

Gauer, Miriam; Kristen, Cornelia

## Research Report

A guide to using the SOEP for research on individuals of immigrant origin. Samples and contents of the SOEP-Core, including the IAB-SOEP Migration Samples and the IAB-BAMF-SOEP Survey of Refugees, through a migration and immigrant integration lens

SOEP Survey Papers, No. 1332

## Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Gauer, Miriam; Kristen, Cornelia (2023) : A guide to using the SOEP for research on individuals of immigrant origin. Samples and contents of the SOEP-Core, including the IAB-SOEP Migration Samples and the IAB-BAMF-SOEP Survey of Refugees, through a migration and immigrant integration lens, SOEP Survey Papers, No. 1332, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/278807>

### Standard-Nutzungsbedingungen:

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

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

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

### Terms of use:

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

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

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



<https://creativecommons.org/licenses/by-sa/4.0/>

# 1332<sup>2023</sup>

**SOEP** Survey Papers

Series C - Data Documentations (Datendokumentationen)

## A guide to using the SOEP for research on individuals of immigrant origin.

Samples and contents of the SOEP-Core, including the IAB-SOEP Migration Samples and the IAB-BAMF-SOEP Survey of Refugees, through a migration and immigrant integration lens

Miriam Gauer and Cornelia Kristen

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/soepsurveyspapers>

#### **Editors:**

Dr. Carina Cornesse, DIW Berlin and University of Bremen

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Cornelia Kristen, University of Bamberg and DIW Berlin

Prof. Dr. Philipp Lersch, DIW Berlin and Humboldt-Universität zu Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Prof. Dr. Sabine Zinn, DIW Berlin and Humboldt-Universität zu Berlin

Please cite this paper as follows:

Miriam Gauer and Cornelia Kristen. 2023. A guide to using the Socio-Economic Panel for research on individuals of immigrant origin. SOEP Survey Papers 1332. Series C. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2023 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin  
German Socio-Economic Panel (SOEP)  
Mohrenstr. 58  
10117 Berlin  
Germany

[soeppapers@diw.de](mailto:soeppapers@diw.de)

# A guide to using the Socio-Economic Panel for research on individuals of immigrant origin.

Samples and contents of the SOEP-Core, including the IAB-SOEP Migration Samples and the IAB-BAMF-SOEP Survey of Refugees, through a migration and immigrant integration lens

**Miriam Gauer and Cornelia Kristen**

*Berlin, 2023*

# 1 Introduction

The Socio-Economic Panel (SOEP) was established in 1984 as a longitudinal study of private households in Germany. It collected data on migration and immigrant integration from the outset. Over time, samples and contents were expanded, resulting in an extensive database that enables researchers to examine immigration-related processes over time (Brücker et al. 2014; Brücker et al. 2016; Goebel et al. 2019; Krupp 2008; Liebau and Tucci 2015).

For nearly three decades, from 1984 to 2012, the German Socio-Economic Panel at the German Institute for Economic Research (DIW Berlin) was responsible for designing and conducting the SOEP. In 2013, in collaboration with the Institute for Employment Research (IAB), the IAB-SOEP Migration Sample was created and implemented. Subsequent IAB-SOEP Migration Samples were also jointly designed and drawn in later years (2015 and 2020). In 2015, a collaborative effort involving three institutes – the Institute for Employment Research (IAB), the Research Centre of the Federal Office for Migration and Refugees (BAMF-FZ), and the German Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin) – led to the establishment of the IAB-BAMF-SOEP Survey of Refugees. Together, these institutes planned, developed, and conducted the data collections specifically focused on refugees.

As we approach four decades since the SOEP's establishment, this guide takes stock of these accomplishments and documents the various samples and contents that are relevant for research on immigrants and their descendants. We also provide detailed information on when and how these contents were measured. Our guide serves as a tool for scholars to quickly assess whether the data is suitable for their research objectives.

This compendium draws upon two prior works that have identified and organized key contents relevant to research on individuals of immigrant origin: Liebau and Tucci's (2015) summary of migration-specific topics up until that time, and Jacobsen, Krieger, Schikora, and Schupp's (2021) outline of the SOEP's potential for migration research. Our documentation builds on these works, updating and modifying the descriptions to reflect ongoing developments and providing detailed information on the measurement of key contents. While following in their footsteps, we also deviate to some extent from these earlier contributions in our presentation of the various constructs and the ways in which we organize and assign them to higher-level topics.

Section 2 presents an overview of the various migration subsamples. The centerpiece of this guide is Section 3, where the migration and integration constructs covered in the SOEP are introduced. In Section 4, detailed information on the measurement of these constructs is provided. Finally, Section 5 offers practical guidance on how to utilize the SOEP's migration and integration data.

## 2 Individuals of immigrant origin in the various SOEP subsamples

### 2.1 Sampling individuals of immigrant origin

The SOEP is comprised of multiple subsamples of private households that were added sequentially. Most subsamples are two-stage random samples (Jacobsen et al. 2021: 530). In the first stage, primary sampling units (PSUs) are defined, such as addresses within regional clusters. Stratification by region is then typically used to sample PSUs. In the second stage, individuals or addresses are randomly sampled from the selected units, which are referred to as secondary sampling units (SSUs). Disproportionate sampling is often employed at this stage to ensure sufficient sample sizes for smaller populations. Because the SOEP is a household study, all adult household members are included, not just the sampled respondent (i.e., the anchor person; SOEP Group 2001). Design weights are utilized to account for the sampling design and to integrate the various SOEP subsamples (Kroh, Siegers and Kühne 2015). It is generally recommended to use the SOEP as a whole and not just specific subsamples (Jacobsen et al. 2021: 531).

All subsamples are named with a letter, starting with A for the first sample drawn in 1984, and new subsamples have been added almost every year. Some of these subsamples represent specific subpopulations, such as L2 for single parent families, P for top earners, Q for LGBTQ+ people, or M for immigrants and refugees, while others are general population samples.

Immigrants and their descendants can be included in the SOEP subsamples in several ways (Liebau and Tucci 2015: 3-4). Firstly, they may be part of a general population sample. Secondly, they may be oversampled within a general population sample, as was done for subsamples F, I, J, and L1. Thirdly, they may be included in specific migration samples that use sampling frames of migrants instead of the general population. These subsamples are of particular interest to researchers on migration and immigrant integration, as they contain significant numbers of individuals of immigrant origin and cover migration-specific topics. Currently, the SOEP includes the following migration subsamples: B, D, M1, M2, M3, M4, M5, M6, M7, and M8.

### 2.2 Sample descriptions

Below, we provide a brief overview of the different migration subsamples and general population subsamples in which immigrants were oversampled. For each subsample, we will specify the target population, summarize the sampling design, and report on the translations of the German questionnaires into other languages. These translations are particularly important for newly arrived immigrants who may not yet be proficient in German and could otherwise be excluded from participating in the survey.

The incremental addition of various subsamples reflects a range of migratory developments and events in Germany over time. The SOEP initially covered labor migrants and their family members from Turkey and southern European countries. Later, it included immigrants from

Eastern European countries after the fall of the Iron Curtain and the enlargement of the EU. More recently, it addressed the surge in refugee migration. With each new subsample, the database reflects the increasing diversification of the immigrant population in terms of composition and structure.

### **The early migration subsamples (B, D)**

#### **Subsample B, 1984**

- *Target population:* Sample B “Foreigners in the Federal Republic of Germany” is one of the two initial samples of the SOEP. It covers so-called “guest-workers” (“Gastarbeiter”) who arrived in West Germany before 1984 following the recruitment agreements (“Anwerbeabkommen”) between Germany and several European countries (Hanefeld, 1987). The sample consists of private households with a Greek, Italian, Spanish, Turkish or Yugoslavian household head (Siegers, Steinhauer and Dührsen 2021: 10).
- *Sampling:* Sample B is based on a multistage stratified sampling procedure using the Central Register of Foreign Nationals (“Ausländerzentralregister”, AZR) in each county (Siegers et al. 2021: 10).
- *Translations:* Greek, Italian, Spanish, Turkish, Yugoslavian

#### **Subsample D, 1994/1995**

- *Target population:* Sample D “Immigrants” includes private households in which at least one household member had moved from abroad to West Germany between 1984 and 1994 (Siegers et al. 2021: 12). It added recently arrived individuals from a wide range of countries. The largest group were immigrants of German descent who had left one of the states that, until 1991, had belonged to the Soviet Union (Liebau and Tucci 2015: 7, 38-40; Schupp and Wagner 1995).
- *Sampling:* Sample D includes two subsamples that were drawn independently in 1994 and in 1995. The households were identified in representative surveys of the German population driven by Infratest and following the ADM-Design (Association of German Market and Social Research Institutes; Siegers et al. 2021: 12).
- *Translations:* Greek, Italian, Spanish, Turkish, Yugoslavian

### **General population samples with an oversampling of individuals of immigrant origin (F, I, J, L1)**

#### **Subsample F, 2000**

- *Target population:* Sample F “Refreshment II” covers private households in Germany and was implemented to increase the sample size of the SOEP. As migrant households display lower response probabilities, households with at least one adult not having German nationality were oversampled (Siegers et al. 2021: 15). The refreshment covered individuals of various origins, reflecting the increasingly heterogeneous composition of the migrant population in Germany (Liebau and Tucci 2015: 9, 38-40).

- *Sampling*: To draw sample F, a multistage stratified sampling procedure based on the ADM-Design was applied (Siegers et al. 2021: 15).
- *Translations*: Greek, Italian, Spanish, Turkish, Yugoslavian

### **Sample I, 2009**

- *Target population*: Sample I, also known as “Innovation Sample”, was implemented to increase the representation of migrant households in the SOEP (Liebau and Tucci 2015: 10). It includes individuals with an immigration history linked to a wide range of countries (Liebau and Tucci 2015: 38-40) residing in private households in Germany.
- *Sampling*: Sample I was drawn using a multistage stratified sampling procedure based on the ADM-Design (Siegers et al. 2021: 18). The identification of individuals of immigrant origin was based on an onomastic method. In 2012, Sample I was fully integrated into SOEP-IS and is no longer part of the SOEP-Core.
- *Translations*: English, Greek, Italian, Spanish, Turkish, Yugoslavian

### **Sample J, 2011**

- *Target population*: Sample J, also known as “Refreshment IV”, was aimed at increasing the representation of migrant households in the SOEP (Siegers et al. 2021: 19). It covers private households in Germany and employs a disproportional sampling design. The sample includes individuals of various origins (Liebau and Tucci 2015: 38-40).
- *Sampling*: Sample J was drawn using a multistage stratified sampling procedure based on the ADM-Design (Siegers et al. 2021: 19).
- *Translations*: English, Russian, Turkish

### **Sample L1, 2010**

- *Target population*: Sample L1, also known as the “Cohort Sample”, covers private households in Germany, in which at least one child that was born between January 2007 and March 2010 (Siegers et al. 2021: 19). Again individuals of immigrant origin are oversampled, and the sample includes individuals from various backgrounds (Liebau and Tucci 2015: 38-40).
- *Sampling*: Sample L1 was drawn using a multistage stratified sampling procedure based on information from local registration offices (“Einwohnermeldeämter”). Individuals of immigrant origin were identified employing an onomastic method.
- *Translations*: English, Turkish

### **IAB-SOEP Migration Samples (M1, M2, M7, and M8)**

The IAB-SOEP Migration Samples were collaboratively developed and conducted by the Institute for Employment Research (IAB) and the German Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin).



### **Subsample M1, 2013**

- *Target population:* The 2013 “IAB-SOEP Migration Sample” was designed to increase the representation of first-generation immigrants who came to Germany between 1995 and 2010, as well as second-generation individuals who were born in Germany after 1976 (Brücker et al. 2014; Siegers et al. 2021: 24). In response to changes in the scale and composition of immigration to Germany following the EU’s Eastern enlargement and the financial and economic crisis, this sample added a significant number of individuals with an immigration history from countries such as Poland, Romania, Kazakhstan, and other states of the former Soviet Union, as well as Turkey and Southern European countries (Brücker et al. 2014: 3-4, 10).
- *Sampling:* The Integrated Employment Biographies (IEB) dataset from the Federal Employment Agency (“Bundesagentur für Arbeit”, BA) was used as the sampling frame (Brücker et al. 2014; Siegers et al. 2021: 24). Sample M1 was drawn using a multistage stratified sampling procedure (Kroh, Kühne et al. 2015).
- *Translations:* English, Polish, Romanian, Russian, Turkish

### **Subsample M2, 2015**

- *Target population:* Study M2 follows the 2013 “IAB-SOEP Migration Sample” M1. The M2 sample was specifically designed to collect information on recently arrived immigrants who came to Germany between 2009 and 2013 (Siegers et al. 2021: 25). This sample encompasses a range of immigrant groups, including Eastern Europeans such as Poles and Romanians as well as new immigrants from Southern European countries (Kühne and Kroh 2017).
- *Sampling:* Similar to M1, the Integrated Employment Biographies (IEB) dataset from the Federal Employment Agency (“Bundesagentur für Arbeit”, BA) served as the sampling frame (Siegers et al. 2021: 24). The M2 sample was drawn using a multistage stratified sampling procedure (Kühne and Kroh 2017).
- *Translations:* English, Polish, Romanian, Russian, Turkish

### **Subsample M7, 2020**

- *Target population:* M7 was developed to enhance the “IAB-SOEP Migration Sample”. It aimed to capture migration dynamics from 2016 to 2018, with a particular focus on EU migration and the three most prominent countries of origin during that period: Bulgaria, Poland and Romania (Rathje and Glemser 2021: 8, 143; Siegers, Steinhauer and Schütt 2022: 34; Steinhauer, Trübswetter and Zinn 2022: 1).
- *Sampling:* Similar to the M1 and M2 samples, the Integrated Employment Biographies (IEB) dataset from the Federal Employment Agency (“Bundesagentur für Arbeit”, BA) served as the sampling frame (Siegers et al. 2022: 34; Steinhauer et al. 2022). The M7 sample was drawn using a multistage stratified sampling procedure (Siegers et al. 2022: 34; Steinhauer et al. 2022: 2-3).
- *Translations:* Bulgarian, English, Romanian, Polish

### **Subsample M8, 2020**

- *Target population:* The 2020 boost sample M8 supplements the existing “IAB-SOEP Migration Sample”. It focuses on third-country nationals from outside the European Union who applied to work as professionals (“Fachkräfte”) under the Residence Act (“Zuwanderungsgesetz”) and were granted permission between January 2019 and January 2020 (Siegers et al. 2022: 35; Steinhauer et al. 2022: 1). This sample provides a foundation for evaluating the “Fachkräfteeinwanderungsgesetz” which came into effect in March 2020.
- *Sampling:* To sample third-country nationals, the study used the Integrated Employment Biographies (IEB) dataset provided by the Federal Employment Agency (BA; Siegers et al. 2022: 35; Steinhauer et al. 2022). Similar to the previous samples, the M8 sample was drawn using a multistage stratified sampling procedure (Siegers et al. 2022: 35; Steinhauer et al. 2022: 2-3).
- *Translations:* Bulgarian, English, Romanian, Polish

### **IAB-BAMF-SOEP Refugee Samples (M3, M4, M5, and M6)**

The IAB-BAMF-SOEP Survey of Refugees was collaboratively developed and conducted by the Institute for Employment Research (IAB), the Research Centre of the Federal Office for Migration and Refugees (BAMF-FZ), and the German Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin).

The survey was specifically designed to gather information on refugees (Brücker, Rother and Schupp 2016; Brücker, Rother and Schupp 2017; Kroh et al. 2016; Kroh et al. 2017). It initially began with two sub-samples known as M3 and M4 (following the IAB-SOEP Migration Samples M1 and M2). Subsequently, additional refresher and enlargement samples were drawn, resulting in the creation of the M5 and M6 samples (Jacobsen et al. 2019; Steinhauer et al. 2022).

### **Subsample M3/M4, 2016**

- *Target population:* The M3 and M4 samples were drawn from a target population comprising households that included individuals who arrived in Germany between January 2013 and January 2016 and either applied for asylum or were accommodated through specific programs implemented by the federal states (Brücker, Rother and Schupp 2017; Kroh et al. 2017: 6; Siegers et al., 2021: 26). Significant refugee groups with notable representation included those originating from Syria, Afghanistan, and Iraq (Kroh et al. 2016: 5; Kroh et al. 2017: 30).
- *Sampling:* The sampling strategy consisted of a multistage stratified sampling design based on the German Central Register of Foreign Nationals (“Ausländerzentralregister – AZR”; Siegers et al. 2021: 26; Kroh et al. 2017: 6-8), which is a register database of information on foreigners living in Germany.
- *Translations:* Arabic, English, Farsi/Dari, Kurmanji, Pashto, Urdu

### **Subsample M5, 2017**

- *Target population:* Sample M5 served as both an enlargement and a refreshment of the previous samples M3 and M4 (Jacobsen et al. 2019; Siegers et al. 2021: 27). It encompassed individuals who migrated to Germany between January 2013 and January 2016 and subsequently made a claim for asylum between April 2016 and January 2017, which accounted for late registrations (refreshment). These respondents were entered in the German Central Register of Foreign Nationals (“Ausländerzentralregister – AZR”) after April 2016 and were therefore not part of the target population of M3 or M4 (Jacobsen et al. 2019: 7). In addition to these late registrations, the M5 sample included individuals who immigrated to Germany between February 2016 and December 2016 and made an asylum claim until January 2017, representing new arrivals (enlargement; Jacobsen et al. 2019: 7; Siegers et al. 2021: 27).
- *Sampling:* The sampling methodology for the M5 sample was similar to that of the M3 and M4 samples. It involved the utilization of a multistage stratified sampling design, which was based on the German Central Register of Foreign Nationals (“Ausländerzentralregister – AZR”; Jacobsen et al. 2019: 7-10).
- *Translations:* Arabic, English, Farsi/Dari, Kurmanji, Pashto, Urdu

### **Subsample M6, 2020**

- *Target population:* The M6 sample was specifically designed to enhance the previous refugee samples (Siegers et al. 2022: 33; Steinhauer et al. 2022: 1). It comprised two distinct groups (Siegers et al. 2022: 33; Steinhauer et al. 2022: 1). The first group consisted of individuals who arrived in Germany between January 2013 up to the end of December 2016, filed an asylum application, and had their last change of asylum status between 2013 and the end of 2016 (refreshment). The second group comprised individuals who entered Germany between January 2013 and June 2019, filed an asylum application, and had their last change of asylum status between 2017 and June 2019 (enlargement). As a result, the previous samples M3, M4 and M5 were both refreshed and expanded to include individuals whose last asylum status entry fell between January 2017 and June 2019 (Steinhauer et al. 2022: 2).
- *Sampling:* Similar to the previous samples, the M6 sample employed a multistage stratified sampling design, utilizing the German Central Register of Foreign Nationals (“Ausländerzentralregister – AZR”; Steinhauer et al. 2022: 2).
- *Translations:* Arabic, English, Farsi/Dari, French, Kurmanji, Pashto, Urdu

## **2.3 Representation of individuals of immigrant origin in the SOEP**

This section provides an overview of the size of the immigrant population in the various SOEP samples. It includes immigrants and their descendants who have been part of the previously mentioned subsamples, as well as individuals from the general population samples who were not mentioned in the previous section.

Table 1: Number of respondents of immigrant origin in the various SOEP subsamples

Sample	Year	Households	Individuals	First generation	Second generation	Origin groups with at least 50 cases	
						First generation	Second generation
A	1984	4.524	11.366	530	148	n.a.	n.a.
B	1984	1.393	4.807	3.408	1.305	BA, HR, GR, IT, ES, TR	TR, GR, IT, ES
C	1990	2.179	6.044	155	52	n.a.	n.a.
D1	1994	236	719	447	68	KZ, PL, RO, RU	n.a.
D2	1995	295	905	620	108	KZ, PL, RO, RU	n.a.
E	1998	1.056	2.376	208	62	X	n.a.
F	2000	6.043	13.871	1.543	704	IT, KZ, PL, RO, RU, TR	n.a.
G	2002	1.224	3.364	140	117	n.a.	n.a.
H	2006	1.506	3.239	264	145	n.a.	n.a.
I	2009	1.495	3.052	395	242	TR	n.a.
J	2011	3.136	6.308	794	491	KZ, PL, RU, TR	n.a.
K	2012	1.526	3.036	384	270	PL, RU	n.a.
L1	2010	2.074	7.670	1.116	1.663	KZ, PL, RU, TR, XK	TR
L2	2010	2.500	8.838	677	1.094	KZ, PL, RU, TR	n.a.
L3	2011	924	3.579	155	376	n.a.	n.a.
M1	2013	2.723	7.445	4.667	2.976	BA, HR, GR, IQ, IT, KZ, KG, PL, RO, RU, RS, ES, SY, TR, UA, XK	TR, GR, IT
M2	2015	1.096	2.638	2.412	451	AF, BG, GR, HU, IQ, IT, PL, RO, RO, RS, ES, SY,	n.a.
M3	2016	3.273	3.959	3.690	266	AF, AL, ER, IR, IQ, PK, RU, RS, SY, XK	n.a.
M4	2016	3.273	5.897	5.403	493	AF, AL, ER, PS, IR, IQ, RU, RS, SY, XK	n.a.
M5	2017	1.519	4.161	3.826	332	AF, ER, IR, IQ, PK, SO, SY	n.a.
M6	2020	1.141	3.177	2.641	526	AF, ER, GQ, IR, IQ, NG, SO, SY	n.a.
M7	2020	783	1.993	1.854	113	BG, PL, RO	n.a.
M8	2020	1.096	1.979	1.836	87	AL, BA, BR, IN, RU, RS, MK, XK	n.a.
N	2017	2.378	4.807	404	464	PL, RU	n.a.
O	2018	935	1.730	544	359	PL, SY, TR	n.a.
P	2019	1.960	3.589	303	317	n.a.	n.a.
Q	2019	477	636	52	51	n.a.	n.a.

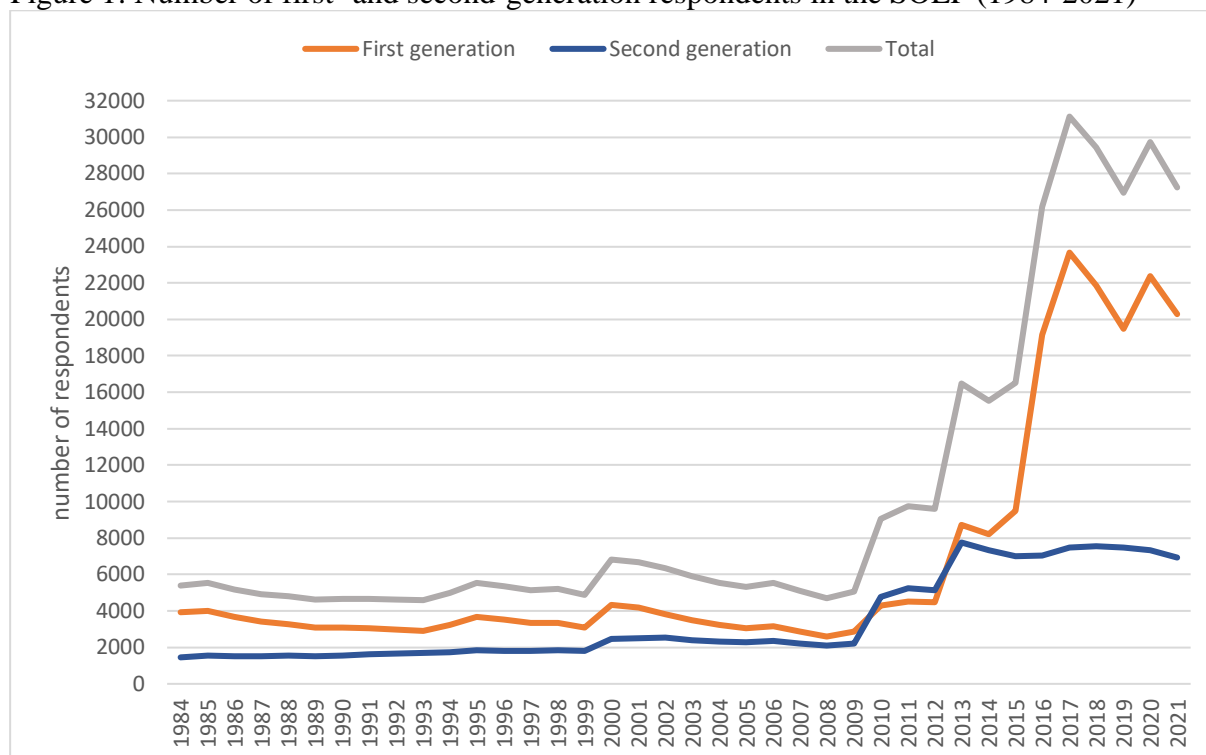
Notes: Siegers et al. (2022) and own calculations based on SOEP v38.

n.a. = not applicable; in these samples, none of the origin groups reaches 50 cases.

Origin groups: AF = Afghanistan, AL = Albania, BA = Bosnia and Herzegovina; BG = Bulgaria, BR = Brazil, ER = Eritrea, ES = Spain, GQ = Equatorial Guinea, GR = Greece, HR = Croatia, HU = Hungary, IN = India, IQ = Iraq, IR = Iran, IT = Italy, KG = Kyrgyzstan, KZ = Kazakhstan, MK = North Macedonia, NG = Nigeria, PK = Pakistan, PL = Poland, PS = Palestine, RO = Romania, RU = Russia, RS = Serbia, SO = Somalia, SY = Syria, TR = Turkey, UA = Ukraine, XK = Kosovo.

Table 1 presents the number of individuals of immigrant origin included in the different samples. It offers information about the number of first- and second-generation respondents and identifies which origin groups have a significant presence. The criteria used to assign individuals to the first and second generation will be explained in Section 3.1. It is worth noting that, due to household sampling, there can be individuals without a migration history in the general population samples.

Figure 1: Number of first- and second-generation respondents in the SOEP (1984-2021)



Notes: Own calculations based on SOEP v38.

Figure 1 illustrates the trends in the numbers of first and second-generation respondents in the SOEP over the years. Figures 2, 3, and 4 present the quantitative trends for various immigrant groups by country of origin and generation status since 1984. Figure 2 depicts first-generation immigrants of various origins. Figure 3 represents first-generation refugees. Figure 4 illustrates second-generation respondents categorized based on their parents' origins. For detailed information on the numbers underlying these figures, please refer to Tables 5, 6, and 7 in the Annex.

Due to the specific focus on recent immigrants, as evidenced by the establishment of a range of migration-specific subsamples since 2013 (see section 2.2), the number of first-generation immigrants covered in the SOEP substantially exceeds that of second-generation respondents.

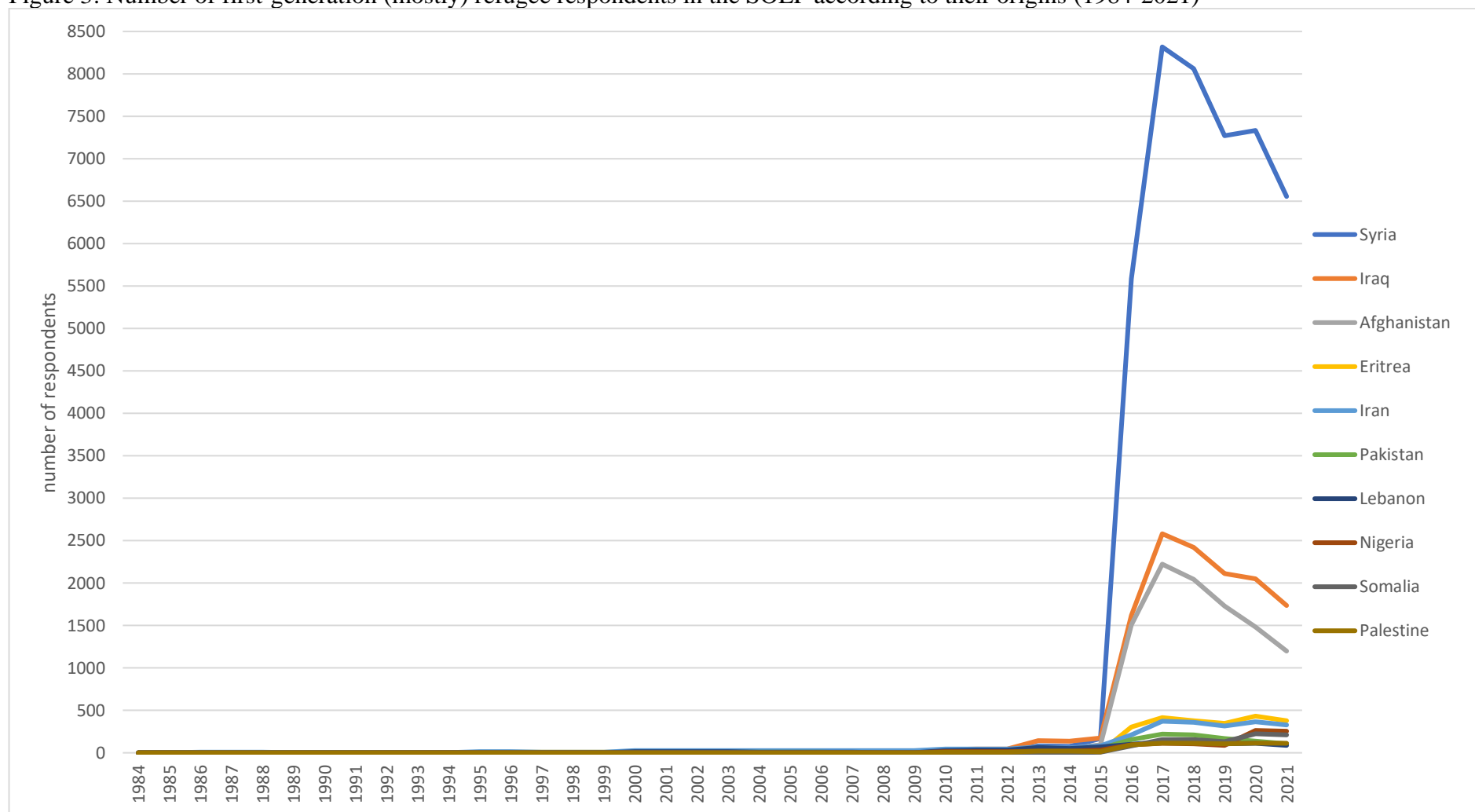
The population considered for these figures comprises individuals who have been sampled in any of the samples discussed before (see section 2.2 and Table 1). To be included in the figures, origin groups had to meet two criteria. Firstly, there had to be at least 500 interviews conducted since 1984 for the respective origin group. Secondly, in at least one year, there had to be at least 50 interviews for that origin group.

The chart displays the number of respondents for 25 different countries over a 37-year period. The y-axis, labeled 'number of respondents', ranges from 0 to 1300 in increments of 100. The x-axis shows years from 1984 to 2021. The data series are as follows:

- Turkey (Orange line): Starts at ~1260 in 1984, peaks at ~1290 in 1985, then declines to ~650 by 2021.
- Former Yugoslavia (Grey line): Starts at ~230 in 1984, remains relatively flat until 2012, then rises sharply to ~1180 in 2020 before a slight dip in 2021.
- Poland (Yellow line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~1150 in 2020 before a slight dip in 2021.
- Russia (Light Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~1100 in 2013, peaks at ~1080 in 2016, and declines to ~630 by 2021.
- Italy (Green line): Starts at ~630 in 1984, peaks at ~650 in 1985, then declines to ~220 by 2021.
- Kazakhstan (Dark Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~880 in 2020 before a slight dip in 2021.
- Greece (Brown line): Starts at ~470 in 1984, declines to ~130 by 2021.
- Romania (Dark Grey line): Starts at ~230 in 1984, remains low until 2012, then rises sharply to ~950 in 2013, peaks at ~740 in 2015, and declines to ~460 by 2021.
- Spain (Olive line): Starts at ~410 in 1984, declines to ~100 by 2021.
- Bulgaria (Medium Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~940 in 2013, peaks at ~740 in 2015, and declines to ~430 by 2021.
- Ukraine (Dark Green line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~240 in 2013, peaks at ~270 in 2015, and declines to ~210 by 2021.
- Czech (Light Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~510 in 2013, peaks at ~510 in 2015, and declines to ~240 by 2021.
- Austria (Light Orange line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Hungary (Grey line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Albania (Yellow line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- France (Light Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- USA (Green line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Netherlands (Dark Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Kyrgyzstan (Brown line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Morocco (Dark Grey line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- India (Olive line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Portugal (Medium Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Brazil (Dark Green line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.
- Azerbaijan (Light Blue line): Starts at ~40 in 1984, remains low until 2012, then rises sharply to ~110 in 2013, peaks at ~110 in 2015, and declines to ~100 by 2021.

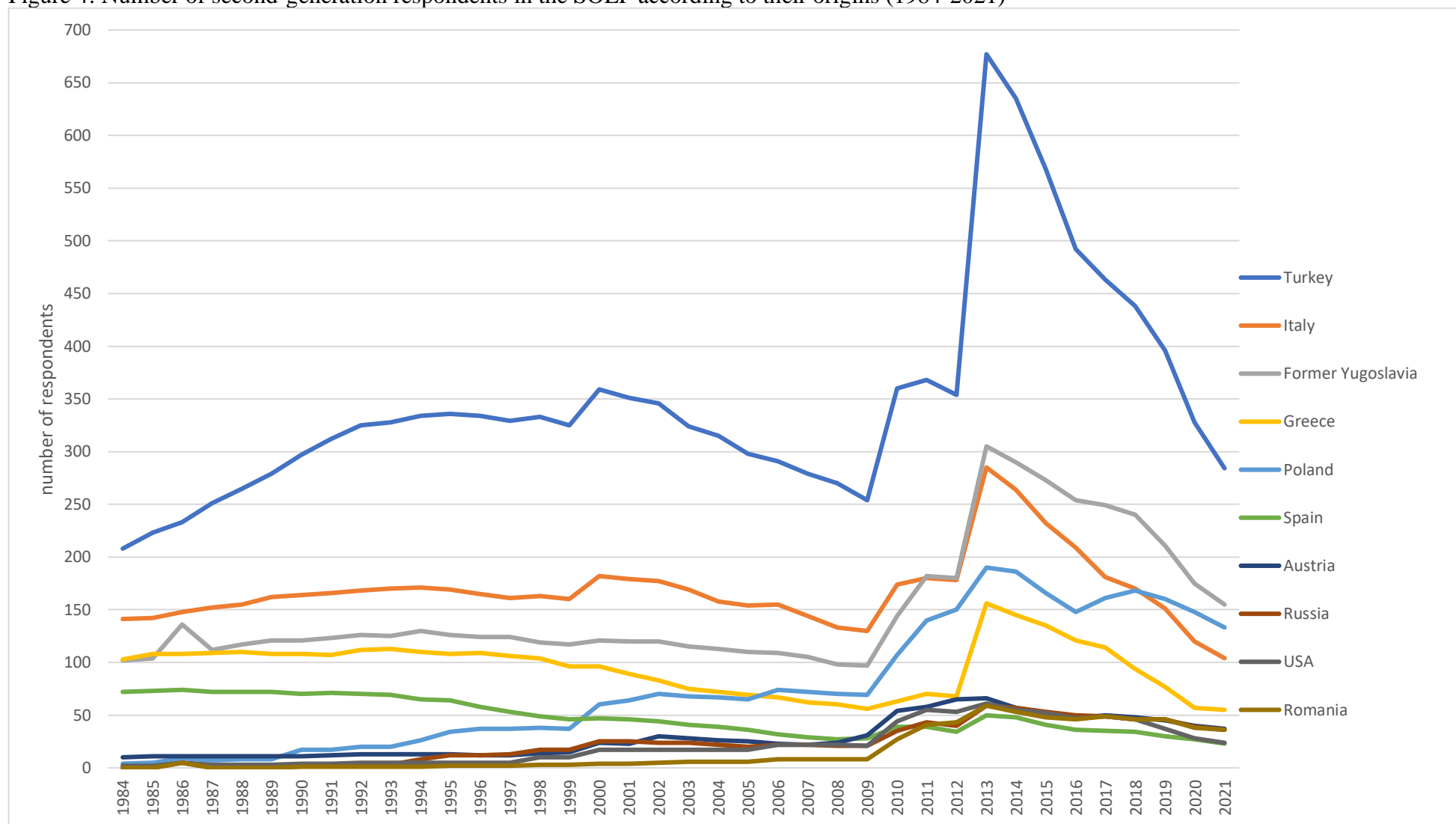
Notes: Own calculations based on SOEP v38. [Table 5](#) in the Annex presents the data used to generate this figure.

Figure 3: Number of first-generation (mostly) refugee respondents in the SOEP according to their origins (1984-2021)



Notes: Own calculations based on SOEP v38. [Table 6](#) in the Annex presents the data used to generate this figure. The majority of individuals covered here are refugees. However, it is important to note that the calculations are based on country of origin information rather than legal status. As a result, the table also includes individuals who did not immigrate as refugees.

Figure 4: Number of second-generation respondents in the SOEP according to their origins (1984-2021)



Notes: Own calculations based on SOEP v38. [Table 7](#) in the Annex presents the data used to generate this figure.



### 3 Contents relevant for studying individuals of immigrant origin

This guide provides an overview of the migration and integration constructs covered in the SOEP. A crucial prerequisite for researching immigrants and their descendants is to identify individuals with a migration history in the data. In this regard, we will outline the common indicators and procedures used for this purpose (“Immigrant Origin”, Section 3.1). We will then introduce measures that capture immigrants’ movements, such as leaving their country of origin and the early settlement period (“Migration”; Section 3.2). The most significant part of this guide focuses on the contents of immigrant integration (“Immigrant Integration”; Section 3.3). Drawing on a general conceptual framework, we will introduce various constructs implemented in the different subsamples. These constructs can be used to study the integration trajectories of individuals of immigrant origin across various outcomes, including cultural, structural, social, and emotional aspects. Table 2 provides an overview of the structure of this guide and specifies the various constructs covered in the SOEP-Core, including the IAB-SOEP Migration Samples and the IAB-SOEP-BAMF Survey of Refugees.<sup>1</sup>

Our presentation is informed by Liebau and Tucci’s (2015: 23-24) discussion of migration-specific topics and Jacobsen, Krieger, Schikora, and Schupp’s (2021: 539-546) outline of the SOEP’s potential for migration research. While drawing inspiration from their work, this overview differs from these earlier contributions in its structure, the inclusion of specific constructs, and the organization of these constructs under higher-level topics. Each construct is captured by a variety of variables. In Section 4, we offer a comprehensive supplementary account detailing these variables and their measurements.

---

<sup>1</sup> Most of the constructs discussed here involve multiple variables. These variables have been implemented in various subsamples since the establishment of the SOEP in 1984 (see section 2). To align with our goal of highlighting essential content pertinent to researchers investigating individuals of immigrant origin, we opt not to enumerate specific variables or specify the initial subsample where the variable was introduced. This approach eliminates the necessity for an exhaustive list of references to variables and subsamples, allowing us to maintain a focused presentation of the contents. For individuals seeking detailed information about a specific variable and its measurement, please refer to section 4.

Table 2: Constructs relevant for studying individuals of immigrant origin

<b>Immigrant origin</b>	
1. Country of birth	1.1. Employment and occupation before migration
1.1. Country of birth – target person	1.2. Employment and occupation upon arrival
1.2. Country of birth – family members	1.2.1. Job offer before migration
2. Immigration	1.2.2. First employment
2.1. Immigration – target person (first gen.)	1.2.3. Consecutive employment
2.2. Immigration – family members	1.2.4. Expectations for future employment
3. Immigrant origin (generated variable)	2. Experience with German authorities (first gen.)
<b>Migration</b>	
1. Migration motives (first gen.)	2.1. Official support and consultation offers
2. Escape biography (refugees only; first gen.)	2.2. Experience and satisfaction with German authorities
2.1. Financial costs	
2.2. Escape route and transport	
2.3. Experiences during escape	
3. Moving biography (first gen.)	
3.1. Moving patterns	
3.2. Residential preferences in Germany	
3.3. Living situation and location in Germany	
4. Intentions for family reunification	
<b>Immigrant integration</b>	
<b>Cultural</b>	
1. Language	
1.1. Language proficiency	
1.1.1. Speaking, reading, and writing the language of origin (L1)	
1.1.2. Speaking, reading, and writing German (L2)	
1.1.3. Proficiency in other languages	
1.2. Language acquisition	
1.2.1. Course attendance	
1.2.2. Improvement of German language skills	
1.3. Language use	
2. Religion	
3. Other cultural aspects	
<b>Structural</b>	
1. Legal situation	
1.1. Legal status at immigration (first gen.)	
1.2. Residence permit	
1.3. Asylum procedure (first gen.)	
2. Citizenship and naturalization	
2.1. Citizenship – target person	
2.1.1. Changes in citizenship, naturalization	
2.1.2. Dual citizenship	
2.2. Citizenship – parents	
3. Education (first gen.)	
3.1. Education in Germany	
3.2. Education abroad – target person, parents	
3.3. Recognition of educational qualifications	
3.4. Educational expectations	
3.5. Preference for further education or job in Germany (Vignette)	
4. Employment situation before and after migration (first gen.)	
4.1. Time-use before migration	
4.2. Financial situation before migration	
	<b>Social</b>
	1. Social capital: Bonding and bridging ethnic ties
	1.1. Pre-migration contacts and support (first gen.)
	1.2. Social capital and finding the first job in Germany (first gen.)
	1.3. Steady relationship before and after migration (first gen.)
	1.4. Co-/inter-ethnic and inter-ethnic contacts and friendships in Germany
	1.5. Transnational ties and remittances
	1.6. Tandem “Start with a friend” (program evaluation; first gen.)
	2. Ethnic composition
	2.1. School
	2.2. Work
	2.3. Accommodation
	2.4. Circle of friends
	3. Feelings of acceptance and perceived discrimination
	3.1. School
	3.2. Work
	3.3. Accommodation
	3.4. Public services and institutions
	3.5. Everyday life
	4. Attitudes towards immigration and refugees
	4.1. Attitudes towards refugees
	4.2. Support of refugees
	4.3. Worries about immigration and anti-immigrant attitudes
	<b>Emotional, attitudinal</b>
	1. Orientations
	1.1. Identifications and feelings of belonging
	1.2. Intention to stay
	1.3. Intention to migrate
	2. Life satisfaction and worries (first gen.)
	2.1. Life satisfaction before and after migration
	2.2. Worries about staying/leaving
	3. Political orientations, attitudes, and knowledge
	3.1. Identification with a party in country of origin (first gen.)
	3.2. Perceived state of democracy in country of origin (first gen.)
	3.3. Attitudes toward democracy
	3.4. Attitudes toward gender equality
	3.5. Tolerance and liberal-minded attitudes
	3.6. Knowledge of the German political system
	3.7. Inter-ethnic trust

### 3.1 Immigrant origin

The study of immigrants and their descendants requires the identification of this population. The SOEP provides data on various relevant characteristics such as country of birth, immigration biography, legal situation, citizenship, and naturalization (Table 2; Liebau and Tucci 2015: 18-22). Researchers can use this information to generate new variables that represent an individual's immigration history, ethnic background, or other related characteristics. Additionally, scholars can opt to use several user-friendly variables created by the SOEP (Scheller 2011; SOEP Group 2021: 17-30).

GEMBORN and CORIGIN are used to provide information about a person's country of birth. GEMBORN specifies whether a person was born in Germany or in another country (SOEP Group 2021: 17-20). For individuals born in another country, CORIGIN designates the country of origin (SOEP Group 2021: 20-22). IMMIYEAR specifies the year of immigration to Germany (SOEP Group 2021: 23-27), and MIGBACK can be used to identify an individual's history of immigration. It uses information on the respondents' country of birth and (grand-)parental information such as their country of birth and citizenship (SOEP Group 2021: 28-29). The SOEP dataset uses the terms "direct migration background" and "indirect migration background" to refer to respondents' migration histories. Respondents who were born in a country other than Germany are considered to have a "direct migration background", while those born in Germany may have either "no migration background" or an "indirect migration background". Respondents with no migration background have both parents born in Germany, while those with an indirect migration background have at least one parent born outside of Germany. In terms of research on immigrants and their descendants, a "direct migration background" typically refers to the first generation, while an "indirect migration background" can designate either the second generation (both parents born abroad) or the 2.5th generation (one parent born abroad). However, the SOEP dataset's generated variable MIGBACK does not capture this differentiation.

### 3.2 Migration

The SOEP also includes information about why and how immigrants came to their destination, although these topics have received less attention compared to the contents related to immigrant integration (see Table 2). Important data collected for all first-generation immigrants over the years include migration motives and the moving biography. The moving biography covers details such as moving patterns, relocation history, residential preferences, and living situations upon arrival. With the increase in refugee migration in 2015, migration-related data collections have gained further significance, and refugees have been extensively surveyed in the IAB-BAMF-SOEP Survey of Refugees. This survey provides comprehensive information on escape biographies, routes to Germany, migration costs, and experiences.

### 3.3 Immigrant integration

Immigrant integration can be considered as the process through which newcomers to a country become part of the receiving society (Castles et al. 2002: 14-15). This process is

gradual, can vary in pace and may extend over several generations. When examining how immigrants and their children become “full members of the societies where they now live” (Alba and Foner 2015: 1), existing literature often focuses on specific areas of life that are relevant for individuals’ life chances. This suggests that integration encompasses various aspects (see e.g., Gordon 1964; Alba and Nee 1997) and can thus be understood as a multidimensional concept (Jonsson, Kalter and van Tubergen 2018: 6).

Since its inception, the SOEP has incorporated various constructs that closely align with Esser’s considerations on social integration (Esser 1980; 2000; 2001; 2004; 2006). Social integration refers to the inclusion or exclusion of individuals in an existing social system and, as a result, a more or less equal distribution of characteristics across social groups or categories of people. Esser distinguishes between several dimensions of social integration (Esser 1980: 22-23; 2000: 289; 2006: 287). The first is the *cultural* dimension, which involves the acquisition of knowledge and skills. The *structural* dimension pertains to individuals’ positioning or placement, for example, within the educational system or the labor market. The *social* dimension focuses on social relations and the establishment of connections with others within and across social groups, including social acceptance. Finally, the *emotional* dimension encompasses identifications and orientations. These four dimensions of social integration are captured in the SOEP questionnaires (Esser 2008; Jacobsen et al. 2021: 534; Liebau and Tucci 2015: 22).

Table 2 displays the multiple constructs that the SOEP covers for each dimension. A crucial element linked to the cultural dimension is proficiency in both the language of origin and the language of destination. Alongside proficiency, the data covers conditions relevant to language acquisition. The SOEP also allows for an examination of other cultural aspects such as religion, specifically in terms of denomination, religiosity, and religious practices.

The structural dimension pertains to positions and placements within society. The data in question contains information regarding immigrants’ legal circumstances, including aspects such as legal status, type of residence permit, and the asylum procedure. Other legal features covered are citizenship and naturalization. Moreover, education and employment occupy a central role within the structural dimensions. As such, the SOEP gathers data on educational degrees attained both in Germany and abroad, the recognition of educational qualifications, and educational expectations. Employment and labor market-oriented constructs encompass information on individuals’ employment and occupational situation before and after migration.

The social dimension involves social relations and the resources they provide (Jonsson et al. 2018: 7). These social ties can be established within one’s ethnic group, known as bonding ties, as well as across ethnic groups, particularly with the ethnic majority, known as bridging ties. The SOEP includes various measures related to contacts, friendships, and other aspects of social capital. It also provides information on the ethnic composition in everyday contexts such as the workplace, school, or within one’s circle of friends. Furthermore, social acceptance from the ethnic majority is a crucial aspect of the social dimension. The SOEP covers both the perceptions of immigrants and their descendants regarding the level of

acceptance and discrimination they experience, as well as the attitudes individuals hold towards immigrants and immigration.

Lastly, the emotional dimension encompasses orientations, feelings of belonging, and identifications. These aspects are measured with regard to both the society of origin and the receiving society. The SOEP also provides information on the intention to stay and on attitudes, such as attitudes towards democracy, tolerance, and gender equality.

## 4 Measurements

To illustrate the measurement of the various constructs presented in the preceding section, we provide a comprehensive table that can be accessed online via the following link: [https://www.diw.de/sixcms/detail.php?id=diw\\_01.c.879779.de](https://www.diw.de/sixcms/detail.php?id=diw_01.c.879779.de). Due to its extensive size and the amount of information covered, it is not feasible to present the table in print. However, the online format allows us to easily update and modify it with future data releases.

For illustrative purposes, we present a simplified and condensed version of the online table as Table 3. This table serves to highlight and explain its key features and contains exemplary information on the “citizenship – target person” construct, which was assigned above to the structural dimension (see Table 2). Table 3 as well as the online table include the following columns:

- *Variable name*: The name of the variable as recorded in the SOEP datasets.
- *Variable label*: The variable label as recorded in the SOEP datasets in both English and German.
- *Question text*: The text of the question that was last used and is stored on [paneldata.org](https://paneldata.org). For some questions, the text is only available in German and for others, it is only available in English.
- *Questionnaire type*: This column indicates the type of questionnaire instrument used to ask the question most recently.
- *Year*: This column indicates the most recent year in which the question was asked.
- *Dataset*: This column indicates whether the variable is recorded in the individual datasets (“pl”) or in the biographical datasets (“boil”) of the SOEP-Core.
- *Question reference*: This column provides a link where more information on the question’s history, as well as answer categories, is available.
- *Variable reference*: This column provides a link where more information on the variable is available, including the various waves for which the variable is available.
- *1984-2021*: Each of the remaining columns indicates in which year and for which subsample the variable is available. An empty cell means that the variable has not been recorded in any subsample and is thus not available for the respective year.

The original table provided online has been generated from other sources, such as [paneldata.org](https://paneldata.org)<sup>2</sup> and information stored in the individual (pl) and biographical (biol) datasets of the SOEP-Core<sup>3</sup>, rather than being manually entered. Some of the seemingly peculiar characteristics of the table are due to the information taken from [paneldata.org](https://paneldata.org), including erratic switches between English and German, incomplete or missing question texts, erroneous spelling, or missing information. The wording of the question is only reported in the way it was asked in the latest data collection. Additionally, the table does not specify answer categories and value labels. To obtain information on answer categories, value labels, and potential changes in the wording of a question and/or its answer categories over time, data users can use the links provided in the table, which refer them to additional information stored on [paneldata.org](https://paneldata.org). Researchers can also obtain more information by cross-checking the questionnaires or the datasets.

Despite these limitations, the table provides an important tool that allows data users to access detailed information on a wide range of variables relevant to studying immigrants and their offspring. For an efficient assessment of the potential of the data for addressing their research questions, researchers may be particularly interested in those parts of the compilation that contain information on the years and the subsamples in which the respective variable has been measured. The information is based on the longitudinal individual and biographical datasets of SOEP-Core.<sup>4</sup> However, it does not include information on SOEP-IS variables that are not yet integrated with SOEP-Core, as well as data from parent and child datasets and youth datasets.

---

<sup>2</sup> Question text, questionnaire type, year, question reference, variable reference.

<sup>3</sup> Variable name, variable label, dataset, 1984-2021.

<sup>4</sup> The SOEP-Core comprises various subsamples that have been consolidated into a single dataset. These subsamples encompass the SOEP-Samples A-Q, the IAB-SOEP Migration Samples, and the IAB-BAMF-SOEP Survey of Refugees.

Table 3: Excerpt from the online table on variables and measurements of relevant constructs

Variable name	Variable label German	Variable label English	Question text	Questionnaire type	Year	Dataset	Question reference	Variable reference	1984
plj0013	Staatsangehörigkeit	Nationality	Welche Staatsangehörigkeit haben Sie?	n/a	n/a	pl		<a href="https://paneldata.org/soep-core/data/pl/plj0013">https://paneldata.org/soep-core/data/pl/plj0013</a>	
plj0014_v1	Staatsangehörigkeit [1984-1993,1996-2019]	[de] Staatsangehörigkeit [1984-1993,1996-2019]		Individual (CAPI) 2019	2019	pl	<a href="https://paneldata.org/soep-core/inst/soep-core-2019-pe-lgb/178">https://paneldata.org/soep-core/inst/soep-core-2019-pe-lgb/178</a>	<a href="https://paneldata.org/soep-core/data/pl/plj0014_v1">https://paneldata.org/soep-core/data/pl/plj0014_v1</a>	A B
plj0014_v2	Staatsangehörigkeit [1984-1995]	[de] Staatsangehörigkeit [1984-1995]	Welche Staatsangehörigkeit haben Sie?	n/a	n/a	pl		<a href="https://paneldata.org/soep-core/data/pl/plj0014_v2">https://paneldata.org/soep-core/data/pl/plj0014_v2</a>	B

*Note:* This table illustrates the documentation for the first three variables of the “citizenship – target person” construct. Access the comprehensive online table at [https://www.diw.de/sixcms/detail.php?id=diw\\_01.c.879779.de](https://www.diw.de/sixcms/detail.php?id=diw_01.c.879779.de).

## 5 Working with the SOEP migration and integration data

The SOEP group offers multiple datasets that can be combined in a flexible manner to meet individual research needs. In this section, we will offer insights into the most significant datasets for migration and integration research and provide guidance on how to create a customized dataset.

The SOEP data includes wave-specific datasets for cross-sectional analysis and, since 2012, a user-friendly longitudinal dataset in “long” format that integrates all subsamples into one dataset and provides harmonized and generated variables over time. Additionally, there are several datasets that offer generated variables, spell data, and biographical data, providing insights into the respondents’ life courses before their first SOEP interview. While the cross-sectional and longitudinal datasets contain all SOEP variables collected either in the respective year or for the entire period covered by the SOEP, generated, spell, and biographical datasets focus on specific themes. An overview of all SOEP datasets, along with short descriptions, can be found in the [SOEP Companion](#).

Table 4 provides a summary of important datasets for scholars conducting research on migration and integration, although it is not intended to be exhaustive. Consistent with Jacobsen et al. (2021: 533), it categorizes datasets according to various types.

Table 4: Overview of important datasets for migration and integration research

Dataset	Type	Label	Format
Files providing the originally surveyed data for each year			
\$p	Original data	Individual questionnaire	Cross-sectional
\$h	Original data	Household questionnaire	Cross-sectional
Files providing the originally surveyed data in long format			
pl	Original data	Individual questionnaire	Long
hl	Original data	Household questionnaire	Long
Tracking files about the development of the sample			
ppathl	Tracking data	Individual Tracking file	Long
hpathl	Tracking data	Household tracking file	Long
Files providing respondents’ biographies prior to their first survey year			
biol	Original data	Biographical information	Long
bioimmig	Generated data	Biographical information	Long
Files providing additionally generated variables that are coded consistently over time			
pgen	Generated data	Individual data	Long
hgen	Generated data	Household data	Long
Files providing spell data on immigrants			
migspell	Spell data	Migration history	Spell
refugspell	Spell data	Migration history (refugees)	Spell

Notes: For brief descriptions of these datasets, please refer to the [SOEP Companion](#). Additionally, for specifications regarding the types of SOEP datasets, see Jacobsen et al. (2021: 533).

As research interests and needs are diverse, we cannot offer a one-size-fits-all approach to building a migration research dataset. Nonetheless, we would like to highlight some steps that can serve as a starting point for researchers.



To build an individual-level analysis dataset from the various SOEP datasets, it is recommended to start with the general dataset that contains all adult respondents, called `ppathl`. `ppathl` includes basic demographic variables, such as sex and year of birth, and five generated migration variables that researchers can use to restrict their sample to the desired population. The five migration variables are: `IMMIYEAR`, which represents the year in which an individual migrated to Germany; `GERMBORN`, which indicates whether the respondent was born in Germany or another country; `CORIGIN`, which identifies the country where the respondent was born; `MIGBACK`, which expresses the respondents' immigrant origin as "direct or indirect migration background", essentially indicating whether the respondent belongs to the first or second generation or has no history of immigration; and `AREFBACK`, which indicates whether respondents are refugees. For more information on these generated variables, see Scheller (2011) and SOEP Group (2021).

Once the sample population is defined based on these variables, it is possible to merge the modified `ppathl` dataset with other SOEP datasets to add selected survey information of the respondents. The variables `PID` (Person ID) and `SYEAR` (survey year) are unique identifiers in `ppathl` that can be used to merge additional SOEP datasets. The variable `HID` (Household ID) is also available in most datasets and is necessary when merging household information to the individual-level dataset.

The SOEP datasets can also be linked with other data sources, for example regional or administrative data (Jacobsen et al., 2021: 535-536). An important project, the SOEP-CMI-ADIAB, may be of particular interest to researchers in the field of migration and integration research (Antoni et al., 2023). It allows them to add administrative information from the Integrated Employment Biographies (IEB) to the IAB-SOEP Migration Samples (Brücker et al., 2014) and to the IAB-BAMF-SOEP Survey of Refugees (Keita and Trübswetter, 2020).

More information and documentation on the datasets mentioned above can be found on the [SOEP's Research Data Center](#) homepage. Researchers can access information and documentation on the latest dataset version, all SOEP Survey Papers covering data documentation, survey instruments, survey reports, variable description and coding, and further information on linking external data sources to the SOEP.

## 6 References

- Alba, R., and Foner, N. (2015): *Strangers no more. Immigration and the challenges of integration in North America and Western Europe*. Princeton and Oxford: Princeton University Press.
- Alba, R., and Nee, V. (1997): Rethinking assimilation theory for a new era of immigration. *International Migration Review* 31(4): 826-874.
- Antoni, M., Beckmannshagen, M., Grabka, M., Keita, S., and Trübswetter, P. (2023): Survey data of SOEP Core, IAB-SOEP Migration Sample, IAB-BAMF-SOEP Survey of Refugees and SOEP Innovation Sample linked to administrative data of the IAB (SOEP-CMI-ADIAB) 1975-2020. *FDZ-Datenreport*, 03/2023(en), Nuremberg: IAB.
- Brücker, H., Kroh, M., Bartsch, S., Goebel, J., Kühne, S., Liebau, E., and Schupp, J. (2014). The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. *SOEP Survey Papers 216: Series C*, Berlin: DIW/SOEP.
- Brücker, H., Rother, N., and Schupp, J. (2016). IAB-BAMF-SOEP-Befragung von Geflüchteten. Überblick und erste Ergebnisse. *Politikberatung kompakt* 116, Berlin: DIW/SOEP.
- Brücker, H., Rother, N., and Schupp, J. (2017; korrigierte Fassung 2018). IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. *Politikberatung kompakt* 123, Berlin: DIW/SOEP.
- Castles, S., Korac, M., Vasta, E., and Vertovec, S. (2002): Integration: mapping the field. *Home Office Online Report* 29(3). London: Home Office.
- Esser, H. (1980): Aspekte der Wanderungssoziologie. Assimilation und Integration von Wanderern ethnischen Gruppen und Minderheiten. Darmstadt: Luchterhand.
- Esser, Hartmut (2000): *Soziologie. Spezielle Grundlagen. Band 2: Die Konstruktion der Gesellschaft*. Frankfurt: Campus.
- Esser, H. (2001): Integration und ethnische Schichtung. *Working papers* 40. Mannheim: MZES.
- Esser, H. (2004): Does the “new” immigration require a “new” theory of intergenerational integration? *International Migration Review* 38(3): 1126-1159.
- Esser, H. (2006): Migration, language and integration. *AKI Research Review* 4, Berlin: WZB.
- Esser, H. (2008): Das SOEP und die Lindenstraße. *Vierteljahrshefte zur Wirtschaftsforschung* 77(3): 181-186.
- Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* 239(2): 345-360.

- Gordon, M. M. (1964): *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press.
- Hanefeld, U. (1987). *Das Sozioökonomische Panel. Grundlagen und Konzeption*. Frankfurt am Main: Campus.
- Jacobsen, J., Kroh, M., Kühne, S., Scheible, J. A., Siegers, R., and Siegert, M. (2019). Supplementary of the IAB-BAMF-SOEP Survey of Refugees in Germany (M5) 2017. *SOEP Survey Papers 605: Series C*, Berlin: DIW/SOEP.
- Jacobsen, J., Krieger, M., Schikora, F., and Schupp, J. (2021). Growing potentials for migration research using the German Socio-Economic Panel Study. *Journal of Economics and Statistics* 241(4): 527-549.
- Jonsson, J. O., Kalter, F., and van Tubergen, F. (2018): Studying integration: ethnic minority and majority youth in comparative perspective. In: Kalter, F., Jonsson, J. O., van Tubergen, F., and Heath, A. (eds.) *Growing Up in Diverse Societies: The Integration of Children of Immigrants in England, Germany, the Netherlands and Sweden*, pp. 3-39, Oxford: Oxford University Press.
- Keita, S., and Trübswetter, P. (2020). IAB-BAMF-SOEP Survey of Refugees linked to administrative data of the IAB. *FDZ Datenreport* 15/2020. Nürnberg: IAB.
- 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 271: Series C*, Berlin: DIW/SOEP.
- Kroh, M., Siegers, R., and Kühne, S. (2015). Gewichtung und Integration von Auffrischungstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP). In: Schupp, J., Wolf, C. (eds.) *Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen*, pp. 408-444, Wiesbaden: Springer VS.
- Kroh, M., Brücker, H., Kühne, S., Liebau, E., Schupp, J., Siegert, M., and Trübswetter, P. (2016). Das Studiendesign der IAB-BAMF-SOEP-Befragung von Geflüchteten. *SOEP Survey Papers 365: Series C*. Berlin: DIW/SOEP.
- Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., and Siegers, R. (2017). Sampling, nonresponse, and integrated weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4) – revised version. *SOEP Survey Papers 477: Series C*. Berlin: DIW/SOEP.
- Krupp, H.-J. (2008). Die Anfänge: Zur Entstehungsgeschichte des SOEP. *Vierteljahrshefte zur Wirtschaftsforschung*. 77(3): 15-26.
- Kühne, S., and Kroh, M. (2017). The 2015 IAB-SOEP migration study M2: Sampling design, nonresponse, and weighting adjustment. *SOEP Survey Papers 473: Series C*. Berlin: DIW/SOEP.
- Liebau, E., and Tucci, I. (2015). Migrations- und Integrationsforschung mit dem SOEP von 1984-2012: Erhebung, Indikatoren und Potentiale. *SOEP Survey Papers 270: Series C*, Berlin: DIW/SOEP.

- Rathje, M., and Glemser A. (2021). SOEP-Core – 2020: Report of survey methodology and fieldwork. *SOEP Survey Papers 1050: Series B*. Berlin: DIW/SOEP.
- Scheller, F. (2011). Bestimmung der Herkunftsnationen von Teilnehmer des Sozio-oekonomischen Panels (SOEP) mit Migrationshintergrund. *SOEPpapers on Multidisciplinary Panel Data Research 407*, Berlin: DIW/SOEP.
- Siegers, R., Steinhauer, H. W., and Dührsen, L. (2021). SOEP-Core v36 – Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP) (1984 until 2019). *SOEP Survey Papers 960: Series C*, Berlin: DIW/SOEP.
- Siegers, R., Steinhauer, 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 1106: Series C*, Berlin: DIW/SOEP.
- SOEP Group (2001). The German Socio-Economic Panel (GSOEP) after more than 15 years: Overview. *Vierteljahreshefte zur Wirtschaftsforschung* 70(1): 7-14.
- SOEP Group (2021). SOEP-Core v36 – PPATHL: Person-related meta-dataset. *SOEP Survey Papers 1048: Series D*, Berlin: DIW/SOEP.
- Schupp, J., and Wagner, G. G. (1995). Die Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung* 64(1): 16-25.
- Steinhauer, W., Trübswetter, P., and Zinn, S. (2022). SOEP-Core – 2020: Sampling, nonresponse and weighting in the IAB-SOEP Migration Studies M7 and M8. *SOEP Survey Papers 1105: Series C*, Berlin: DIW/SOEP.

## 7 Annex

Table 5: Number of first-generation respondents in the SOEP according to their origins (1984-2021)

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Turkey	1265	1292	1154	1086	1065	1013	968	949	915	907	869	853	801	751	707	660	855	827	732
Former Yugoslavia	235	238	239	242	246	249	253	255	256	278	285	315	318	320	327	301	349	331	305
Poland	47	47	46	48	47	49	76	79	79	81	253	395	384	356	366	344	507	489	456
Russia	3	3	3	3	3	3	8	8	8	8	103	186	190	182	204	181	327	325	306
Italy	635	648	583	527	503	477	447	437	425	399	376	363	341	321	294	276	327	304	275
Kazakhstan	0	0	0	0	0	0	0	0	0	0	68	147	151	148	182	177	322	318	304
Greece	472	484	454	404	372	329	314	308	298	296	274	262	245	223	218	198	204	187	161
Romania	11	13	13	13	13	13	19	19	19	19	77	119	114	111	109	107	165	167	153
Spain	414	421	367	315	289	266	238	209	193	177	163	142	126	107	104	86	85	82	81
Bulgaria	11	11	11	11	11	10	11	12	14	13	11	12	11	10	12	12	11	9	9
Ukraine	0	0	0	0	0	0	0	0	0	0	0	13	14	14	22	22	51	51	50
Czech	42	42	42	42	42	42	64	65	65	65	66	68	66	67	71	69	110	110	102
Austria	40	40	40	38	32	34	35	36	33	33	33	36	35	37	41	37	70	68	77
Hungary	15	15	15	15	15	15	20	20	19	20	20	21	22	22	23	22	46	46	44
Albania	5	5	5	5	5	6	6	6	6	6	15	26	27	21	24	19	28	22	21
France	18	19	18	14	10	9	8	8	11	12	17	19	19	20	19	17	39	40	41
USA	26	28	21	18	17	16	15	14	15	11	11	13	14	11	13	13	37	39	38
Netherlands	6	6	6	6	6	6	6	7	7	7	7	11	11	11	15	15	22	21	28
Kyrgyzstan	0	0	0	0	0	0	0	0	0	0	8	16	16	9	8	9	20	19	20
Morocco	0	0	0	0	0	0	0	0	0	0	1	1	1	1	3	6	18	17	6
India	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	3	3	3
Portugal	5	5	1	1	1	1	1	0	0	0	0	8	9	9	20	23	27	27	17
Brazil	2	2	3	3	3	3	3	3	3	3	4	4	4	4	5	5	14	14	10
Azerbaijan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4

Table 5: Number of first-generation respondents in the SOEP according to their origins (1984-2021) (Continued)

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
Turkey	668	623	589	572	516	460	483	657	595	540	988	904	797	663	630	642	600	544	473	29613
Former Yugoslavia	291	270	259	250	232	210	231	343	382	360	821	797	859	1222	1203	944	769	1185	1121	17091
Poland	412	380	354	363	326	296	304	432	487	507	951	898	1100	1020	941	857	794	1170	1117	16858
Russia	273	258	250	279	250	229	249	454	509	523	1109	1041	934	1079	1075	932	772	717	633	13620
Italy	253	224	205	202	180	160	175	210	198	190	394	375	419	386	328	286	264	240	214	12861
Kazakhstan	284	252	247	255	230	215	208	394	468	460	945	908	825	730	668	616	552	481	427	10982
Greece	143	124	117	116	103	94	101	103	100	97	265	258	316	284	248	209	175	141	132	8829
Romania	141	131	124	131	126	124	138	169	190	188	521	508	741	695	617	513	465	958	917	8671
Spain	65	57	54	56	50	42	54	52	56	56	121	116	159	152	117	94	86	82	78	5412
Bulgaria	7	5	5	7	7	6	8	11	23	24	47	42	301	298	259	187	168	876	876	3369
Ukraine	49	41	38	44	42	38	42	72	71	77	241	234	246	273	251	217	202	232	221	2868
Czech	94	90	82	82	78	70	73	75	84	84	94	82	87	84	79	75	68	63	59	2743
Austria	72	62	59	61	53	54	62	71	82	96	110	106	116	105	102	100	113	107	100	2426
Hungary	43	40	40	42	37	33	35	54	57	55	53	48	136	134	126	131	118	110	97	1824
Albania	21	20	20	18	15	13	13	27	31	25	40	34	34	178	202	108	87	250	245	1639
France	40	37	34	33	33	31	40	54	53	57	72	68	78	79	73	66	78	75	67	1426
USA	35	33	33	35	35	35	33	44	47	53	63	58	70	62	56	48	54	97	94	1355
Netherlands	27	26	25	26	26	26	28	42	54	53	64	58	68	64	69	66	76	69	67	1138
Kyrgyzstan	20	20	20	21	20	16	18	40	49	55	108	101	90	78	63	60	49	39	33	1025
Morocco	2	2	2	7	7	7	2	23	39	36	83	76	92	88	81	81	77	91	89	939
India	4	4	4	4	2	2	4	16	17	17	38	38	56	59	64	51	50	234	230	914
Portugal	16	13	12	13	9	9	12	23	22	23	33	31	65	59	62	49	46	41	36	729
Brazil	9	9	8	9	9	6	5	23	26	27	32	30	38	33	31	30	32	86	86	621
Azerbaijan	4	5	5	6	6	5	6	14	14	16	43	40	32	50	67	59	58	59	57	558

Notes: Own calculations based on SOEP v38. The last column ("Total") represents the number of completed interviews between 1984 and 2021.

Table 6: Number of first-generation (mostly) refugee respondents in the SOEP according to their origins (1984-2021)

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	
Syria	1	1	1	1	1	1	1	1	1	1	1	2	2	2	1	0	3	4	2	
Iraq	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	8	6	
Afghanistan	4	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	9	9	5	
Eritrea	0	0	0	0	0	0	0	0	0	0	0	7	7	8	8	7	7	5	5	
Iran	3	3	5	5	5	4	2	3	2	2	2	13	12	8	5	5	26	26	26	
Pakistan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	8	6	6	
Lebanon	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	15	15	12	
Nigeria	0	0	0	0	0	0	0	0	1	1	1	1	1	1	3	3	3	2	2	
Somalia	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	
Palestine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
Syria	1	0	0	1	2	2	1	25	27	25	80	74	165	5587	8318	8063	7274	7331	6557	43560
Iraq	7	7	7	22	20	7	13	36	42	46	142	138	173	1621	2581	2420	2113	2052	1737	13206
Afghanistan	2	3	3	3	2	2	2	17	18	21	57	52	94	1508	2223	2047	1730	1480	1199	10502
Eritrea	3	3	3	5	5	4	4	4	5	5	7	6	6	303	412	376	348	432	375	2360
Iran	26	26	27	26	23	23	25	47	46	46	66	55	79	213	372	358	314	367	330	2626
Pakistan	6	6	6	7	3	2	3	14	19	19	40	36	52	154	220	209	168	137	105	1230
Lebanon	12	10	8	8	6	4	2	25	34	29	62	52	67	99	130	129	105	111	90	1027
Nigeria	2	2	1	1	1	1	2	15	16	11	21	17	32	91	112	105	90	263	257	1059
Somalia	1	1	1	1	1	1	1	1	1	1	1	1	8	82	154	153	128	224	208	977
Palestine	0	0	0	2	2	0	0	7	6	9	18	17	14	92	114	112	107	114	113	727

*Notes:* Own calculations based on SOEP v38. The last column (“Total”) represents the number of completed interviews between 1984 and 2021. The majority of individuals listed here are refugees. However, it is important to note that the calculations are based on country of origin information rather than legal status. As a result, the table also includes individuals who did not immigrate as refugees.

Table 7: Number of second-generation respondents in the SOEP according to their origins (1984-2021)

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	
Turkey	208	223	233	251	265	279	297	312	325	328	334	336	334	329	333	325	359	351	346	
Italy	141	142	148	152	155	162	164	166	168	170	171	169	165	161	163	160	182	179	177	
Former Yugoslavia	102	104	136	112	117	121	121	123	126	125	130	126	124	124	119	117	121	120	120	
Greece	103	108	108	109	110	108	108	107	112	113	110	108	109	106	104	96	96	89	83	
Poland	4	5	9	7	8	8	17	17	20	20	26	34	37	37	38	37	60	64	70	
Spain	72	73	74	72	72	72	70	71	70	69	65	64	58	53	49	46	47	46	44	
Austria	10	11	11	11	11	11	11	12	13	13	13	13	12	12	14	15	24	23	30	
Russia	0	0	5	0	0	0	2	3	3	3	8	12	12	13	17	17	25	25	24	
USA	2	2	5	3	3	3	4	4	5	5	5	5	5	5	10	10	17	17	17	
Romania	0	0	5	0	0	0	1	1	1	1	1	2	2	2	3	3	4	4	5	
Kazakhstan	0	0	5	0	0	0	0	0	0	0	3	4	4	5	6	6	9	9	10	
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
Turkey	324	315	298	291	279	270	254	360	368	354	677	635	568	492	463	438	396	328	284	13162
Italy	169	158	154	155	144	133	130	174	180	178	285	264	232	209	181	170	151	120	104	6386
Former Yugoslavia	115	113	110	109	105	98	97	144	182	180	305	290	273	254	249	240	211	175	155	5693
Greece	75	72	69	67	62	60	56	63	70	68	156	145	135	121	114	94	77	57	55	3603
Poland	68	67	65	74	72	70	69	107	140	150	190	186	166	148	161	168	160	148	133	2860
Spain	41	39	36	32	29	27	28	39	39	34	50	48	41	36	35	34	30	27	23	1855
Austria	28	26	25	23	22	24	31	54	58	65	66	57	52	47	50	48	45	40	37	1068
Russia	24	22	20	22	22	21	21	35	43	40	59	57	53	50	49	46	46	39	36	874
USA	17	17	17	22	22	22	21	44	55	53	61	55	52	47	49	46	37	28	24	816
Romania	6	6	6	8	8	8	8	27	41	43	59	53	48	46	49	46	46	38	36	617
Kazakhstan	10	10	10	10	11	11	11	25	37	39	53	50	48	44	45	37	35	31	27	605

Notes: Own calculations based on SOEP v38. The last column ("Total") represents the number of completed interviews between 1984 and 2021.