contact, the interviewer characteristics, specific attributes of the residence and neighborhood, as well as socio-economic characteristics and attributes of the anchor person from the IEB. Besides that, we separate the two main survey modes: CAWI and CAPI for modelling contact and participation. Naturally, interviewer characteristics can only be considered in CAPI mode.

4.1.1 Contact

Looking at the households that have been approached in CAPI mode we find an in person first contact attempt to increase the likelihood of successful contact. A first contact attempt in morning time is less likely to result in a successful contact. When it comes to attributes of the neighborhood a high number of persons aged 15 years and older with a secondary school certificate leads to higher rates of successful contacts. In neighborhoods with a high purchase index of property the success rates for contact are lower.

In CAWI mode it is mostly attributes of the anchor person that predict a successful contact attempt. Having a higher level of school education (*Highest school education*)as well as having a higher educational degree (*Highest education*) positively influences the chances of a successful contact. Given the population of the sample being skilled labor immigrants, this finding is little surprising. Having a first entry in the IEB data before 2021 as well as being present in the IEB data for less than one year has a negative impact on successfully contacting the household. Finally, residing in a residential area where the dominant type of building blocks is another than residential or mixed also lowers the probability for a successful contact.

4.1.2 Participation

Households administered by an interviewer are more likely to participate in the panel, if the interviewer has an educational degree from secondary education, i.e. has a CASMIN in the ranges of 2a, 2b, 2c_gen, and 2c_voc. Also, interviewers having another occupation than working full time or being a pensioner are more likely to achieve complete interviews and panel consent. For CAPI mode households in neighborhoods with a high number of persons aged 15 years and older without professional qualification have a higher propensity to participate in the panel.

In CAWI mode households with an anchor person having a master or diploma as highest educational degree were more likely to participate in the panel. When the anchor person's first entry in the IEB data was from before 2021 the household has a lower participation propensity. Finally, households residing in a neighborhood with a low number of persons with a Turkish migration background have a higher propensity to participate in the panel.

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

	Contact		Participation	
	CAPI	CAWI	CAPI	CAWI
Variable Value	Estimate (Std. Error)			
Intercept	0.217**	−1.949***	−3.240***	−0.382***
	(0.075)	(0.088)	(0.249)	(0.083)
1. Attributes of first contact attempt				
Contact time	−0.540**			
morning	(0.186)			
Contact	0.283***			
in person	(0.081)			
2. Attributes of the anchor person				
Highest education		0.464***		
Vocational training		(0.109)		
Highest education		0.563***		
Bachelor		(0.147)		
Highest education				0.749***
Master / Diploma				(0.199)
Highest school education		0.627***		
Advanced technical college entrance qualification		(0.119)		
Highest school education		0.498***		
Abitur (A level)		(0.127)		
First entry into IEB		−1.211***		−0.594**
before 2021		(0.129)		(0.197)
Number of years in IEB		−0.423***		
less than 1		(0.107)		
3. Attributes of the interviewer (CAPI only)				
CASMIN			0.790**	
2a, 2b, 2c_gen, 2c_voc			(0.245)	
Occupation			0.697**	
other[1]			(0.238)	
4. Attributes of the residential area (block of buildings)				
Type of settlement block		−0.984***		
other[2]		(0.255)		
5. Attributes of the residential area (neighborhood)				
Number of persons aged 15+ without professional qualification			0.772**	

Table 3 continued.

	Contact		Participation	
	CAPI	CAWI	CAPI	CAWI
Variable	Estimate			
Value	(Std. Error)			
highest quartile			(0.244)	
Number of persons aged 15+ with secondary school certificate	0.355^{***}			
highest quartile	(0.085)			
Purchase index (€/m^2)	-0.342^{***}			
highest quartile	(0.083)			
Number of persons with a Turkish migration background				0.561^{***}
lowest quartile				(0.135)

Notes: Dependent variable: Success in contacting the household (Contact: 1 = yes, 0 = no), participation of the household (Participation: 1 = yes, 0 = no). Abbreviations are: CAPI = Computer Assisted Personal Interview, CAWI = Computer Assisted Web Interview, IEB = Integrated Employment Biographies. [1] other than working or pensioner. [2] other than residential or mixed. 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 link function in R (R Core Team, 2025).

4.2 Population Weighting Adjustments

In the final step of the weighting process, we employ post-stratification and raking techniques to adjust the weights obtained in the previous step. This adjustment is necessary to align the weights with known population totals, as well as joint and marginal distributions. The specific method chosen for this adjustment depends on the available data for the population. A comprehensive overview of these methods is provided by Kalton and Flores-Cervantes (2003). The resulting weights from this step serve as the foundation for deriving cross-sectional and longitudinal weights for subsequent waves, starting from Wave 2 onwards.

The population parameters and distributions utilized in the population weighting adjustments are provided by the Federal Statistical Office, drawing upon data from the German Microcensus 2024. Margins used in the post-stratification process are:

Number of households with at least one person of the M8d-population who immigrated to Germany in 2023 by

- household typology (single vs. other);
- municipality size; and
- Regions (federal states categorized by north, east, south, and west)

Number of persons of the M8d-population who immigrated to Germany in 2023 by

- sex,
- nationality, and
- age groups

5 Characteristics of Weights

Due to the sampling design, there is some variance in the design weights. Multiplying design weights with the inverse of estimated participation probabilities increases variation in the second weighting step; compare Table 4. The population weighting adjustments then again add to the variation and magnitude of weights. Resulting weights are provided in the variable `hhrf0` included in the data set `hpathl` as well as in the variable `bohhrfm8d` in the data set `hhrf`.

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	2	2	3	4	7	7	7	5	2
SWA	6	16	23	40	61	111	541	56	58
PWA	16	57	88	134	246	482	1485	205	203

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

After the integration step, a further post-stratification step was carried out in which the weights (previously nonresponse-adjusted, if necessary post-stratified and integrated) of all SOEP samples were adjusted with respect to the standard marginal distributions used by SOEP, as taken from the Microcensus 2024. Using the resulting standard SOEP weighting factors (`hhrf` included in `hpathl` and `phrf` included in `ppathl`), the sample M8d cases can then be analyzed jointly and comparatively in combination with all other SOEP cases.

6 Summary

The new sample M8d is a refresher sample adding 348 households to the IAB-SOEP-Migration survey. Like to previous migration samples in this survey it is drawn from the IEBs. Concerning field work, we find a few possible selectivities in the CAPI as well as in the CAWI field. However, they differ in terms of their characteristics but are similar to previous migration samples. In the CAPI field, we mainly find that attributes of the residential area affect contact. In contrast, in the CAWI field, we find that successful contacts are primarily driven by the attributes of the anchor person. We find similar patterns for participation. Using the information available from regional attributes of the household as well as attributes of the anchor person from the IEB, we account for possible selectivities due to contact and participation in our weighting strategy.

Sample M8d augments the samples of the IAB-SOEP-Migration survey by non-EU nationals joining the German labor force. The sample secures and expands the previous analysis potential of samples M1, M2, M7, M8a, M8b, and M8c. Thereby, it can be used to evaluate the Skilled Labor Immigration Act (Fachkräfteeinwanderungsgesetz) together with

samples M8a, M8b, and M8c. The households of this new sample have been seamlessly integrated.

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

Brücker, H., Kroh, M., Bartsch, S., Goebel, J., Kühne, S., Liebau, E., … Schupp, J. (2014). *The new IAB-SOEP migration sample. An introduction into the methodology and the contents.* (SOEP Survey Papers No. 216). Berlin: DIW-Berlin.

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., Siegers, R., & Kühne, S. (2015). Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP). In *Nonresponse bias* (pp. 409–444). Springer.

Oberschachtsiek, D., Scioch, P., Seysen, C., & Heining, J. (2009). *Stichprobe der Integrierten Erwerbsbiografien IEBS* (FDZ-Datenreport No. 03/2009). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung. Retrieved from http://doku.iab.de/fdz/reporte/2009/DR_03-09.pdf

R Core Team. (2025). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Steinhauer, H. W., Trübswetter, P., Fendel, T., Ivanov, B., Cardozo Silva, A., Süttmann, F., … Zinn, S. (2025a). *IAB-SOEP-Migration – 2022: Sampling, Nonresponse, and Weighting in Sample M8b* (SOEP Survey Papers No. 1497). Berlin: DIW-Berlin.

Steinhauer, H. W., Trübswetter, P., Fendel, T., Ivanov, B., Cardozo Silva, A., Süttmann, F., … Zinn, S. (2025b). *IAB-SOEP-Migration – 2023: Sampling, Nonresponse, and Weighting in Sample M8c* (SOEP Survey Papers No. 1536). Berlin: DIW-Berlin.

Steinhauer, H. W., Trübswetter, P., & Zinn, S. (2022). *SOEP-Core - 2020: Sampling, Nonresponse, and Weighting in the IAB-SOEP Migration Studies M7 and M8* (SOEP Survey Papers No. 1105). Berlin: DIW-Berlin.