

SOECBIAS-COVREF Data Set: Income, (Mis-)Perceptions, Redistributive Preferences and Economic Affectedness in times of the COVID-19 Pandemic

Method Report

Authors:

Henning Lohmann

Hequn Wang

Nico Eggers

This report is an extended and updated version of:

Beblo, Miriam/ Bublitz, Elisabeth/ Jäger, Julian/ Lohmann, Henning/ Wang, Hequn (2023): SOECBIAS Data Set Socioeconomic Data on Income (Mis-)Perceptions and Redistributive Preferences in Four EU Member States, Method Report, Hamburg.

Correspondence:

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

Fachbereich Sozialökonomie

Welckerstr. 8

D-20354 Hamburg

henning.lohmann@uni-hamburg.de

Table of Contents

1	Introduction.....	1
2	Survey Structure and Design of Experiment	1
2.1	Wave 1 (SOECBIAS Main Survey).....	2
2.1.1	Questions before the Treatments.....	2
2.1.2	Questions after the Treatments	3
2.2	Wave 2 (SOECBIAS follow-up)	4
2.3	Wave 3 (COVREF)	5
2.4	Wave 4 (COVREF)	5
2.5	Regional-level Identifiers.....	6
2.6	Data on Actual Income Positions.....	6
3	Sampling and Data Collection	6
3.1	Sample	6
3.2	Pre-Test	7
3.3	SOECBIAS Wave 1 and 2	8
3.4	COVREF Wave 3 and 4	8
4	Data Cleaning Process and Data Quality	8
5	Analysis Potential	9
	Acknowledgments	12

1 Introduction

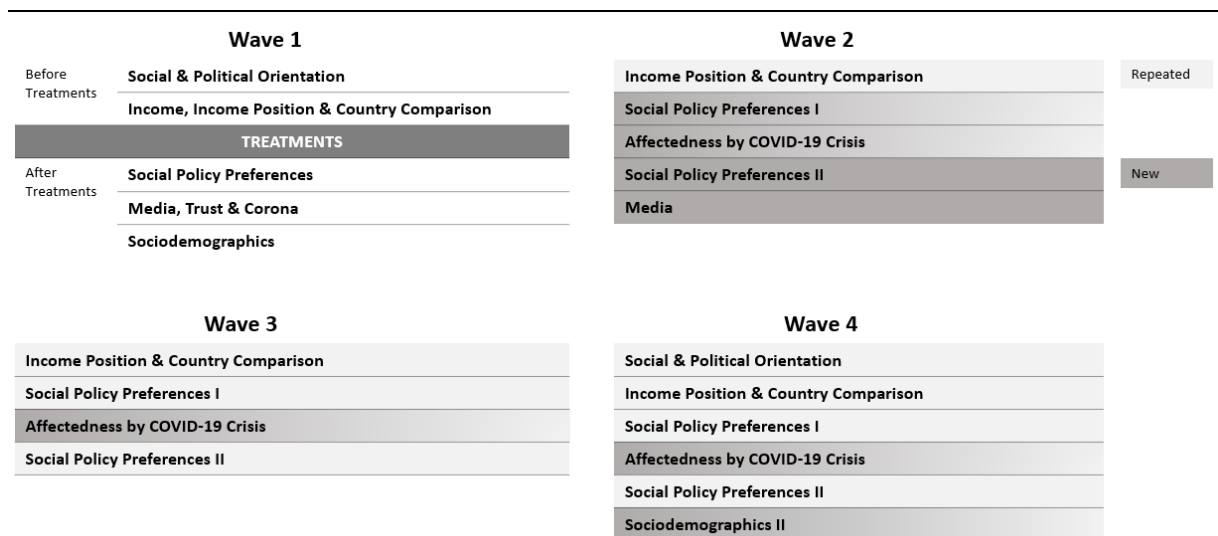
The SOECBIAS-COVREF data set is an output of two research projects: The interdisciplinary research project SOECBIAS, funded by the Federal Ministry of Labour and Social Affairs in Germany and the research project COVREF, funded by the German Research Foundation, both conducted at the Universität Hamburg. SOECBIAS started in May 2019 and ended in December 2021 and studied income perceptions and redistributive preferences combining inequality research with social policy and welfare state research in economics and sociology. SOECBIAS addressed three main questions: How do Europeans perceive national and European social policy? What explains the perception of one's own income position within the EU income distribution? What are the consequences of these perceptions for the assessment of redistribution measures?

Since the SOECBIAS survey period overlapped with the onset of the COVID-19 pandemic it became an additional focus of the project and evolved into a further research project (COVREF) which started in June 2022 and ended in June 2024.¹ This research project focused on the individual affectedness by the COVID-19 pandemic from an objective and subjective point of view. The two leading questions for COVREF were: What influence do objectively measurable factors like illness, job loss or loss of income during the COVID-19 pandemic have on the subjective assessment of economic affectedness and the perception of one's own income position? What impact do comparisons with reference groups or reference points have on the perception and assessment of economic affectedness?

2 Survey Structure and Design of Experiment

The SOECBIAS-COVREF data set comprises the original SOECBIAS main survey (wave 1) with an experiment and a follow-up survey (wave 2). Both were collected in 2020 (March and June/July). Wave 3 was collected in April 2021 covering specifically the outcomes of the pandemic. Wave 4 on longer term consequences of the pandemic was collected in the framework of the COVREF project in summer 2022. All waves were implemented as standardized online surveys. **Figure 1** depicts the design and the contents of the four waves. In the following sections, we provide an overview on each of the topics and the components of the respective waves.

¹ For more information on the SOECBIAS and COVREF projects see: <https://www.wiso.uni-hamburg.de/soecbias> & <https://www.wiso.uni-hamburg.de/fachbereich-sozoek/professuren/lohmann/forschung/01-forschungsprojekte/covref.html>



Notes: Modules on the right in light grey include repeated questions from the previous wave. Modules in dark grey include new questions. The Modules “Social Policy Preferences I” and “Affectedness by COVID-19 Crisis” for example include both new questions and repeated questions from the first wave.

Figure 1: Design and contents of the Four Waves

2.1 Wave 1 (SOECBIAS Main Survey)

In the main survey we collected data in four countries: Germany, Italy, Poland and Sweden. The core of the main survey was a survey experiment in which respondents were informed about their national and/or EU income position. Consequently, questions that were supposed to remain unaffected by the intervention were asked before the treatments. These include, among others, questions on the perceived own income positions. Questions following the treatments constitute either possible outcome variables (mainly preferences for social policy measures) or characteristics and traits that the treatment should have no effect on, such as socio-demographics and political trust. The structure of the main survey will be discussed in the following sub-sections in more detail. For more detailed information on the theoretical background and the experiment see the documentation of the two-wave SOECBIAS dataset.

2.1.1 Questions before the Treatments

Social & Political Orientation: Respondents were asked whether and to what degree they identified as an EU citizen, about their political orientation, their welfare attitudes and their own social status as well as the social status of their friends and neighbours.

Income, Income Position & Country Comparison: First, respondents were asked to name all sources of income for each member within their household during the past year. This was supposed to help respondents recall all their income from last year. It additionally ensured that respondents would understand the subsequent questions that referred to the full income of their household and not simply earnings. After that, respondents were asked to indicate their last year's net household income. Based on their stated income, respondents were then asked to guess which percentage of households they thought were poorer than their own household, first within their country and second within the EU. The module concluded with a question about where respondents thought their country stood compared to the EU average.

2.1.2 Questions after the Treatments

Social Policy Preferences: The core of this module consists of questions how respondents assess their own country and social policy related issues. A common question from cross-national surveys, such as the *International Social Survey Programme (ISSP)*, was implemented on whether respondents evaluate the income distribution in their country as just or unjust. As an extension of the question, respondents were also asked to evaluate the income distribution within the EU. The module also includes a question from the *OECD's Risks that Matter Survey* (<https://www.oecd.org/social/risks-that-matter.htm>) on whether respondents are willing to pay an additional 2% in taxes or social security contributions of their income for a better provision of different benefits and services. Respondents were also asked if they supported the introduction of an EU-wide minimum wage (adjusted to the living costs in each EU Member State). The wording corresponds for the most part to a question on support of an EU-wide social benefit scheme in Wave 8 of the *European Social Survey*. Another question stems from the *World Value Survey* and asks whether respondents think that the government should take more responsibility to care for its people or rather the people should take care for themselves. As an extension of the question, respondents should additionally state whether they thought that the EU should take more responsibility to care of EU citizens or whether each Member State is foremost responsible to take care of its citizens. The module concludes with a question at what point respondents thought EU immigrants should obtain the same rights to social benefits as citizens and two questions on their ideas about parenting.

Media, Trust & COVID-19: This module asked respondents about their media usage (in particular, which media sources they used and how often) and their degree of trust in national

and EU political institutions. Since the survey was conducted during the beginning of the first wave of the COVID-19 pandemic, a question was implemented on whether the respondents or somebody in their surroundings showed symptoms of COVID-19 or had been tested positive.

Socio-Demographics: Questions comprise the employment status, occupation, place of residence, parenthood, and household composition.

2.2 Wave 2 (SOECBIAS follow-up)

The follow-up survey was carried out in Germany only. It repeats questions from the main survey and additionally asks about topics on COVID-19. The data can be analysed as a two-wave panel for the repeated questions. It contains variables on the following topics:

Income Position & Country Comparison: This module repeats questions from the main survey on the respondents' perceived own income positions (national and within the EU) and on their country's ranking compared to the EU average.

Social Policy Preferences I: This module contains repeated questions from the main survey on paying additional taxes or social contributions, the EU-wide minimum wage, the responsibility of national government and the EU, the rights of immigrants to social benefits, and on parenting. The module also has three new questions about whether the respondents thought that immigration is economically more beneficial or more harmful, both for themselves and for the German state, and whether they thought that future generations would fare better compared to their own generation.

Affectedness by COVID-19 Crisis: The module includes a series of questions that measure direct and indirect affectedness by the pandemic. As in the main survey, a simple question asked about COVID-19 symptoms and positive tests among friends, relatives, neighbours and work colleagues. Relying mainly on questions from *SOEP-CoV*, respondents were also questioned in more detail about these issues and whether any of their household members belonged to the risk groups. Respondents were also asked whether the pandemic has had a direct economic impact on them, i.e., on their employment status (including short-time work), their work situation, their net household income, and to which degree they thought they might lose their job, face financial difficulties, or they had to lower their standard of living in the next 12 months.

Social Policy Preferences II: Respondents were asked about their approval of the crisis management to curb the spread of COVID-19 of different entities: the federal government, the government of their federal state (Bundesland), and the EU. They also needed to evaluate how strongly they perceived the following groups to be economically affected by the COVID-19 crisis: themselves, friend/family/acquaintances, people in Germany, and people in the EU. The module concludes with a question which – if any – of the different policies implemented during the pandemic (e.g., short-time work) should continue to exist after the crisis in this form.

Media: In this module, respondents were asked to identify the most important information sources for different topics and to indicate whether they believed they could identify fake news.

2.3 Wave 3 (COVREF)

The data were collected in preparation for the COVREF project. In line with the main questions of the project, this survey wave is comprised of additional questions regarding affectedness by the COVID-19 crisis while repeating the modules on social policy preferences, income positions and country comparisons.

Affectedness by COVID-19 Crisis: In addition to repeating questions from the SOECBIAS follow-up survey, this module additionally included questions about test results and, if tested positive, about the date of the first positive test result.

2.4 Wave 4 (COVREF)

This wave reintroduces questions about social and political orientation as well as the sociodemographics module to collect baseline data for the refreshment samples drawn in waves 3 & 4 and repeats the social policy and income position and country comparison modules to complete the fourth panel wave for repeated questions. Additionally, the module on the COVID-19 crisis was repeated with some changes to complete the third panel wave on those questions.

Affectedness by COVID-19 Crisis: Questions from previous waves in this module have been repeated. In addition to that, retrospective data regarding one's own and one's household members history with COVID-19 infections and tests since the onset of the pandemic has

been collected. If respondents have not been participating in the labour market within the last seven days, they've been asked about their previous main occupation.

Sociodemographics II: This module includes repeated questions from the main survey on sociodemographics as well as new questions from the SOEP and ALLBUS. The SOEP questions are: Whether respondents have German citizenship and if they have it since birth, if their parents were born in Germany, how satisfied they are with their life overall and with their health, their household duties, the income of their household, their family life, democracy in Germany and, if applicable, their work and the available childcare.

2.5 Regional-level Identifiers

The SOECBIAS-COVREF data set includes a regional identifier for each respondent based on their place of residence: For Poland and Italy at NUTS 2 level for Sweden and Germany at NUTS 3 level.² This allows for merging other regional level indicators to the survey data.

2.6 Data on Actual Income Positions

The data set also includes variables on the respondents' actual national and EU income positions that was used in the treatments. These variables are based on EU-SILC yearly net household income data. We utilized the most recent data at our disposal, i.e., the 2018 data for the national position and the 2017 data for the EU position. These variables can be used, together with data on the income positions assessed by respondents themselves, to analyse the misperception of the respondents regarding their own income positions.

3 Sampling and Data Collection

About 1,500 persons were sampled in each country in wave 1 and re-contacted in Germany for three more waves. In addition, refreshment samples were drawn in wave 3 and wave 4.

3.1 Sample

For the main survey (wave 1) respondents were selected from the population of persons aged 18 and above in the respective countries by quota sampling in an online access panel. All waves were implemented and conducted by YouGov Deutschland (<https://yougov.de/>), a

² For Italy, the data set includes a substantial number of missings on the regions that respondents reside in (around 23%). The available data set has been subjected to anonymization, which has resulted in the coarsening of certain variables, including the regional identifier for Germany.

commercial survey company with a large pool of panellists who are invited to participate in surveys by email. The median completion time for all waves was about 13 minutes each. Since the data were collected by quota sampling, respondents were first channelled through four screening questions that ensured balanced samples as regards gender (male, female), age (18-33, 35-54, 55 or older), education (low, middle, high), and income.³ **Figure 2** depicts the design of the data collection process.

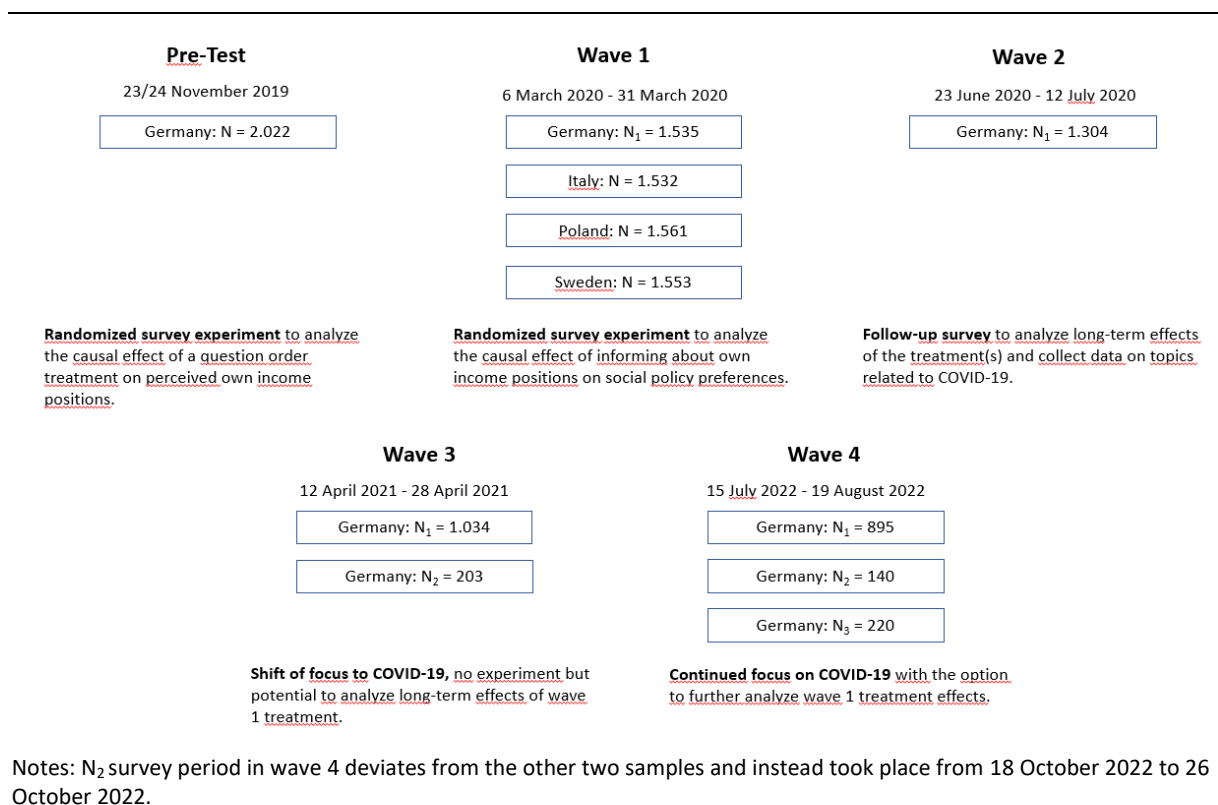


Figure 2: Data Collection

3.2 Pre-Test

Prior to the main survey, we performed a pre-test in Germany (N=2022) in November 2019, also conducted by YouGov Deutschland. The aim of the pre-test was two-fold: First, we tested the feasibility of the question on the own income position within the EU. The great majority

³ YouGov used country specific income brackets for the screening:

Germany (monthly net household income): 500€-brackets starting from 0 up to 4999€; 5000€-9999€; 10000€ or higher;

Italy (yearly gross household income): 5000€-brackets starting from 0 up to 49999€; 50000€-74999€; 75000€-99999€; 100000€ or higher;

Poland (monthly net household income): 0zł-999zł; 1000zł-1499zł; 1500zł-1999zł; 2000zł-2999zł; 3000zł-3999zł; 4000zł-4999zł; 5000-5999; 6000-6999zł; 8000zł-8999zł; 9000zł-9999zł; 10000zł-14999zł; 15000zł or higher;

Sweden (yearly net household income): 5000SEK-brackets starting from 0 up to 199999SEK; 10000SEK-brackets from 200000SEK to 799999SEK; 800000SEK-999999SEK; 1000000SEK or higher.

of respondents gave consistent answers on their own national and EU income position. Second, we implemented a question order treatment. The consistency of answers improved if respondents were *primed* first to think about their national income position before being asked about their EU income position.

3.3 SOECBIAS Wave 1 and 2

The main survey was conducted in four EU countries (Germany, Italy, Poland, and Sweden) in their respective main language. The sample sizes are 1535 respondents for Germany, 1532 for Italy, 1561 for Poland, and 1553 for Sweden, summing up to 6181 respondents. Data collection mostly took part in the second half of March; 255 of the respondents (58 in Germany, 64 in Italy, 65 in Poland, and 68 in Sweden) participated during a soft-launch phase in early March. To investigate potential long-term effects of the experimental treatments, we conducted a follow-up survey in Germany during June and July 2020. 1304 of the 1535 respondents of the main survey participated (recontact rate: 85%). The follow-up survey allowed us to include additional questions regarding how respondents are affected by the COVID-19 pandemic.

3.4 COVREF Wave 3 and 4

In April 2021 we recontacted participants from previous waves (N=1034) and additionally drew a refreshment sample from the same population (N=203). Following that, we conducted another survey in June and July 2022, recontacting respondents from the main survey (N=895) and drawing another refreshment sample of 220 respondents. It was intended to recontact the respondents of the wave 3 refreshment sample together with the other sample members. Due to a mistake at the level of the survey operations the respondents of the refreshment sample were only contacted in October 2022. 140 of the 203 respondents participated (recontact rate: 69%).

4 Data Cleaning Process and Data Quality

In the data cleaning process, we renamed and relabelled the variables in the raw data delivered by YouGov Deutschland for easier identification. For each variable, invalid values due to different reasons (e.g., respondents picked “don’t know” as their answer) were recoded into missings and labelled with the respective reason.

For Germany, we cross-referenced the reported yearly income with the income bracket selected by the respondents during the screening process for their monthly income. We found some inconsistency in the incomes and that ca. 15% of the respondents may have reported their monthly income instead. Thus, we added a flag variable that marks the inconsistency in the yearly income as well as the probability of reporting the monthly income by mistake. Additionally, we adjusted the yearly income of the German respondents who highly likely, possibly, or somewhat less likely reported their monthly income by multiplying the income by 12. As a result, we created three new variables with the adjusted yearly income (variant 1: for those who highly likely reported their monthly income; variant 2: for those who highly likely or possibly reported their monthly income; variant 3: for those who highly likely, possibly, or somewhat less likely reported their monthly income).

Finally, weights are available in the data set to adjust for small sample deviations from the overall population in each country concerning the quotation characteristics (i.e., gender, age, education, and income). Weights were constructed for all waves, based on data from the respective national censuses or similar sources (regarding gender, age, and education) and from the EU-SILC (regarding income). **Table 1** shows summary statistics of wave 1 and 2 compared with nationally representative statistics by country. Overall, our samples show a typical representation of the population along the dimensions of gender, age, and education. One exception is Sweden, where the middle-education group is slightly overrepresented. Our samples are also balanced as regards income and the non-targeted dimension of employment, except for Italy, where the third income bracket is underrepresented. After using weights, deviations are adjusted. The additional two columns for Germany in the table reveal that the sample of the follow-up survey is very similar to the main survey sample (for both the weighted and non-weighted data).

5 Analysis Potential

The SOECBIAS-COVREF data set offers a variety of opportunities for analysis. To name only a few:

- 1) *Social policy analyses*: The questions on social policy preferences allow a fine-grained analysis of support for social policy measures across the EU.
- 2) *Perception analyses*: A core feature of the data is the possibility to analyze (mis-) perceptions and their origins. Misperceptions may exist with regard to respondents'

own income perceptions (nationally and EU), the position of their country, their social status and the status of their surroundings. Origins for these misperceptions can be analyzed with regard to sociodemographic factors (age cohorts, gender, and education level), political affiliation, or media usage.

- 3) *Analyses of treatment effects*: The main survey (wave 1) includes a survey experiment to analyze the impact of informing respondents about their own income position in the (national and EU) income distribution on their social policy preferences. Within a Randomized Controlled Trial (RCT), any difference between treatment and control groups can be attributed to the information provided to respondents. The wide range of questions allow analyzing treatment effects on several different issues.
- 4) *Analyses on the consequences of COVID-19 pandemic*: Wave 2 to 4 include a variety of questions on the assessment of different social policies that played an important role during the pandemic. In contrast to most current studies, SOECBIAS-COVREF data contain additional information about public opinion on whether some of these policies (e.g. short-term work or stricter border controls) should persist beyond the crisis. Therefore, the data have direct policy implications that demand further elaboration and are of public interest beyond the confines of a specialist audience.
- 5) *Evaluations of self and others*: A main set of questions on the consequences of the COVID-19 pandemic is directed at how respondents evaluate their own situation and the affectedness of others. Similar questions were asked on subjective social status. This allows for analyses on the role of comparisons with others.
- 6) *Spatial and temporal analyses*: The nature of the data provides opportunities to analyze differences in the respective variables not only at the national level (or cross-country comparisons) but also at the regional-level as well as changes over time. Hence, with additional data, regional variation can be addressed. This seems particularly important since the field phase took place during the COVID-19 pandemic. Regional data on infections, unemployment and other indicators can be used to control for differences in affectedness across regions and countries.
- 7) *Comparative analyses*: SOECBIAS-COVREF data includes several questions from widely recognized cross-country surveys. This allows testing the external validity of the results obtained from the dataset.

Table 1: Sample Characteristics Wave 1 & 2 (in %)

	Germany					Italy			Poland			Sweden		
	Wave 1	Wave 1 (wt.)	Wave 2	Wave 2 (wt.)	EU-SILC	Wave 1	Wave 1 (wt.)	EU-SILC	Wave 1	Wave 1 (wt.)	EU-SILC	Wave 1	Wave 1 (wt.)	EU-SILC
<i>Female</i>	52.3	51.6	51.8	51.7	50.9	51.9	52.3	52.0	54.2	52.2	52.3	53.6	50.5	50.1
<i>18-34 years old</i>	22.8	24.1	19.2	24.1	22.9	22.0	22.9	20.8	33.8	32.4	26.3	20.0	27.6	27.2
<i>35-54 years old</i>	36.4	36.2	37.4	36.2	32.9	38.2	37.0	34.9	38.1	33.7	34.1	35.5	33.3	32.0
<i>55 years or older</i>	40.8	39.7	43.5	39.7	44.2	39.8	40.2	44.2	28.1	33.9	39.6	44.5	39.1	40.7
<i>Low educ.</i>	23.4	20.9	23.5	20.9	17.5	44.3	51.1	46.8	10.6	20.6	16.1	22.1	22.2	24.9
<i>Middle educ.</i>	47.8	53.2	46.9	53.2	56.7	41.3	36.2	37.2	65.0	60.1	57.7	53.0	51.0	42.8
<i>High educ.</i>	28.8	25.9	29.6	25.9	25.8	14.4	12.7	15.9	24.4	19.3	26.3	24.9	26.8	32.3
<i>Income bracket 1</i>	38.2	39.0	37.8	39.0	37.6	34.3	26.5	26.5	25.6	33.0	34.1	23.8	27.0	27.5
<i>Income bracket 2</i>	36.2	34.0	37.0	34.0	31.4	53.4	46.2	46.2	44.9	38.0	38.7	47.3	48.0	45.8
<i>Income bracket 3</i>	25.6	27.0	25.2	27.0	31.1	12.3	27.3	27.3	29.5	29.0	27.3	28.8	25.0	26.7
<i>Employed</i>	54.9	55.4	54.0	55.2	55.9	46.0	44.9	45.9	60.0	54.6	52.9	53.0	49.0	58.0

Notes: This table displays unweighted and weighted (wt.) summary statistics from the Main Survey (Wave 1) and the Follow-up Survey (Wave 2) alongside nationally representative statistics (using EU-SILC 2018). Weights were constructed based on variables in the screening. Country-specific income brackets are: 1) for Germany (monthly net household income in Euro): 0-1999, 2000-3499, 3500 or higher; 2) for Italy (yearly gross household income, in Euro): 0-19999, 20000-49999, 50000 or higher; 3) for Poland (monthly net household income, in Złoty): 0-2999, 3000-5999, 6000 or higher; 4) for Sweden (yearly net household income, in Swedish krona (SEK)): 0-199999, 200000-499999, 500000 or higher.

Acknowledgments

SOECBIAS project:



Gefördert durch:



COVREF project:



Seed funding was granted in 2021 by the Faculty of Business, Economics and Social Sciences / University of Hamburg for data collection in preparation for the COVREF project.