

Brücker, Herbert et al.

**Research Report**

## The new IAB-SOEP migration sample: An introduction into the methodology and the contents

SOEP Survey Papers, No. 216

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Brücker, Herbert et al. (2014) : The new IAB-SOEP migration sample: An introduction into the methodology and the contents, SOEP Survey Papers, No. 216, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<http://hdl.handle.net/10419/103964>

**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.*

## SOEP Survey Papers

Series C - Data Documentations

SOEP – The German Socio-Economic Panel Study at DIW Berlin

216-2013

# The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents

Herbert Brücker, et al.

Running since 1984, the German Socio-Economic Panel Study (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 Documentations (Datendokumentationen)

**Series D** – Variable Descriptions and Coding

**Series E** – SOEPmonitors

**Series F** – SOEP Newsletters

**Series G** – General Issues and Teaching Materials

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

**Editors:**

Prof. Dr. Gert G. Wagner, DIW Berlin and Technische Universität Berlin

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

Please cite this paper as follows:

Herbert Brücker, Martin Kroh, Simone Bartsch, Jan Goebel, Simon Kühne, Elisabeth Liebau, Parvati Trübswetter, Ingrid Tucci, Jürgen Schupp. 2014. The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. SOEP Survey Papers 216: Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin  
SOEP  
Mohrenstr. 58  
10117 Berlin

Email: [soepapers@diw.de](mailto:soepapers@diw.de)

**HERBERT BRÜCKER<sup>1)</sup>, MARTIN KROH<sup>2)</sup>, SIMONE  
BARTSCH<sup>2)</sup>, JAN GOEBEL<sup>2)</sup>, SIMON KÜHNE<sup>2)</sup>,  
ELISABETH LIEBAU<sup>2)</sup>, PARVATI TRÜBSWETTER<sup>1)</sup>,  
INGRID TUCCI<sup>2)</sup>, JÜRGEN SCHUPP<sup>2)</sup>**

**THE NEW IAB-SOEP MIGRATION SAMPLE: AN  
INTRODUCTION INTO THE METHODOLOGY AND THE  
CONTENTS**

1) IAB – Institute for Employment Research, The Research Institute of the Federal  
Employment Agency, Nuremberg/Germany

2) DIW Berlin, The Research Infrastructure ‘Socio-Economic Panel (SOEP)’, Berlin/Germany

# The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents

Herbert Brücker, Martin Kroh,  
Simone Bartsch, Jan Goebel, Simon Kühne,  
Elisabeth Liebau, Parvati Trübswetter,  
Ingrid Tucci & Jürgen Schupp

## 1 Introduction

The IAB-SOEP Migration Sample is a joint project of the Institute for Employment Research (IAB) and the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW). The project tempts to overcome limitations of previous datasets by drawing a sample that takes into account changes in the structure of migration to Germany since 1995. The dataset opens up new perspectives for migration research and gives insights on the living situations of new immigrants to Germany. The IAB-SOEP Migration Sample is not a simple update of previous surveys. It has four key features:

First, for a subsample of the IAB-SOEP Migration Sample —and only if the respondents provide explicit consent—their individual data are linked to register data from the Integrated Employment Biographies (IEB) database of the Institute for Employment Research (IAB), which contains the entire labour market history of individuals in Germany. The linkage makes it possible to obtain precise information on wages and salaries; employment, unemployment and benefit receipt; as well as many other variables that are particularly relevant to labor market issues. Record linkage combines the advantages of both survey and register data and opens up diverse new research perspectives. This is important not only for the research on migration and the lives of immigrants but also for the SOEP study, where it is currently being implemented for the first time. To assess fluctuations in response rates and other phenomena in later survey waves, we have included only part of the respondents in the record linkage procedure. Moreover, the year in which respondents are asked for their consent to the record linkage has been randomly assigned. The methodological implications of this experiment can be therefore investigated in detail.

Second, the questionnaire used in the IAB-SOEP Migration Sample covers the complete migration, education, and labor market histories of respondents in both their country of origin and in all countries to which they have lived in. This is an important innovation over previous biographical surveys of migrants in the SOEP: It allows us to determine whether important life events occurred in the respondent's home country, in Germany, or in other countries. This highly demanding survey strategy is motivated by the idea that immigrants' biographies and their processes of integration into the host country's labor market, educational system, culture, and other aspects of society can only be understood with sufficient information on their lives in their countries of origin and on the conditions under which they immigrated. It is also informed by the recognition that migration is no longer a one-time, irreversible event in a person's life, but that individual biographies are becoming increasingly

“transnational,” often involving several migration episodes over the course of a lifetime and with ties in several different countries.

Third, based on recent advances in the research on migration and immigration, the IAB-SOEP Migration Sample questionnaire includes several new batteries of questions that have not previously been considered in the SOEP or other household surveys in Germany, or not in the necessary depth. Examples include questions on earnings and labor market integration and occupational status before migration, migration decisions in the family and partnership context, and purposes and transfer channels of remittances. The dataset will be continually developed in dialogue with the research community using the IAB and SOEP infrastructures as well as their respective user communities and advisory boards. Furthermore, a special project council consisting of academic scholars, policy makers and other social actors provides advice to the project and contributes to the further development of the questionnaire.

Finally, the IAB-SOEP Migration Sample substantially increases the sample size for research on migration and the lives of immigrants in Germany. 4,964 persons residing in 2,723 households participated in the first wave of the survey. In the SOEP, which oversampled foreigners from the former guestworker recruitment countries at the beginning of the survey (1984) and ethnic Germans (for more details on migration in the previous SOEP samples, see Section 2), the maximum number of immigrants surveyed previously was 3,369 individuals in 2000. In the 2012 wave—the last wave before the new IAB-SOEP Migration Sample went into the field—this figure declined to 1,945. To the best of our knowledge, the IAB-SOEP Migration Sample is the largest longitudinal household survey of immigrants in Germany.<sup>1</sup> Moreover, since the survey is completely harmonized with and integrated into the SOEP (as sample “M”), migrants from the other SOEP samples can be included in analyses, increasing the number of observations further.

Both partners, the IAB and the SOEP, contribute similar levels of expertise and financial resources to this joint project. The IAB’s contribution to the project’s funding is provided out of the research budget of the Federal Employment Agency (BA) and the Federal Ministry of Labor and Social Affairs (BMAS). The SOEP’s contribution comes out of the institutional funding provided to the SOEP by the German federal and state governments. Moreover, an additional grant has been provided through the Pact for Research and Innovation of the Leibniz Competition.<sup>2</sup> Both partners, IAB and SOEP, gratefully acknowledge the generous support of these donors that has enabled the creation of the IAB-SOEP Migration Sample.

The remainder of this paper describes the development and special features of the IAB-SOEP Migration Sample and gives an overview of the questionnaire. The IAB-SOEP Migration Sample builds both on previous experience surveying migrants in the SOEP as well as on past experience with migration research based on register data such as the Integrated Employment Biographies (IEB) data set. Section 2 discusses how the changing socio-economic structure of migration to Germany and the limitations of register data and previous surveys have created a growing need for a survey of this kind—in particular, one that links survey and register data. Section 3 describes the procedure used to draw the sample and the weighting procedure. To the best of our knowledge, the IAB-SOEP Migration

---

<sup>1</sup> Of course, the German Microcensus and the German Labour Force Survey have more observations than the IAB-SOEP Migration Sample, but they cover a limited number of waves per individual. They also do not contain a limited number of migration- and integration-specific indicators.

<sup>2</sup> The grant (title: “SOEP Record Linkage: Longitudinal Survey of Migrants from the Social Insurance Statistics (SOEP-REC-LINK)—Sampling of Administrative Data and Linkage with Survey Data on Migration”) is earmarked for investigating the methodological implications of linking survey and register data.

Sample constitutes the first use of IEB data as a sampling frame. This has a number of implications for the survey. Section 4 presents the content of the questionnaire used in the IAB-SOEP Migration Sample and the key similarities and differences compared to the standard SOEP questionnaire. Section 5 describes the linkage of the survey data and the IEB register data, as well as the data products that are produced. Since we cannot present all issues in detail here, we have included references to in-depth survey papers in the relevant sections.

## **2 The IAB-SOEP Migration Sample: building on register data to sample immigrants**

The IAB-SOEP Migration Sample draws on different data sources that are already available in Germany. A natural starting point for the new migration sample is the most comprehensive household survey in Germany, the longitudinal study “Living in Germany” conducted by the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin). This study has surveyed immigrants at its beginning in 1984 with an oversampling of Turkish, Greek, Italian, Spanish and (former) Yugoslavian households. This longitudinal study also oversampled ethnic Germans (“Spätaussiedler”) who arrived after the fall of the Iron Curtain. Nevertheless, despite efforts to reach immigrants in each SOEP wave (see Liebau/Tucci 2014), recent immigrant groups, particularly those from Central and Eastern Europe, are not sufficiently covered because they often create new households, which have a low probability of being included in regular SOEP refresher samples. The topics covered by the IAB-SOEP Migration Sample questionnaire are similar to those in the core SOEP and include the household and family context, socio-structural indicators (income, labour market position, etc.), social and cultural values and preferences, language skills, health outcomes and subjective indicators of well-being (Wagner et al. 2007). Furthermore, new questions that are important for research are introduced regularly and in some cases developed further.

The IAB-SOEP Migration Sample builds not only on survey data but also on the IEB register data, which contain precise information on a wide array of labour market variables on a daily basis for the universe of the population covered. Although register data are becoming more and more important for migration research, they are not collected for research purposes and therefore have a number of limitations. First, register data do not contain some indicators that might be decisive in understanding the integration process of migrants—for instance, their knowledge of German, their social contact with people of German origin, the conditions under which they came to Germany. Nevertheless, the new IAB-SOEP Migration Sample seeks to combine the advantages of both survey and register data and therefore also draws on past experience in the use of administrative data such as the IEB for migration research.

Before focusing on the data sources that underlie the new IAB-SOEP Migration Sample, we briefly describe how migration patterns have changed in Germany in order to explain the demand to improve and expand existing data sources.

### ***Changing migration patterns and their implications for the sample design***

The economic, legal, and institutional conditions for migration to Germany have changed substantially since the 1990s due to the economic and financial crisis in Europe, falling costs of travel and communication, the eastern enlargement of the EU, and various amendments to the immigration legislation. With these developments, the scale and structure of migration to Germany in recent years have changed substantially. After a decade of fairly modest net immigration at around 90,000 persons

p.a., a surge in immigration began in 2010: net migration reached 438,000 in 2013 and has been on the increase since then. Moreover, the forms of migration have also changed: more and more migrants are only moving abroad temporarily and many migrate more than once over the course of their life.

At the same time, the composition of immigration flows has changed: the EU's Eastern enlargement together, with the financial and economic crisis, triggered a diversion of European migration flows to Germany. In 2013, 66 percent of the new arrivals came from other EU Member States, 44 percent from the new Member States of Central and Eastern Europe, and 13 percent from the Southern EU Member States, which were affected particularly severely by the crisis.<sup>3</sup> Immigration from the latter two groups of countries has increased substantially in recent years, while immigration from the "traditional" countries of origin—Turkey and the former Yugoslavia—has declined to a negligible level.

The changing composition of immigration and the global trend toward increasing mobility of high-skilled workers (Boeri et al., 2012) is reflected in the skill structure of new arrivals to Germany: the share of individuals with tertiary education has increased by a factor of two from 2000 to 2010, while the share of individuals without a vocational degree has decreased substantially over the same period. Although the skill level of immigrants has declined slightly in the most recent immigration surge, it is still well above the level 10 or 15 years ago (Brenke & Neubecker 2013, Brücker 2014).

Gradually, these changing migration patterns are affecting the structure of the immigrant population in Germany overall. Although traces of "guestworker" immigration and the subsequent immigration of their family members are still evident in the social and economic structure of Germany's immigrant population, immigration from the new EU Member States of Central and Eastern Europe and the former USSR is steadily rising. The same holds true for other recent migration trends such as the increasing immigration of high-skilled individuals from the EU and other OECD countries and of individuals from Arabic countries in the wake of political crises there (although the number of immigrants from these regions is less sizeable). This has had important implications for the design of the SOEP-IAB Migration Sample: The new survey had to reflect recent changes in both the ethnic and skill composition of recent immigration to allow migration research to analyze these developments and their implications for the social and economic structure of the immigrant population in Germany.

### *Representation of migrants in previous SOEP samples and in the IAB-SOEP Migration Sample*

In 2012, before the IAB-SOEP Migration Sample was drawn, the SOEP consisted of 11 randomly selected subsamples. A range of different methods were used to sample individuals with a migration background. Subsamples can be distinguished first by whether they were designed as random samples (Samples A, C, E, F, G, H, I<sup>4</sup>, J, and K) or created specifically to sample migrants and their descendants (Samples B and D) (for a detailed description, see Liebau & Tucci 2014). In the latter samples, different sampling procedures were used for the different migrant populations. In Sample B, the information on nationality reported to the official registry office was used to sample migrants from the five largest former "guestworker" countries (Turkey, Italy, Greece, the former Yugoslavia, and Spain), which were the primary sources of immigration to the former West Germany. Sample D is

---

<sup>3</sup> Own calculations based on the migration statistics of the Federal Statistical Office of Germany (DESTATIS, 2014).

<sup>4</sup> Sample I was drawn in 2009 as a subsample of the SOEP but has now been transferred, together with the purely computer-assisted personal interviews (CAPI) conducted by interviewers in subsample E, into the independent SOEP Innovation Sample (SOEP-IS) and is no longer part of the core SOEP survey (see Richter & Schupp 2012).

designed to cover all immigration starting in 1984, which consisted primarily of ethnic German immigrants, asylum seekers from across the world (but mainly from the war-torn regions of the former Yugoslavia), as well as family reunification (immigration of family members of immigrants from “guest worker” countries who had already settled in Germany). Since ethnic Germans also were granted German citizenship upon arrival in Germany, in their case it was impossible to use nationality data from the official registry offices for purposes of sampling. Instead, screening interviews were carried out for Sample D, augmented initially by a snowball procedure (Schupp & Wagner 1995).

The remaining subsamples contain different numbers of immigrants depending on the respective target population and any additional measures taken to ensure proportional coverage of the immigrant population.<sup>5</sup> Subsamples A, C, E, G, H, and K are either representative of the total population or focus on other subpopulations (e.g., East Germans in case of Sample C and high-income earners in the case of Sample G). The proportion of foreigners or immigrants and thus the number of immigrant households in these samples is therefore relatively low. In reaction to this, measures were taken in subsamples F, I, and J to ensure a representative cross-section of the net sample through slight oversampling of the immigrant population. In Sample F, the random route walk taken in the pre-selection of addresses was extended (on the random walk procedure, see Thompson 2006) and in the extended part, screening interviews were carried out to determine the nationalities of household members. The foreign households identified in the extended part of the random walk procedure were then also added to the gross sample instead of adding additional German households. In order to carry out the slight oversampling of individuals with a migration background that was also planned for subsamples I and J, an onomastic (name recognition) procedure was used to identify individuals with possible migration backgrounds in the process of an extended address listing. This procedure is used to identify individuals that may have a migration background based on their first and last names (see Humpert and Schneiderheinze 2013, and Section 3 of this paper).

Despite the sampling approaches used in past years to obtain precise coverage of people with a migration background, the number of immigrants surveyed in the SOEP has been in continual decline due both to return migration to home countries and due to refusals to participate over the course of the panel study (see Figure 1).<sup>6</sup>

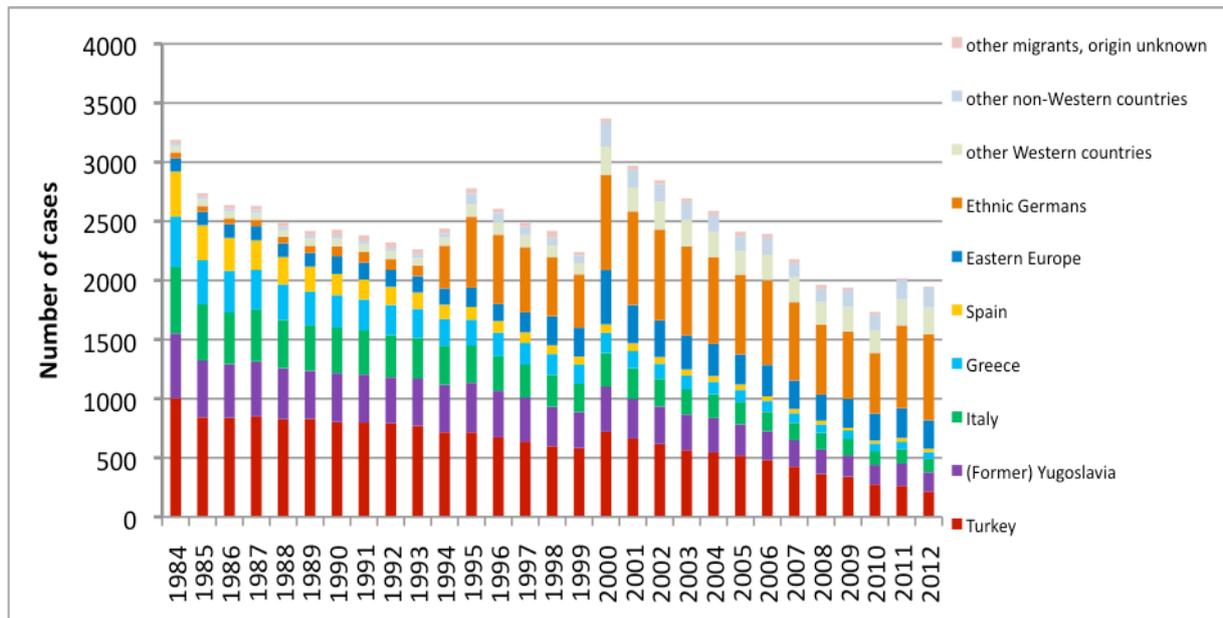
The decline in the number of respondents with a migration background in the SOEP has made it increasingly difficult to use the SOEP data to analyze significant changes in Germany’s migrant population. Particularly underrepresented groups in the SOEP are the aforementioned newer groups and waves of immigration from Central and Eastern Europe and the southern EU countries in crisis. One goal of the IAB-SOEP Migration Sample was therefore to take these changes in the immigrant population adequately into account and to improve coverage of more recent immigrant cohorts in the SOEP.

---

<sup>5</sup> Immigrants are people who were born in another country and moved to Germany. These individuals have a migration background that includes a personal experience of migration, whereas there are also people with a migration background who do *not* have personal experience of migration (i.e., individuals whose parents immigrated). The concept of “migration background” therefore includes both immigrants and their descendants.

<sup>6</sup> In addition, new immigrant groups such as the ethnic Germans of the 1990s are only being covered by new samples and sampling methods. As clearly evident in Figure 1, the addition of a new immigration sample (Sample D) in 1994-1995 was accompanied by an increase in the share of individuals with a migration background and among them, the groups of ethnic Germans in particular. Thereafter, the share declined gradually over time. Not included, however, are the data collected since 2011 as part of the Familien in Deutschland (FiD, Families in Germany) survey, which covers around 4,500 families with children in the household (see Schröder et al. 2013). In the FiD wave 2013, around 2,500 adults respondents have a migration background and almost 3000 children.

**Figure 1: Case numbers of selected immigrant groups, SOEP Samples A-K – Respondents aged 16 and over**



Source: SOEP A-K, own calculations.

The sample was selected, based on an analysis of immigration figures, to overrepresent certain immigrant groups. In the selection of these groups, various criteria were used to ensure that an adequate case number would be available for analysis of the countries that are becoming increasingly important sources of immigration. The countries were identified based on both long-term migration trends since 1995 as well as more recent migration trends. The latter procedure was chosen in order to consider groups of countries in the sample that could become more important for immigration to Germany in the future. These groups include Poland and Romania, the two new EU member states that are the most important sources of migration to Germany, and the successor states of the former Soviet Union, including ethnic Germans, who come mainly from this group of countries. Because of the economic and financial crisis, immigration from the southern European EU member states of Greece, Italy, Spain, and Portugal has increased significantly in recent years. Because new immigrants differ starkly in both economic and social terms from those who came to Germany during the period of guest worker recruitment, the new immigrants from this group of countries are overrepresented in the sample. Finally, it is likely that the political upheavals in Arab and predominantly Islamic countries will cause immigration from this region to increase in the future, and there has been acute interest in the public, as well as from quantitative migration researchers, in understanding this specific form of immigration better. The sample was therefore selected to also overrepresent immigrants from Arab and Islamic countries. Finally, to ensure that the SOEP will have adequate numbers of cases from the “traditional” source countries of immigration in the future, the IAB-SOEP Migration Sample has an overproportional representation of these immigrants’ groups. This is also motivated by the fact that many descendants of the so-called guestworkers have meanwhile entered the labor market and dominate the so-called second generation in Germany today.

### *Register data for migration research*

Administrative register data are taking an increasingly important place alongside longitudinal survey data such as the SOEP in the research on migration and immigration, as well as in other areas of labor

market research. In Germany, one such dataset, the Integrated Employment Biographies (IEB), provides comprehensive information on employment biographies, earnings, unemployment, benefit receipt, and participation in active labor market policies (see below). In general, register data such as the IEB have the advantage of providing precise information on these issues on a daily basis. Moreover, since they cover the universe of the respective population smaller changes in the social and economic structure can be better tracked than with survey data. They also entail disadvantages, however: register data are not collected for research purposes and therefore provide only a limited number of variables. The most important disadvantage for the research on migration and the lives of immigrants in Germany is that the IEB does not offer information on the migration background of individuals and on their date of arrival in Germany.<sup>7</sup> Other important information is missing as well: whether educational degrees were obtained in Germany or abroad, whether degrees or qualifications earned abroad have been recognized in Germany, and information on language proficiency and family and partnership contexts, to name a few.

Nevertheless, register data are being used increasingly in migration research due to the large number of observations and the precision of the information—for instance, to measure the wage and employment effects of immigration (Brücker & Jahn 2011, Brücker et al., 2014; D’Amuri et al., 2010; Glitz, 2012), to investigate wage discrimination against migrants in a setting with monopsonistic competition (Hirsch & Jahn, 2014), and to assess wage convergence between immigrants and natives (Lehmer & Ludsteck, 2014). These studies usually attempt to approximate the migration status by using information from register data on the present and previous nationality of individuals or on the participation of ethnic Germans in specific programs (e.g., Brücker & Jahn 2011). Analogously, the arrival date is approximated by the first observation of labour market participants in the IEB (Lehmer & Ludsteck, 2014). Nevertheless, the limited information provided by datasets such as the IEB severely limits the potential for migration research based on register data in Germany. In light of this, the IAB-SOEP Migration Sample has been developed using the approach of linking survey and register data—of course, under strict adherence to German data protection provisions.

### 3 Sampling and weighting

In the previous section, we described the efforts taken to represent different target groups of the overall immigrant population in Germany in the SOEP-IAB Migration Sample. A variety of obstacles and demands for a possible sampling frame had to be taken into account in creating the sample. Germany lacks a centralized register of persons with a migration background. Although there are municipal registry offices (*Melderegister*) that generally include information on nationality in their records, it is essentially impossible to use this information in sampling naturalized persons with a migration background or specific immigrant cohorts. Alternative sampling strategies, which have been used in the SOEP in the past, either use a large number of screening interviews, for instance, (Sample D and F), or onomastic procedures, for instance, using address information (Sample I and J).

---

<sup>7</sup> While the official statistics of the Federal Employment Agency (BA) include the migration background of unemployed persons for statistical purposes, German legislation does not allow the use of this information for research purposes by using the data. Only aggregate data are available. Moreover, data on the migration background of employees are not collected by the Federal Employment Agency, only information on nationality is available.

The IAB-SOEP Migration Sample is to our knowledge the first sample of persons with migration background in Germany using the Integrated Employment Biographies (IEB) as a sampling frame. The following sections provide a cursory overview of the sampling procedure, non-response analysis, and construction of the weighting variables. Detailed information can be obtained from Kroh et al. (2014).

### *The Integrated Employment Biographies (IEB) as a sampling frame for migrant populations*

As mentioned before, the IEB is a database provided by the IAB. The IEB covers employees, unemployed persons, job seekers, recipients of mean-tested benefits (unemployment benefit II) and participants in active labor market programs on a daily basis from 1975 onward. Data are collected on employees in Germany who are subject to social security contributions, which describes almost all private sector employment. Public sector employees are only covered if they are obliged to pay social security contributions; Civil servants who are not covered by the social security insurance system (so-called “Beamte”) are not. Employers are requested to submit information on starting and ending dates of all their employees’ job spells as well as total earnings received (censored at the maximum taxable earnings level) on an annual basis. All changes in earnings and other aspects of labor contracts are reported within a year as well. In addition, the establishment identification number and some job characteristics are recorded. In total, the IEB contains 83,521,672 individuals with 1,894,018,836 spells. Furthermore, information on unemployment spells, benefit receipt, participation in active labor market policies, and job-search status are directly matched from the different sources of the social security system to form a complete picture of the labour market history of individuals.

Although the IEB does not collect information that allows migrants or migrants’ descendants to be identified precisely at a given point of time, its longitudinal structure makes it possible to identify whether or not a person ever had non-German nationality. Moreover, the IEB records whether individuals participated in specific programs provided by the Federal Employment Agency such as language classes that are designed for migrants.<sup>8</sup> Also, family names are used in onomastic procedures to further identify persons with a possible migration background. Hence, the IEB allows identification of migrants much better than more common sampling frames such as the phone directory and data stored in local registry offices.

Three additional advantages come with the IEB as a sampling frame. First, a practical advantage is that the IEB is a centralized sampling frame whereas register offices in Germany work at the local level and national sampling requires collaboration with each of the sampled municipalities. Second, the wealth of information on the labor market participation of individuals, their wages, as well as information on their employers enables researchers to model non-response processes more fully than is possible with many alternative sampling frames. The (model-based) weighting of the data obtained in the IAB-SOEP Migration Sample—for instance, on wages—thus corrects for any deviation in registered wages in the IEB between the gross and the net sample. Third, the IEB sampling frame allows linkage of survey and register data in subsequent research projects.

---

<sup>8</sup> Those programs have been aimed particularly at ethnic Germans (so-called “Spätaussiedler”). These programs are provided to the overwhelming majority of this group, which numbers several million individuals, most of who immigrated in the 1990s.

The IEB has some disadvantages as well. Although the database represents a great share of the target population, some groups are not covered (on “undercoverage”, see Jacobebbinghaus & Seth 2007). In particular, civil servants who are not obliged to pay social security contributions and self-employed people who have never held a job that is subject to social security contributions and have never received unemployment benefits or attended an active labor market policy measure are not covered in the IEB. An estimation based on the SOEP and the German Microcensus has shown that by choosing the IEB as a sampling frame, 5 to 8 percent of the target population is excluded. However, the excluded groups, such as the self-employed, students, and refugees, may enter the survey as household members of anchor persons (see below), but possibly less than proportionally.

### *Overview of the sampling procedure*

In a first step, we aggregated the spell information from the IEB at the level of individuals and restricted the population to individuals who first appeared in the IEB after 1994. The database comprised 17.4 million individual records at this stage, not excluding any person without (presumed) migration background. To make the subsequent fieldwork (face-to-face interviews) easier, we then clustered available address information into groups of about 2,500 persons to create 6,725 geographically distinct sample points (primary sampling units, PSUs) based on a clustering of geographically proximate addresses.<sup>9</sup>

The second step in the sampling process was to identify target population members. Individuals who at some point in time had been reported as having foreign, i.e. non-German citizenship, as well as those who had taken part in measures of the Federal Employment Agency specifically designed for persons with a migration background (e.g., language classes) were defined as persons with a migration background. This definition applied to about 4.1 million individuals. However, the identification does not cover individuals who became German citizens before their first entry into the IEB. This applies in particular to naturalized second-generation migrants. To assure that these subgroups had a chance to be sampled as well, we made use of the onomastic procedure in which names are classified by country of origin (Humpert & Schneiderheinze 2013). First and last names of individuals without an unequivocal migration background need to be preprocessed and then compared to large databases containing lists of names specific to country and ethnic origin. German nationals were assigned to a country of origin based on a probabilistic matching procedure. For financial reasons, the onomastic procedure was only conducted on a subsample of 600 out of the total of 6,725 PSUs. PSU selection was based on stratified random sampling by federal states (Länder) and region type. The number of PSUs to be sampled in each stratum was proportional to the number of individuals with an unequivocal migration background in a given stratum. The 600 sampled PSUs comprise about 1.5 million individuals whose names were analyzed in the onomastic process. After the onomastic procedure, we removed all individuals from the data who had always been German nationals, never participated in an immigration measure, and who had “typical” German names.

---

<sup>9</sup> Clustering based on geocoded addresses. Geocoding and clustering was conducted by the team of the German Record Linkage Center ([www.record-linkage.de](http://www.record-linkage.de), Schnell 2013). Note that clustering operated at the county level, i.e., primary sampling units do not cover persons from more than one county.

In the next step, 250 of the 600 sample points were sampled for fieldwork. Sampling probabilities depend on the number of migrants identified using both the information provided by the IEB and onomastic analysis. Afterwards, secondary sampling units were sampled from the 250 PSUs. The gross sample comprised 80 addresses (households) from each of the selected PSUs, i.e., 20,000 records in total. A distance-based and entirely simulated random walk procedure was implemented in the sampling process. Furthermore, the algorithm was based on a disproportional sampling scheme, which assigned higher sampling probabilities to the aforementioned migrant groups (see Section 2). This approach ensures pre-defined minimum sample sizes for each of the migrant groups, although some population subgroups are rather small. For further details on the complex sampling design of the migrant sample, see Kroh et al (2014).

**Table 1. Composition of population and samples by nationality/country of origin and generation**

Nationality/ Country of Origin	IEB since 1995		250 PSUs		Gross Sample		Net Sample	
	N	%	n	%	n	%	n	%
Italy (All)	168,733	4.6	10,961	4.3	1,764	8.8	202	7.4
1			2,909	1.3	868	4.3	94	3.5
2			6,669	2.6	683	3.4	91	3.3
n.d.			1,353	0.5	213	1.1	17	0.6
Spain & Greece (All)	119,365	3.3	7,236	3.0	1,892	9.5	229	8.4
1			2,660	1.1	929	4.6	108	4.0
2			3,428	1.4	706	3.5	87	3.2
n.d.			1,148	0.5	257	1.3	34	1.4
Turkey (All)	860,442	23.5	53,176	21.1	3,038	15.2	381	14.0
1			17,425	6.9	1,587	7.9	169	6.2
2			28,957	11.5	1,094	5.5	163	6.0
n.d.			6,794	2.7	357	1.8	49	1.8
Former Yugoslavia (All)	351,338	9.6	18,718	7.4	1,828	9.1	262	9.6
1			8,127	3.2	858	4.3	121	4.4
2			8,394	3.3	742	3.7	99	3.6
n.d.			2,197	0.9	228	1.1	42	2.6
Ethnic Germans	-	-	31,252	12.4	2,123	10.6	347	12.7
Poland	290,891	7.9	16,375	6.5	2,108	10.5	269	9.9
Romania	104,190	2.8	4,480	1.8	1,309	6.5	189	6.9
CIS	448,140	12.2	18,276	7.2	1,977	9.9	367	13.5
Muslim and Arabic countries	209,667	5.7	22,597	9.0	1,238	6.2	167	6.1
Rest of World	1,115,076	30.4	69,547	27.5	2,723	13.6	310	11.4
Subtotal	3,666,852	100.0	-	-	-	-	-	-
Germany	13,479,090	-	-	-	-	-	-	-
Total sample	17,145,942	-	252,618	100.0	20,000	100.0	2,723	100.0

Table 1 displays the composition of the IEB population (column 2) as well as the composition of the gross and net sample (columns 3 and 4) across sampling groups. As can be seen from the comparison between the 252,618 individuals in the 250 sample points and the 20,000 individuals sampled for fieldwork, selected migrant groups were to be sampled with a disproportional, i.e., higher or lower sample probability. For instance, we aimed for much higher percentage of migrants from Romania, although Romanian migrants have represented just 1.8 percent of the IEB database since 1995.

A short screening interview was conducted with each cooperating anchor person in order to identify eligible individuals and therefore households. Around thirty percent of all households were screened out, as anchor persons turned out not to be part of the target population. In more than half of the cases, screen-out was due to immigration before 1995 and in about one-third of the cases, the screening interview classified anchor persons as natives.

## Weighting

While the construction and the sampling of the 250 (geographical) primary sampling units results in a self-weighting scheme, i.e., all sampling units in the target population have the same sampling probability, the selection of secondary sampling units, i.e., the anchor persons from these PSUs, clearly follows a disproportional design. The differences in sampling probabilities are captured in the design weight of the sample.

Moreover, sampled anchor persons may refuse to participate. Based on regional information as well as interviewer records of the sampled addresses and—most importantly—based on the register information of the IEB, we analyze differences in response propensities across groups of the sample (for complete documentation of the non-response analysis, see Kroh et al. 2014). The product of different selection probabilities by country of origin as well as the different response probabilities by characteristics of anchor persons constitutes the raw weight of the IAB-SOEP Migration Sample. Using totals on the target population based on the German Microcensus, we finally poststratify the raw weight to obtain a weighting variable that replicates known margins of the target population in terms of year of immigration, regional distribution, country of origin, etc.

## 4 Questionnaire

The questionnaire of the IAB-SOEP Migration Sample has two main aims: First, it attempts to provide a comprehensive picture of the social, economic, and cultural aspects of immigrants' lives in Germany as well as their complete biographies in their home countries, in Germany, and in any other countries where they lived for at least three months. In doing so, it also considers determinants of migration and many other aspects of migration that are relevant for the frontiers of research on immigration and the lives of immigrants. Second, since the sample is also constructed as a subsample of the SOEP, the migration survey is harmonized with the standard SOEP survey, and the questionnaire therefore builds on the regular SOEP questionnaire. Note that many topics relevant for the research on migration and the lives of immigrants are already part of the regular SOEP. In the remainder of this section, we briefly describe the parts of the standard SOEP program that are also covered by the IAB-SOEP Migration Sample questionnaire, and then outline the novel aspects of the new questionnaire.

### *Overview on the standard SOEP survey program*

The following are among the core thematic areas of the standard SOEP survey program. They also reflect the diverse interests and perspectives of social scientific research in economics, sociology, psychology, and many other disciplines. In our view, these topics are not only of general interest but also highly relevant for the research on migration and the lives of immigrants and their descendants:

- *Household composition:* gender, age structure, births, deaths, marital status, family structure, position within the household.
- *Socio-economic structure:* earning status, professional position, prestige scales of occupations.
- *Labor market and employment conditions:* labor market participation, occupational mobility, firm characteristics, work quality, occupational qualification requirements, job search, unemployment, reservation wages.

- *Educational characteristics*: highest formal educational attainment and occupational qualifications, current enrolment in educational institutions, attainment of educational degrees and qualifications, further education, and parental education.
- *Income types and levels, household transfers, social security*.
- *Time and activity budgets as well as do-it-yourself work within private households*: average time-use indicators.
- *Living*: life satisfaction, standard of living, quality of life, expenditures, geographic mobility.
- *Health*: self-reported indicators on health conditions.
- *Preferences, political and cultural values*: Risk attitudes, political and cultural values, political participation.

Also, since the beginning, the SOEP has sought to do justice to the multidimensional living situations of immigrants and descendants of immigrants by including questions on different aspects of their lives and their integration into the labor market, educational system, society, and social networks in the regular SOEP survey program and replicated at regular intervals. The SOEP contains numerous indicators that are relevant to the analysis of migration and the integration of immigrants and their descendants, including indicators of proficiency in German and in the language of the country of origin (cognitive dimension of integration), on acquisition of German citizenship (structural dimension), on contact with people of German origin (social dimension), and on identification with Germany and the country of origin (identificative dimension). A table with the indicators and the respective surveying methodology can be found in Liebau & Tucci (2014).

### *Migration-specific modules used in the IAB-SOEP questionnaire*

The survey program of the IAB-SOEP Migration Sample follows the standard structure of the SOEP questionnaires, the first wave of which usually contains a biographical, an individual, and a household questionnaire.<sup>10</sup> For the IAB-SOEP Migration Sample, as for the extension samples of 2011 and 2012, an integrated individual biographical questionnaire was conceptualized to avoid duplication and to improve the flow of questions in the interview. In order to adapt the questionnaire more closely to the target group and in consideration of the interests of researchers in immigrants as a group, several migration-specific modules and questions were developed and are being used in this subsample of the SOEP for the first time. These are:

- Migration biography
- Recognition of foreign educational degrees
- Inclusion of last job prior to migration and first job after migration in the occupational biography
- Situation in partnership at the point in time of migration
- Religion and religious practice

In addition, in this sample, the following sets of questions from the standard SOEP survey are asked, although some sets of questions have been modified or expanded<sup>11</sup>:

- Current language proficiency (also SOEP 2013), prior to migration (Sample M only)

---

<sup>10</sup> Additional instruments like the mother-child questionnaires, the youth questionnaire, and the questionnaire “The deceased respondent” were used starting with the second wave in 2014.

<sup>11</sup> In parentheses is information on whether the questions were also asked in the same year in the other samples of the main SOEP survey.

- Experience of discrimination (also SOEP 2013), by area (Sample M only)
- Contacts between immigrants and Germans (SOEP 2013 also)
- Intentions to stay (SOEP 2013 also)
- Visits to country of origin (Sample M only)
- Identification with Germany, with country of origin, and with Europe (Europe: Sample M only)

In order to be able to ask these additional questions, part of the questionnaire had to be left out that was given to respondents in the other SOEP subsamples in 2013.<sup>12</sup> In the following, the five new modules are described in more detail.

### Migration biography

Up to now, the SOEP only asked when the respondent (most recently) moved to Germany. There is little information on previous episodes of migration: on the one hand, the question repeated at two-year intervals of whether the individual had visited his or her country of origin in the last two years, and on the other hand, the question so previous periods of living abroad for more than three months, which was asked in 2009 but can only give a vague idea of whether further episodes of migration occurred<sup>13</sup>. By creating a gapless migration biography covering all stays of three or more months between the respondent's first move from his or her country of origin to his or her move to Germany, the SOEP now provides a data base for the analysis of migration processes for the first time in the SOEP context. Migrants' increasingly complex stories have become a significant area of research in recent years surrounding diverse issues such as transnationalism. The SOEP data provide the respondent's migration status and the precise start and end month of each migration episode for up to 15 episodes,<sup>14</sup> each lasting three months or more. In addition, for the respondent's last move to Germany, the respondent is also asked to state what languages he or she spoke before moving to Germany and about any support provided by friends and relatives. The second generation and all other individuals who were born in Germany were also asked whether they spent any periods of more than three months abroad, and these stays abroad are also documented precisely to the month.<sup>15</sup>

### Recognition of foreign educational qualifications

An important factor affecting immigrants' integration into working life is the question of whether any educational degrees they attained before coming to Germany are recognized here. Up to now in the SOEP, the following questions have been asked: "Did you receive a degree or qualification upon successful completion of this education or training?" If the respondent answers yes, they are asked: "Is this qualification recognized in Germany?" Answer options: yes/no. It was distinguished whether the degree was recognized as completely or only partially equivalent to a German degree. The passage of new legislation on April 1, 2012, that changed the criteria for recognition of foreign educational qualifications and the increasing public interest in the acknowledgment of foreign credentials made it necessary to collect additional information for analysis. Further questions were added on the process of obtaining recognition (when and where the respondent applied, when he or she received notification) and on the outcome of the recognition procedure, with more detailed questions about

---

<sup>12</sup> The complete questionnaires can be found under:

[http://www.diw.de/de/diw\\_02.c.222729.de/instrumente\\_feldarbeit.html](http://www.diw.de/de/diw_02.c.222729.de/instrumente_feldarbeit.html).

<sup>13</sup> The question was formulated as follows: "Have you ever lived in another country for more than three months, whether for professional or personal reasons?" Answer options: yes, within the last 10 years; yes, but more than 10 years ago; no.

<sup>14</sup> This is the technically defined upper limit, although it was assured that the last move to Germany was included.

<sup>15</sup> The migration biography information can be found in a separately generated dataset (MIGSPELL).

whether support was provided for the respondent to obtain further qualifications and whether he or she had already completed the respective training program. In addition, individuals who did not apply for recognition of foreign qualifications were asked to state their reasons for not doing so.

#### **Additional information on last job before, first job in Germany and current job situation**

In the SOEP, as part of the collection of biographical information, respondents are asked to provide information on their occupational status, the sector they work in, and the type of work they did in their first job. In the migration sample, a different approach was used with immigrants to Germany. The questions used were not about the respondent's first job in his or her working life, but about his or her last job before moving to Germany and first job in Germany. Because of the already considerable length of the questionnaire, only the following questions were asked:

Last job before moving to Germany:

- Occupational status (in seven categories, in simpler terms than in the SOEP)
- Monthly net income
- Working hours
- Subjective assessment of changes in respondent's occupational situation before and after moving to Germany

First job in Germany<sup>16</sup>:

- How did the respondent find out about the position?
- Does he/she still do the same job / have the same occupational status / have the same employer today as in his/her first job in Germany?
- Have there been job changes since then?
- Subjective assessment of changes in the respondent's occupational situation since moving to Germany

Individuals who were born in Germany were asked to provide the same information as is collected for respondents' first job in Germany, and in addition, to provide information on their occupational status (in seven categories)<sup>17</sup> in their first job.

#### **Relationship with partner since moving to Germany**

An important new field of immigration research deals with the question of what influence relationships have on the decision to immigrate or on possible positive or negative impacts of migration on the individual's occupational situation. In order to explore the latter question, respondents were asked whether they were in a relationship at the time when they moved to Germany and whether they maintained this relationship; which of the two partners moved to Germany first; and who found a job first.

#### **Importance of religion and religious practice**

For the migration sample, a question about religious affiliation was asked first and then broken down into additional subcategories to better reflect the religious diversity of immigrants to Germany. Then all respondents—both those with and those without any religious affiliation—were asked about the importance of their faith, how frequently they pray, and whether they attend religious events and services.

---

<sup>16</sup> In 2014, the second wave of the migration survey included questions on the following information about the respondent's first job in Germany: type of work, educational requirements, public service, economic sector.

<sup>17</sup> The missing information in comparison to the SOEP was collected later, in 2014.

### **Translation assistance**

For the migration sample, a total of five questionnaire translations are provided. Since the beginning of the SOEP survey in 1984, the questionnaires have been translated into different languages (see Liebau & Tucci 2014). Since 2011, translations have been provided in English, Russian, and Turkish. For the migration sample, translations into Polish and Romanian are provided as well—two languages for which no translations had ever been provided previously in the history of the SOEP. The interviewer was also given the option of taking an interpreter along for the household interviews. Overall, the translations and interpreting services were used to a limited extent: only 242 respondents (out of N=4.964), that is, 5%, made use of translation assistance.

**Table 2: Use of translation by language**

Language version	N	%
English	12	5
Turkish	57	24
Russian	111	46
Romanian	33	146
Polish	29	12
Total	242	100

A short film introducing the SOEP to potential respondents, in which longtime SOEP respondents talk about their experience and reasons for participating in the study, has also been translated into the aforementioned languages and made available on the website [www.leben-in-deutschland.info](http://www.leben-in-deutschland.info).

The questionnaire also included the question of whether the respondent would consent to the linkage of their survey data with information from the IEB. According to German data protection law, this consent has to be given in writing, and the respective question is asked at the end of the questionnaire. In the next section, we describe this procedure and the results of the linkage in more detail.

## **5 Linkage with the IEB and data distribution**

As mentioned above, the survey data on the portion of respondents who gave consent are linked with the register data from the Integrated Employment Biographies (IEB) sample. Through this linkage, the survey data are enhanced by the addition of detailed information on the labor market trajectories of immigrants and their descendants. The resulting data set allows for longitudinal analysis starting with the very first wave. At the end of the survey, respondents are given an informational brochure that explains the German data protection regulations and asks for their consent to data linkage. Then they are asked to give their first and last name and in addition their birth name to ameliorate the probability to find the person in the IEB. In order to examine the possible effects of such a formal request for consent on subsequent willingness to participate in the survey, and to answer the question of whether the point at which consent was requested (first or second wave) affects the consent rate, only two-thirds of the respondents were initially asked for consent. In the subsequent two waves, consent will be asked from the other subgroups. Asking for respondents to sign a consent form usually leads to a lower rate of consent than just asking an oral consent due to the formal character of the request, and in the first wave, just 51 percent of all respondents provided consent. With respect to the sample as a whole, this implies a consent rate of 32 percent. This rate is expected to increase over the years because the participants' trust will increase with their participation in the survey.

**Table 3: Record Linkage Rates**

<b>Persons</b>	<b>Linkage consent</b>	<b>No linkage consent</b>	<b>Not asked</b>	<b>All</b>
<b>Anchor</b>	890	912	921	2723
<b>Family</b>	714	761	766	2241
<b>Total</b>	1604	1673	1687	4964

Source: IAB-SOEP Migration Sample, DOI:10.5684/soep.iab-soep-mig.2013

Note: Twelve anchor persons and nine family members provided consent to record linkage and are counted among “linkage consent” although they were not supposed to be asked according to the experiment design.

The 890 anchor persons selected from the IEB who consented to data linkage can be directly linked to the IEB using an existing indicator. Linking the data on the 714 household members is more difficult, since they have to be located in the IEB database based on their names, birthdates, and addresses. Therefore, the names, birthdates, and addresses from the relevant years and counties are taken from the data warehouse on which the IEB is based. In order to minimize the size of the dataset, only addresses recorded from 1995 onwards were included. Furthermore, addresses were only taken from counties in which the respondents’ address was located (147 counties). This dataset as well as the information given by respondents was harmonized with the aid of preprocessing codes provided by the Record Linkage Center at IAB (e.g., writing all words in capital letters, standardization of abbreviations, removal of hyphens, etc.) to make them comparable. The process data were also deduplicated to erase double entries that arise because of multiple entries for individuals over time. In a second step, the data of the survey were merged with the data of the IEB using Stata, allowing identically written names/birthdate/address-combinations to be identified. 42 percent of the household members have already been identified in the IEB in this way. The same “merging” process was carried out again without the zip code and/or street address, since these two variables have a relatively high number of missings in the IEB. Another 28 percent of participants were identified in the IEB in this way. In the last step, a probabilistic merging procedure was carried out with the record linkage software Merge Tool Box (MTB) (Schnell et al. 2004) that is also able to recognize slightly different spellings of names, which occur frequently in the case of foreign names. With 26 percent of household members who were found using the MTB, another 27 individuals, or 4 percent of household members, remain without linkage.

**Table 4: Results of the record linkage between non-anchor persons in the IAB-SOEP Migration Sample and the IEB**

	<b>Identified persons</b>	<b>in percent of 714 non-anchor persons with consent</b>
<b>Direct merge</b>	302	42 %
<b>Merge without postal code</b>	199	28 %
<b>Merge without house number</b>	4	0.5 %
<b>Probabilistic Merge (MTB)</b>	182	25%
<b>Rest</b>	27	4.0 %

Source: IAB-SOEP Migration Sample, DOI:10.5684/soep.iab-soep-mig.2013, own calculations.

### *Data distribution*

The survey data from the IAB-SOEP Migration Sample are collected using questionnaires that are largely harmonized to the SOEP questionnaires administered for sample A-K. The survey will be integrated as “Sample M” into SOEP–Core and therefore will be covered by the same user agreements

on data protection as all other SOEP samples. The survey data will therefore be distributed to all SOEP users automatically as of SOEP Version 30, released in fall 2014.<sup>18</sup>

Nevertheless, differences in the survey program exist between the IAB-SOEP Migration Sample Sample M and the existing SOEP Samples A-K. To make it easier to use the data, but also to achieve the greatest possible transparency regarding the different versions of the survey program, the following approach is pursued: First, all cases in the existing files will be integrated, Second, an additional generated dataset on migration biographies (MIGSPELL) will be created. Third, the users will also be provided with the original dataset to allow them to better identify differences in the questionnaires used. Finally, whenever possible, all relevant variables generated by the SOEP-Team for the distribution of the SOEP-data are also provided to users for the IAB-SOEP Migration Sample. This pool of data on the IAB-SOEP Migration Sample will be available as a scientific use file from the Research Data Center SOEP and the Research Data Center (FDZ) at IAB (<http://fdz.iab.de>). Documentation of the new sample will be published on the Internet (along with descriptive data documentation) under: <http://data.soep.de> and <http://fdz.iab.de>.

At the beginning of 2015, the IAB-SOEP Migration Sample available at the Research Data Center (FDZ) at IAB will in addition also include the linked information of the register based IEB dataset. The data will be weakly anonymized and available for analysis in the framework of a guest visit to the Research Data Center of IAB in Nuremberg or other external locations (see [http://fdz.iab.de/en/FDZ\\_Scope\\_of\\_Services.aspx](http://fdz.iab.de/en/FDZ_Scope_of_Services.aspx) for a list of locations). The dataset will contain several individual and labour market indicators for the survey participants like:

- **Information on the individual:** sex, year of birth, citizenship, marital status, number of children, education and training, schooling, professional qualifications.
- **Information on employment, government benefit receipt, and job search:** start and end of employment, government benefit receipt, and job search, daily pay rate, type of work carried out in (last) job, professional status and working hours, employment status, employment status before job search, type of benefits received, form of job termination, remaining period of unemployment benefit receipt.
- **Employer characteristics:** economic sector, employer's entry into database, last entry on employer in database, total number of employees, number of full-time employees, number of staff in marginal employment.
- **Regional information:** place of residence (federal state, county), place of work (federal state, county).

Furthermore, IAB and SOEP work together on an anonymized dataset with a limited number of variables, which can be distributed as a scientific use file to the research community by mid 2015.

---

<sup>18</sup> The weighting of the SOEP will also include the new sample on a standard basis.

## Bibliography

Brenke, Karl; Neubecker, Nina (2013): Struktur der Zuwanderungen verändert sich deutlich. *DIW-Wochenbericht*, Vol. 80, No. 49, pp. 3-21.

Boeri, Tito; Brücker, Herbert; Docquier, Frédéric; Rapoport, Hillel (eds.) (2012): *Brain drain and brain gain: the global competition to attract high-skilled migrants*. Reports for the Fondazione Rodolfo DeBenedetti, Oxford: Oxford University Press.

Brücker, Herbert; Hauptmann, Andreas; Jahn, Elke J.; Upward, Richard (2014): Migration and imperfect labor markets: theory and cross-country evidence from Denmark, Germany and the UK. In: *European Economic Review*, Vol. 66, No. 1, pp. 205–225.

Brücker, Herbert (2013): *Auswirkungen der Einwanderung auf Arbeitsmarkt und Sozialstaat: Neue Erkenntnisse und Schlussfolgerungen für die Einwanderungspolitik*. Gütersloh: Bertelsmann Stiftung.

Brücker, Herbert; Jahn, Elke J. (2011): Migration and wage-setting: reassessing the labor market effects of migration. In: *The Scandinavian Journal of Economics*, Vol. 113, No. 2, pp. 286-317.

D'Amuri, Francesco; Ottaviano, Gianmarco I.P.; Peri, Giovanni (2010): The labor market impact of immigration in Western Germany in the 1990s. In: *European Economic Review*, Vol. 54, No. 4, pp. 550-570.

DESTATIS (2014): *Vorläufige Wanderungsstatistik 2013*, Fachserie 18, Wiesbaden: Statistisches Bundesamt.

Glitz, Albrecht (2012): The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany, In: *Journal of Labor Economics*, Vol. 30, No. 1, pp. 175-213.

Hirsch, Boris; Jahn, Elke (2014): Is there monopsonistic discrimination against immigrants? First evidence from linked employer-employee data (forthcoming). In: *International Labor Relations Review*.

Humpert, Andreas; Schneiderheinze, Klaus (2013): Linguistic analysis of personal names based on the “Onomastic process”. [http://www.stichproben.de/herunterladen/onomastic\\_process\\_HS\\_GbR.pdf](http://www.stichproben.de/herunterladen/onomastic_process_HS_GbR.pdf).

Lehmer, Florian; Ludsteck, Johannes (2014): Wage assimilation of foreigners: which factors close the gap? Evidence from Germany. In: *Review of Income and Wealth*, Online First.

Liebau, Elisabeth; Tucci, Ingrid (2014): Migrations- und Integrationsforschung mit dem SOEP von 1984 bis 2012: Erhebung, Indikatoren und Potentiale. *SOEP Survey Paper* (im Erscheinen).

Jacobebbinghaus, Peter; Seth, Stefan (2007): The German Integrated Employment Biographies Sample IEBS. In: *Schmollers Jahrbuch*, Vol. 127, No. 2, pp. 335-342.

Kroh, Martin; Goebel, Jan; Kühne, Simon; Preu, Friederike (2014): The 2013 IAB-SOEP Migration Sample (M): Sampling Design and Weighting Adjustment. *SOEP Survey Paper* (im Erscheinen).

Richter, David; Schupp, Jürgen (2012): SOEP Innovation Panel (SOEP-IS) - Description, Structure and Documentation. *SOEPpapers on Multidisciplinary Panel Data Research No. 463*.

Schnell, Rainer (2013): Linking Surveys and Administrative Data. *German RLC Working Paper No. wp-grlc-2013-03*.

- Schnell, Rainer; Bachteler, Tobias; Bender, Stefan (2004): A Toolbox for record linkage. In: *Austrian Journal of Statistics*, Vol. 33, No. 1-2, pp. 125-133.
- Schröder, Mathis; Siegers Rainer; Spieß, C. Katharina (2013): "Familien in Deutschland" - FiD - Enhancing Research on Families in Germany. *SOEPpapers on Multidisciplinary Panel Data Research No. 556*.
- Schupp, Jürgen; Wagner, Gert G. (1995): Die Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung*, Vol. 64, No. 1, pp. 16-25.
- Thompson, Steven K. (2006): Targeted Random Walk Designs. In: *Survey Methodology*, Vol. 32, No. 1, pp.11-24.
- Wagner, Gert G.; Frick, Joachim R.; Schupp, Jürgen (2007): The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. *Schmollers Jahrbuch*, Vol. 127, No. 1, pp.139-169.