

Huber, Simon; Glemser, Axel

**Research Report**

## SOEP-Core - 2016: Report of survey methodology and fieldwork for M3 and M4

SOEP Survey Papers, No. 1049

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Huber, Simon; Glemser, Axel (2021) : SOEP-Core - 2016: Report of survey methodology and fieldwork for M3 and M4, SOEP Survey Papers, No. 1049, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/242105>

**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/>

# 1049<sup>2021</sup>

**SOEP** Survey Papers  
Series B - Survey Reports (Methodenberichte)

## SOEP-Core – 2016: Report of Survey Methodology and Fieldwork for M3 and M4

Simon Huber, Axel Glemser (Kantar Public)

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

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

The SOEP Survey Papers is comprised of the following series:

- Series A** – Survey Instruments (Erhebungsinstrumente)
- Series B** – Survey Reports (Methodenberichte)
- Series C** – Data Documentation (Datendokumentationen)
- Series D** – Variable Descriptions and Coding
- Series E** – SOEPmonitors
- Series F** – SOEP Newsletters
- Series G** – General Issues and Teaching Materials

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

#### Editors:

Dr. Jan Goebel, DIW Berlin  
Prof. Dr. Stefan Liebig, DIW Berlin and Freie Universität Berlin  
Prof. Dr. David Richter, DIW Berlin and Freie Universität 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:

Simon Huber, Axel Glemser (Kantar Public). 2021. SOEP-Core – 2016: Report of Survey Methodology and Fieldwork for M3 and M4. SOEP Survey Papers 1049: Series B. Berlin: DIW/SOEP.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2021 by SOEP

ISSN: 2193-5580 (online)

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

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

# SOEP-Core – 2016: Report of Survey Methodology and Fieldwork for M3 and M4

**Simon Huber, Axel Glemser (Kantar Public)**

*Munich, 2021*

# **KANTAR PUBLIC**

## **SOEP 2016**

### **Wave Report+**

#### **Samples M3/4**

To

DIW Berlin  
Mohrenstraße 58  
10177 Berlin

Submitted by:  
Simon Huber

Kantar GmbH  
Kantar Public Deutschland  
Landsberger Str. 284  
80687 München

Your contact persons:

Simon Huber  
Research Director  
Tel: 089 5600 - 1414  
E-Mail: [simon.huber@kantar.com](mailto:simon.huber@kantar.com)

Axel Glemser  
Senior Director  
Tel: 089 5600 - 1486  
E-Mail: [axel.glemser@kantar.com](mailto:axel.glemser@kantar.com)

Munich, 28 June 2021  
315110938, 315112121

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background Samples M3/4</b>	<b>4</b>
<b>3</b>	<b>Structure of the Gross Sample</b>	<b>6</b>
<b>4</b>	<b>Questionnaires and fieldwork material</b>	<b>8</b>
<b>4.1</b>	<b>Questionnaires</b>	<b>8</b>
<b>4.2</b>	<b>Fieldwork material</b>	<b>8</b>
<b>5</b>	<b>Conducting the survey</b>	<b>11</b>
<b>5.1</b>	<b>Survey mode</b>	<b>11</b>
<b>5.2</b>	<b>Fieldwork timings</b>	<b>11</b>
<b>5.3</b>	<b>Translations</b>	<b>12</b>
<b>5.4</b>	<b>Panel Maintenance</b>	<b>14</b>
<b>5.5</b>	<b>Movers and Tracing</b>	<b>15</b>
<b>5.6</b>	<b>Interviewer Characteristics and Training</b>	<b>15</b>
<b>6</b>	<b>Fieldwork results</b>	<b>17</b>
<b>6.1</b>	<b>Participation on household level</b>	<b>17</b>
<b>6.2</b>	<b>Participation on individual level</b>	<b>20</b>
<b>6.3</b>	<b>Participation by types of questionnaires</b>	<b>21</b>
<b>6.4</b>	<b>Interview length per questionnaire</b>	<b>21</b>
<b>6.5</b>	<b>Consent to Record Linkage</b>	<b>22</b>
<b>6.6</b>	<b>(App-)Follow Up</b>	<b>22</b>
<b>7</b>	<b>Data preparation</b>	<b>24</b>
<b>8</b>	<b>Delivered data</b>	<b>26</b>

## 1 Introduction

**Table 1** gives an overview of the main characteristics of the first wave of the refugee boost samples M3 and M4 of 2016. While a total of 14,913 addresses were provided from the “Ausländerzentralregister” (AZR), the gross sample of M3 consisted of 4,051 households, 1,769 of whom participated in the survey between June and October 2016. This results in an adjusted response rate of 61.8 percent. From 10,850 addresses provided for M4, the gross sample consisted of 3,688 households, 1,769 of whom participated in the survey between August and December 2016. This results in an adjusted response rate of 62.2 percent. As is usually the case for boost and refreshment samples in the SOEP, only questionnaires were fielded in CAPI mode. No other modes were used in both samples. Partial unit non-response (PUNR) was at 53.7 percent (M3) and 57.3 (M4) percent respectively.

Table 1: **Summary fieldwork**

	Sample	
	M3	M4
Fieldwork period	June - October	August - December
Mode (main questionnaires)	CAPI	CAPI
Gross sample (hh)	4,051	3.688
Net sample (hh)	1,769	1,769
Response rate (adjusted; hh) <sup>1</sup>	61.8	62,2
Number of questionnaires	Adults: 2 Youths: 0 Children: 0	Adults: 2 Youths: 0 Children: 0
Net sample (individuals)	Adults: 2,351 Youths: 0 Children: 0	Adults: 2,466 Youths: 0 Children: 0
Questionnaire length (median, in minutes)	Household: 18 Adult: 84	Household: 20 Adult: 76
Partial unit non-response (PUNR)	53.7	57.3

<sup>1</sup> RR = percentage of all households with at least one household and individual interview in the gross sample (gross sample adjusted for households where the last person is deceased or the household moved abroad, is permanently untraceable or dissolved households where the last member moved into another SOEP household).

## 2 Background Samples M3/4

In order to implement an innovative sampling procedure for mapping recent migration and integration dynamics, the SOEP at DIW Berlin, the Institute of Employment Research (IAB Nuremberg), and the Research Centre of the Federal Office for Migration and Refugees (BAMF-FZ) formed a research partnership. This alliance also facilitated drawing samples for research from the Central Register of Foreign Nationals (Ausländerzentralregister, AZR).

M3 is the acronym for the first top-up sample of households that represents adult refugees who entered Germany from January 1, 2013 until January 31, 2016 and applied for asylum in Germany. The sample consists of two tranches. The second tranche was necessary because in the second half of 2015, so many refugees entered Germany that this led to a gap between application for asylum and registration in the Central Register of Foreign Nationals.

M4 is the acronym for the second new refugee top-up sample. It consists of two tranches as well. The first one is a household boost of the M3 sample. For the second tranche, underage children of refugee families were sampled as key informants, but only the adults in their respective households were asked to participate.

The sampling frame of the Central Register of Foreign Nationals provides only basic information about foreigners in Germany, including: name, date of birth, and a registration number linked to the local recordkeeping authority. Thus, the BAMF-FZ was in charge of contacting those local recordkeeping authorities to obtain actual addresses of the refugees. As experts in the SOEP group at DIW Berlin conducted the drawing of the gross samples, we will provide some general information on the sampling procedure. A stratified multistage approach was used to draw the gross sample.

- Each available dataset was flagged to indicate membership in the target group of refugees entering Germany since January 1<sup>st</sup>, 2013 until January 31, 2016 who applied for asylum according to the information of the register.
- All datasets were linked to the local recordkeeping authority. They were the primary sampling units (PSU) in accordance with strata based on the information of the Central Register of Foreign Nationals. Local recordkeeping authorities with smaller number of refugees have been integrated into synthetic PSUs.
- The sampling of 130 PSU, stratification by federal state and administrative district are based on the Central Register of Foreign Nationals.
- The gross overall sampling of eligible registration numbers to be supplied with addresses by the local recordkeeping authority included: 80 addresses per PSU for M3 T1, 40 addresses per PSU for M3 T2, and 45 addresses per PSU for M4 T1 and T2 in each tranche.
- This procedure should have led to n=27,300 addresses in the overall sample. Due to a lack of cooperation by local recordkeeping authorities (in time), refugees leaving the their local recordkeeping authorities' designated area before registering their address, and unaccompanied minor refugees excluded by the local data-keeping authorities for M4 T2, this resulted in a total of 25,763 addresses in the gross samples, M3 and M4.



# KANTAR PUBLIC

- BAMF-FZ provided Kantar Public with these addresses. In order to conduct a sufficient number of interviews, Kantar Public drew a gross sample for fieldwork: 24 addresses per PSU4 in M3 T1, 6 addresses per PSU in M3 T2, 11 addresses per PSU in M4 T1, and 17 addresses per PSU in M4 T2.
- This procedure resulted in 7,739 addresses in gross samples M3 and M4 for the fieldwork.

## 3 Structure of the Gross Sample

The gross sample for the fieldwork of M3/M4 consisted of 7,739 households (4,051 in M3 and 3,688 in M4). **Table 2** shows the distributions of the gross samples by federal state; **Table 3** shows the distributions of these samples with respect to spatial interlocking (BIK types) and community type. With regard to the gross proportion of refugee households, one should bear in mind that refugees are distributed among and within federal states by an official allocation procedure (Königsteiner Schlüssel). Therefore, most of these households were in the states of North Rhine-Westphalia (19.1 percent), Bavaria (12.9 percent), Baden-Wuerttemberg (11.6 percent) and Lower Saxony (9.2 percent) and predominantly in the center of bigger cities.

Table 2: Household characteristics I

Household Level	M3		M4		Total	
	Abs. <sup>1</sup>	In %	Abs. <sup>2</sup>	In %	Abs.	In %
<b>Gross Sample</b>	<b>4,051</b>	<b>100.0</b>	<b>3,688</b>	<b>100.0</b>	<b>7,739</b>	<b>100.0</b>
<b>State</b>						
Schleswig-Holstein	185	4.6	141	3.8	326	4.2
Hamburg	135	3.3	112	3.0	247	3.2
Lower Saxony	356	8.8	358	9.7	714	9.2
Bremen	116	2.9	28	0.8	144	1.9
North Rhine-Westphalia	741	18.3	734	19.9	1475	19.1
Hesse	214	5.3	221	6.0	435	5.6
Rhineland Palatinate	109	2.7	144	3.9	253	3.3
Baden-Wuerttemberg	462	11.4	435	11.8	897	11.6
Bavaria	528	13.0	470	12.7	998	12.9
Saarland	159	3.9	86	2.3	245	3.2
Berlin	297	7.3	279	7.6	576	7.4
Brandenburg	220	5.4	164	4.4	384	5.0
Mecklenburg Western Pomerania	89	2.2	83	2.3	172	2.2
Saxony	87	2.1	226	6.1	313	4.0
Saxony-Anhalt	110	2.7	84	2.3	194	2.5
Thuringia	168	4.1	58	1.6	226	2.9

<sup>1</sup> For 73 households that moved abroad this information is unavailable (1.8% of gross sample of M3); <sup>2</sup> For 65 households that moved abroad this information is unavailable (1.8% of gross sample of M4).

Table 3: Household characteristics II

Household Level	M3		M4		Total	
	Abs. <sup>3</sup>	In %	Abs. <sup>4</sup>	In %	Abs.	In %
<b>Gross Sample</b>	<b>4,051</b>	<b>100.0</b>	<b>3,688</b>	<b>100.0</b>	<b>7,739</b>	<b>100.0</b>
BIK-Type <sup>1</sup>						
0	1,086	26.8	972	26.4	2058	26.6
1	134	3.3	413	11.2	547	7.1
2	826	20.8	684	18.5	1510	19.5
3	451	11.3	459	12.4	910	11.8
4	70	1.8	102	2.8	172	2.2
5	507	12.7	207	5.6	714	9.2
6	493	12.4	387	10.5	880	11.4
7	294	7.4	287	7.8	581	7.5
8	70	1.8	83	2.3	153	2.0
9	47	1.2	29	0.8	76	1.0
Community type <sup>2</sup>						
1	121	3.0	72	2.0	193	2.5
2	205	5.1	260	7.0	465	6.0
3	892	22.0	880	23.9	1772	22.9
4	830	20.5	787	21.3	1617	20.9
5	530	13.1	348	9.4	878	11.3
6	666	16.4	614	16.6	1280	16.5
7	734	18.1	662	18.0	1396	18.0

<sup>1</sup> BIK type: 0 (more than 500,000 inhabitants/center) 1 (more than 500,000 inh./periphery), 2 (100,000 to 499,999 inh./center), 3 (100,000 to 499,999 inh./periphery), 4 (50,000 to 99,999 inh.(center), 5 (50,000 to 99,999 inh./periphery), 6 (20,000 to 49,999 inh.), 7 (5,000 to 19,999 inh.), 8 (2,000 to 4,999 inh.), 9 (fewer than 2,000 inh.)

<sup>2</sup> Community type: 1 (fewer than 2000 inhabitants), 2 (2,000 to 5,000 inh.), 3 (5,000 to 20,000 inh.), 4 (20,000 to 50,000 inh.), 5 (50,000 to 100,000 inh.), 6 (100,000 to. 500,000 inh.), 7 (more than 500,000 inh.)

<sup>3</sup> For 73 households that moved abroad this information is unavailable (1.8% of gross sample of M3).

<sup>4</sup> For 65 households that moved abroad this information is unavailable (1.8% of gross sample of M4).

## 4 Questionnaires and fieldwork material

### 4.1 Questionnaires

In the first wave of M3 and M4 two questionnaires were fielded: the individual questionnaire for first time respondents (including additional biographical questions) for all adult household members and the household questionnaire for the key informant. A special SOEP individual and life-history questionnaire was developed that includes issues specific to refugees. As is the usual approach for boost samples, no youth or child questionnaires were fielded in both samples. All questionnaires were solely available in CAPI mode and provided in six different language versions (see **Section 5.3**). In addition to the questionnaires for respondents, interviewers were asked to complete a short questionnaire about the area the household is located in, the so-called “residential environment questionnaire”.

Table 4: **Questionnaires and modes**

	CAPI
Household questionnaire	✓
Individual questionnaire + life history questionnaire module	✓

As with every other subsample of the migration population in the SOEP (M1 and M2) established previously, there was a clear need for several deviations from SOEP standard questionnaires in order to reflect the special characteristics of the target group. Several additional questions concerning migration and integration were thus integrated into the individual questionnaire to better field the range of research questions and research goals of the cooperating partners involved. This included topics such as: heritage, (experiences on) the way to Germany, language skills, integration classes in Germany, job experience, current occupation, educational background, health, attitudes, and values. The household questionnaire was much more SOEP-related in order to establish longitudinal information on the households.

### 4.2 Fieldwork material

In addition to the questionnaires, a whole range of fieldwork materials such as letters, leaflets, and documents for the interviewers were designed, printed, and sent to households and interviewers. **Table 5** provides an overview of the different material types that are prepared in samples M3 and M4. All materials for respondents are provided in seven different languages (German, English, Arabic, Farsi, Pashtu, Urdu, and Kurmanji).

## **Advance letter**

About two weeks before the start of the fieldwork period, households received an advance letter in which general information on the survey was given and the interviewer's visit was announced. It was always sent in German language as well as in one second of the six other available languages based on the likelihood of the individual speaking a certain language. The letter included links to the SOEP website that provides additional information in the different languages. Due to the subsample of M4 (tranche 2), in which underage children of refugee families were sampled, but only the adults in their respective households were asked to participate, there were two different versions of the advance letter for M3 and parts of M4.

## **Brochure**

Every household in samples M3 and M4 received a four-page brochure with further information on the survey. The brochure in German and the second language was sent with the advance letter. As well as with the advance letter two different versions were provided according to the target population.

## **Declaration on data protection**

Every household gets a two-page declaration on data protection detailing the organizations that are responsible for processing all respondent data along with a description of data handling and data recipients. The declaration on data protection in German and the second language was sent with the advance letter.

## **Consent to record linkage form**

To all respondents a consent to record linkage form was presented that allows an individual linkage between the respondent's data and employment history data available at the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB Nuremberg).

## **Address form and household grid**

The address form provides an overview of the household composition as it was last known to Kantar, in the case of samples M3 and M4 the form contained the key informant only (respectively the key child in tranche two of M4), due to the household composition being unknown in wave one. The interviewers listed every household member and documented when and with whom the survey was conducted or why a sample member did not participate in the current year. They also noted every single contact attempt made. For both samples, interviewers did this electronically in the "Mein Kantar" software. Moreover, the interviewers documented any moves of households and household members.

## **Project instruction book**

In addition to a shorter project description, interviewers in samples M3 and M4 also received a much more detailed instruction manual that is about 60 pages long. This manual contained information on special features of the current wave, specific processing instructions, and questionnaires as well as background information on the project. For tranche two of M4 a slightly different version was provided in order to prepare the interviewers for the different target populations.

## **Further interviewer materials**

Moreover, the interviewers received contact cards for households that could not be reached at home and household information cards with information on key informants, their names, years of birth, types of questionnaires, and notes. Because the language barriers in households of refugee samples are often higher than in other samples, interviewers received a few additional laminated sheets in all seven languages to help explain how a survey is carried out, to address frequently asked questions and to help fill out the household grid. For movers, interviewers also leave a postcard with the households that asks them to send their new address to Kantar.

## Additional interviewer material to use when dealing with shared accommodations

In samples M3 and M4, interviewers might have to deal with employees and security personnel at shared refugee accommodations. For this purpose, the Federal Office for Migration and Refugees (Bundesamt für Migration und Flüchtlinge, BAMF) sent the interviewers a letter and a leaflet describing the survey to be handed over to shared accommodations. Furthermore, a leaflet describing the survey was provided for the accommodations.

Table 5: Fieldwork material

Fieldwork material	Specifics
Advance letter	<ul style="list-style-type: none"> <li>■ For respondents: In 7 languages (German, English, Arabic, Farsi, Pashtu, Urdu, Kurmanji)</li> <li>■ For the management of shared accommodations</li> <li>■ For municipal government officials</li> </ul>
Brochure	<ul style="list-style-type: none"> <li>■ In 7 languages</li> </ul>
Declaration on data protection	<ul style="list-style-type: none"> <li>■ In 7 languages</li> </ul>
Consent to record linkage form <sup>1</sup>	<ul style="list-style-type: none"> <li>■ In 7 languages</li> </ul>
Address form and household grid <sup>2</sup>	<ul style="list-style-type: none"> <li>■ Electronic form ("Mein Kantar")</li> </ul>
Project instruction book	<ul style="list-style-type: none"> <li>■ Versions for M3 and M4</li> </ul>
Other interviewer material	<ul style="list-style-type: none"> <li>■ Project description</li> <li>■ Contact card (in 7 languages)</li> <li>■ HH information card</li> </ul>
Additional interviewer material to use in the households	<p>In 7 languages:</p> <ul style="list-style-type: none"> <li>■ How is a survey carried out?</li> <li>■ FAQ</li> <li>■ Aid to fill out the household grid</li> <li>■ Postcard for movers</li> </ul>
Additional interviewer material to use when dealing with shared accommodations	<ul style="list-style-type: none"> <li>■ Letter from the Federal Office for Migration and Refugees (BAMF)</li> <li>■ Leaflet describing the survey</li> </ul>

<sup>1</sup> Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB Nuremberg): Link to employment history data

<sup>2</sup> Including the so-called "B3 form" used to process address changes.

## 5 Conducting the survey

### 5.1 Survey mode

Both refugee samples were solely conducted face-to-face per CAPI mode.

### 5.2 Fieldwork timings

Face-to-face interviewing for M3 started in mid-June, and fieldwork was scheduled to end in September. Interviewers were allowed to continue interviewing through the first weekend of October such that the final, most delayed interviews were conducted in October. Fieldwork for M4 started in mid-August. While almost two-thirds of the interviews were completed after two-and-a-half months for M3, M4 had a slower start. The focus was on finishing M3 with a sufficient amount of interviews within this overlapping fieldwork period. After finishing M3, the focus obviously shifted to M4, enabling interviewers to concentrate on this project only. This resulted in great impetus to finish interviewing for M4 by the end of December. **Table 6** shows the progress of the fieldwork for the whole period.

Table 6: **Monthly fieldwork progress M3 and M4**

	M3		M4	
	Gross Sample in %	Net Sample in %	Gross Sample in %	Net Sample in %
June	6,2%	6,2%		
July	28,6%	31,5%		
August	57,9%	63,2%	4,1%	4,8%
September	97,5%	98,1%	11,1%	11,6%
Oktober	100,0%	100,0%	29,2%	30,7%
November			66,0%	68,6%
December			100,0%	100,0%

## 5.3 Translations

Language problems during the interviewing process form a potential major challenge for surveys with populations that recently entered the country as refugees. Although some of the deployed interviewers in M3 and M4 speak Arabic, Farsi, or Pashtu, it is generally not feasible to match interviewers with special language skills with respondents in such a large nationwide survey, particularly if in advance there is no knowledge about the respective target person's background and skills. The solution was to use an innovative bilingual CAPI-program. Instead of using translated questionnaires on paper, which served as a reference for the interviewers and the interviewees in recent M1 and M2 samples, the translation was scripted into the CAPI such that German and another language were shown on the screen at the same time. The language was selected at the beginning of the interview. There was no way to switch to another language during the interview, which did not turn out to be a major issue. Moreover, a foreign-language hotline was set up to help interviewer and interviewee arrive at an agreement on the language for their upcoming interview and to help with all other issues and concerns regarding its nature and scope. If interviewee could not read and write in their respective language well enough, audio files were available on each screen for interviewees to listen to the questions and answers in their respective language. As this procedure was quite complex and costly, the number of languages offered to the interviewees was limited to six: English, Arabic, Farsi, Urdu, Pashto and Kurmanji. All written materials and fieldwork documents for the households were translated into these six foreign languages as well. **Figure 1 and 2** show examples of this approach.

Table 7: Available language versions

	German / English	German / Arabic	German / Farsi	German / Pashto	German / Urdu	German / Kurmanji
Household questionnaire	✓	✓	✓	✓	✓	✓
Individual + life history questionnaire	✓	✓	✓	✓	✓	✓

As expected, Arabic was used most frequently in almost two-thirds of all interviews (M3: 63.1 percent, M4: 64.8 percent). The results of using the different language versions are shown in **Table 8**. The language versions of Pashto, Urdu and Kurmanji were only used for a very limited number of interviews. In contrast, English (M3: 18.5 percent, M4: 15.4 percent) and Farsi (M3: 11.1 percent, M4: 12.9 percent) were used with the second highest regularity; with English as the number one choice when none of the provided languages were the interviewees' native language.

**Table 9** shows the extent to which translations were used in interviewing. The English translation, often called upon even when the interview was conducted in German, was as expected not used or hardly used in the majority of the interviews. However, at least one fifth of the interviews required the English translation for each question. For 64 percent of the interviews in the Arabic version, which as mentioned above was used very frequently, the translation was also needed for every question. At 9 percent, the percentage of cases in which the translation was requested, but not required, is comparatively low. A very similar distribution is shown in the Farsi and Urdu versions. In Pashto and Kurmanji, a somewhat different picture emerges; here the response category "not at all" is much higher.



Figure 1: Screenshot bilingual CAPI program – language selection

SP1

Welche Sprachversion soll verwendet werden?

Which language combination should be used for the interview?

ما هو الزوج اللغوي الذي تريد استخدامه لإجراء الحوار؟  
 مايل هستيد كه مصاحبه به چه زبانی انجام بگیرد؟  
 كومه ژبه بايد د مركی لپاره وکارول شي؟  
 انٹرویو كے لئے كون سی زبان كا امتزاج استعمال كيا جانا چاہئے؟

Kijan ziman bikar bê?

1. Deutsch / Englisch  
 1 German / English

2. Deutsch / Arabisch  
 2 ألمانی / عربي

3. Deutsch / Farsi  
 3 ألمانی \ فارسی

4. Deutsch / Paschtu  
 4 جرمنی/پښتو

5. Deutsch / Urdu  
 5 جرمن / اردو

6. Deutsch / Kurmandschi  
 6 Almanî/Kurmancî

Figure 2: Screenshot bilingual CAPI program with audio files

Q64

Wie zufrieden sind Sie in Ihrer gegenwärtigen Wohnsituation mit der Sicherheit in der Nachbarschaft?

Antworten Sie bitte wieder anhand einer Skala.  
 Der Wert 0 bedeutet ganz und gar unzufrieden.  
 Der Wert 10 bedeutet ganz und gar zufrieden.  
 Mit den Werten dazwischen können Sie Ihre Einschätzung abstimmen.

بائنسبە لأوضاع حياتك الحالية إلى أي مدى أنت راض عن سلامة جيرانك؟  
 يرجى ذكر إجاباتك باستخدام التدرج الموضح.  
 قيمة 0 تعني: غير راض تماماً.  
 قيمة 10 تعني: راض تماماً.  
 يمكنك استخدام القيم البينية لتوضيح إجاباتك.

ganz und gar unzufrieden      ganz und gar zufrieden      Keine Angabe

0   1   2   3   4   5   6   7   8   9   10

Text vorlesen      Instruktion vorlesen      Antworten vorlesen

© 2016 THG InHastest

Table 8: Utilization of a certain language version – individual questionnaire

Individual level	M3		M4	
	Abs.	In %	Abs.	In %
<b>Total</b>	<b>2,351</b>	<b>100.0</b>	<b>2.466</b>	<b>100.0</b>
German / English	436	18.5	380	15.4
German / Arabic	1,483	63.1	1.598	64.8
German / Farsi	260	11.1	317	12.9
German / Pashto	39	1.7	49	2.0
German / Urdu	55	2.3	22	0.9
German / Kurmanji	78	3.3	100	4.1

Table 9: Intensity of use of a certain language version M3 plus M4 in percent

Individual level	Englisch	Arabic	Farsi	Pashto	Urdu	Kurmanji
Every question	19.8	63,7	58,0	31,4	54,2	36,5
More than two thirds	7.0	12,0	13,7	10,5	15,3	11,9
More than half	5.9	8,0	8,9	7,0	13,9	7,5
Less than half	9.8	6,9	6,6	9,3	8,3	13,8
Not at all	57.5	9,3	12,8	41,9	8,3	30,2

In contrast to the translations, the audio files were used more rarely and called upon rather selectively in interviews. With longer texts and more demanding questions, especially if the target person had a certain weakness in reading, audio files were helpful in the interview situation. Overall the audio files proved helpful in all language versions.

In addition to the questionnaires and the fieldwork materials that were available in different languages, the interviewers could call a so-called “interpreter hotline” during the process of contacting the households. Then other interviewers that are fluent in either Arabic or Farsi helped the interviewer, e.g., to explain the study background, answer questions and set up an appointment for the interview. But this service could only be used to convince respondents to participate in the study. The interview itself needed to be conducted with the bilingual questionnaires.

## 5.4 Panel Maintenance

In samples M3 and M4, households did not receive any cash incentives or vouchers. As refugees are recipients of state benefits, cash incentives would be obliged to register and might lead to reduction of benefits. Further, qualitative surveys prior to the main study showed that cash incentives might be problematic because most

refugees lived within shared accommodation, which could cause enviousness and conflict between the inhabitants. Instead, interviewers were advised to bring a small gift such as local specialties.

## 5.5 Movers and Tracing

Of the 7,739 households in the administered gross sample of M3 and M4, 24.8 percent of households were identified as movers and therefore their new addresses needed to be traced (see **Table 10**). The tracing of these households was successful in 18.6 percent of all cases. Most new addresses were gained by the interviewers themselves (91.3%), followed by information provided by the Postal Service (7.0%).

Table 10: **Movers and sources of new addresses of administered sample**

Household Level	M3		M4		Total	
	Abs.	In %	Abs.	In %	Abs.	In %
Gross Sample	4,051	100.0	3,688	100.0	7,739	100.0
Movers	1,156	28.5	760	20.6	1,916	24.8
Outcome Tracing	1,156	100.0	760	100.0	1,916	100.0
Tracing successful	202	17.5	154	20.3	356	18.6
Tracing not successful	954	82.5	606	79.7	1,560	81.4
Source	202	100.0	154	100.0	356	100.0
Interviewer	182	90.1	143	92.9	325	91.3
Postal Service	16	7.9	9	5.8	25	7.0
Participant	4	2.0	2	1.3	6	1.7

## 5.6 Interviewer Characteristics and Training

All interviewers were part of a special interviewer staff established exclusively for the first wave of samples M3/4. Many of them can speak Arabic or Farsi. Due to the fact that many of these interviewers were specifically recruited for the refugee samples, the staff is not only notably younger in age compared to the interviewers in other SOEP samples, they are also less experienced as interviewers. Interviewer characteristics are shown in **Table 11**. 72 Interviewers were deployed in sample M3, 85 in sample M4. 44 Interviewers were deployed in both samples. All interviewers in samples M3 and M4 took part in an extensive, project specific one-day training event prior to fieldwork.

Table 11: Interviewer characteristics

Interviewer Level	M3		M4	
	Abs.	In %	Abs.	In %
<b>Number of Interviewers</b>	<b>72</b>	<b>100.0</b>	<b>85</b>	<b>100.0</b>
<b>Gender</b>				
Male	45	62.5	53	62.4
Female	27	37.5	32	37.6
<b>Age</b>				
21-39	25	34.7	27	31.8
40-59	27	37.5	35	41.2
60-79	20	27.8	23	27.1
<b>Experience with Kantar</b>				
0-4 years experience with Kantar	35	73.6	72	84.7
5-9 years	10	13.9	7	8.2
10-19 years	9	12.5	6	7.1
<b>Experience with SOEP</b>				
0-4 years	62	86.1	78	91.8
5-9 years	5	6.9	5	5.9
10-19 years	5	6.9	2	2.4
<b>Number of Households<sup>1</sup></b>				
1 – 99	60	83.3	65	89.0
More than 100	12	16.7	8	11.0

<sup>1</sup> M4:12 Interviewers that were part of the original staff for this sample did not in fact attend to any households.

## 6 Fieldwork results

### 6.1 Participation on household level

There was no prior empirical evidence for creating a refugee sample of persons entering Germany recently for large-scale quantitative social research. Expectations about what might happen during face-to-face interviews were quite unclear. The challenge was to conduct interviews on-site in different settings with maximum flexibility. Most respondents were living independently in their own households. Some respondents were living in federal or community-organized accommodations for asylum seekers. With these accommodations, the first challenge was to obtain permission to gain access to them. Secondly, because many refugees were expected to move around a lot, at least within Germany, the challenge of encountering them at the address provided by the local recordkeeping authorities was anticipated. And thirdly, the specific background of the target population led to the assumption that most of them would not speak German well enough to be interviewed, also making it difficult for interviewers to communicate their questions. Given this situation, only a very vague estimation of the expected cooperation rate was possible in advance. Nevertheless, the first weeks of the fieldwork period clearly indicated that each possible stage of attrition would not affect this survey as much as other surveys targeting the general population in Germany.

**Table 12** shows the fieldwork results for samples M3 and M4. About one-third of the addresses were either not valid, or their validity was unclear. About 20 percent of the addresses were valid, but an interview could not be obtained for various reasons. In both samples, 1,769 households could be interviewed. As the SOEP standard procedural preference of interviewing all household members was not a major objective for the first wave of the two refugee samples, the number of households that were partially interviewed (i.e., at least one individual questionnaire was missing) is comparably high. However, 582 additional people in M3 and 697 in M4 were interviewed, resulting in 2,351 individual interviews in M3 and 2,466 in M4.

**Tables 13 and 14** show the fieldwork results by different gross sample definitions and **Table 15** the different outcome rates for both samples. Only a few cases had to be excluded because the key informant was an unaccompanied refugee minor. In addition, it was seldom the case that the key informant belonged to a household in which another key informant who we had already interviewed also lived. Those cases are listed as QNDs in Tables 13 and 14. With these kinds of attrition and the low number of invalid addresses, there were a lot of addresses left over for fieldwork. And this proved to be very crucial.

In M3, the interviewers could not process 27.4 percent of all addresses (gross sample I). That defines gross sample II as containing 2,941 viable addresses. After adjusting for deceased key informants and those who had moved abroad, 2,863 addresses remained (70.7 percent of gross sample I). Overall, the interviewers were able to contact 2,547 key informants, that is, 62.9 percent of gross sample I. Compared to the recent general refresher samples in the SOEP, the response rate of 61.8 percent, defined as the number of interviews divided by adjusted gross sample II, is very high.

In M4, the interviewers were unable to process 21.1 percent of all addresses (gross sample I). That defines gross sample II as containing 2,910 viable addresses. After adjusting for deceased key informants and those who had moved abroad, 2,845 addresses remained (77.1 percent of gross sample I). Overall, the interviewers were able to contact 2,550 key informants, that is, 69.1 percent of gross sample I. The response rate in M4 of 62.2 percent, defined as the number of interviews divided by adjusted gross sample II, is as remarkable as in M3.

Table 12: **Participation on household level**

Household level	M3		M4		Total	
	Abs.	in %	Abs.	in %	Abs.	in %
<b>Gross sample for fieldwork</b>	<b>4,051</b>	<b>100.0</b>	<b>3,688</b>	<b>100.0</b>	<b>7,739</b>	<b>100.0</b>
<b>Unknown eligibility</b>	<b>1,194</b>	<b>29.5</b>	<b>832</b>	<b>22.6</b>	<b>2,026</b>	<b>26.2</b>
- Unable to reach during fieldwork period	316	7.8	295	8.0	611	7.9
- Key informant moved and unable to obtain address	878	21.7	537	14.6	1,415	18.3
<b>Not eligible (e.g. business address, address does not exist)</b>	<b>232</b>	<b>5.7</b>	<b>241</b>	<b>6.5</b>	<b>473</b>	<b>6.1</b>
<b>Eligible, non-interview</b>	<b>856</b>	<b>21.1</b>	<b>846</b>	<b>22.9</b>	<b>1,702</b>	<b>22.0</b>
- Key informant deceased or permanently living abroad	78	1.9	65	1.8	143	1.8
- Permanently physically or mentally unable / incompetent	37	0.9	24	0.7	61	0.8
- Language problems	153	3.8	158	4.3	311	4.0
- "Soft refusal" (currently not willing / capable)	372	9.2	435	11.8	807	10.4
- Permanent refusals	115	2.8	123	3.3	238	3.1
- Other (e.g. detained, refusal by refugee housing)	101	2.5	41	1.1	142	1.8
<b>Interview (of key informant)</b>	<b>1,769</b>	<b>43.7</b>	<b>1,769</b>	<b>48.0</b>	<b>3,538</b>	<b>45.7</b>
- Household completely interviewed (including single households)	1,291	31.9	1,035	28.1	2,326	30.1
- Household partially interviewd	478	11.8	734	19.9	1,212	15.7

Table 13: Fieldwork results M3 in different gross samples

	Abs.	in % gross sample I	In % gross sample II
Gross sample I (all gross addresses for fieldwork)	4.051	100,0	
Non processable addresses (not attempted; anchor person moved / unable to obtain new address; QNDs)	1.110	27,4	
Gross sample II (processable addresses)	2.941	72,6	
Deceased and moved abroad	78	1,9	
Gross sample II adjusted	2.863	70,7	100,0
Unable to reach during fieldwork period	316	7,8	11,0
Contacted processable addresses	2.547	62,9	89,0
Non Cooperation (Permanently unable / incompetent; Language problems; soft and permanent refusals)	778	19,2	27,2
Valid Interviews	1.769	43,7	61,8

Table 14: Fieldwork results M4 in different gross samples

	Abs.	in % gross sample I	In % gross sample II
Gross sample I (all gross addresses for fieldwork)	3,688	100,0	
Non processable addresses (not attempted; anchor person moved / unable to obtain new address; QNDs)	778	21.1	
Gross sample II (processable addresses)	2,910	78.9	
Deceased and moved abroad	65	1.8	
Gross sample II adjusted	2,845	77.1	100,0
Unable to reach during fieldwork period	295	8.0	10.4
Contacted processable addresses	2,550	69.1	89.6
Non Cooperation (Permanently unable / incompetent; Language problems; soft and permanent refusals)	781	21.2	27.5
Valid Interviews	1,769	48.0	62.2

Table 15: Outcome rates M3 and M4 at household level

Household level	M3		M4	
	In % gross sample I	In % gross sample II adjusted	In % gross sample I	in % gross sample II adjusted
Contact rate (contacted addresses / gross sample)	62,9	89,0	69,1	89,6
Response rate (interviews / gross sample)	43,7	61,8	48,0	62,2

## 6.2 Participation on individual level

**Table 16** presents participation on an individual level. Of altogether 6,394 individuals in the gross sample, 75.3 percent were successfully interviewed, while 24.7 percent either dropped out temporarily (22.4 percent) or permanently (2.3 percent). The low percentage of final dropouts is a good result, especially with regards to the second wave and the potential conversion of soft refusals and their future inclusion in the panel. The response rates on the individual level are provided in **Table 17**.

Table 16: **Participation on individual level**

Individual level	M3		M4		Total	
	Abs.	In %	Abs.	In %	Abs.	In %
<b>Gross Sample<sup>1</sup></b>	<b>3,012</b>	<b>100.0</b>	<b>3,382</b>	<b>100.0</b>	<b>6,394</b>	<b>100.0</b>
<b>Interview</b>	<b>2,351</b>	<b>78.1</b>	<b>2,466</b>	<b>72.9</b>	<b>4,817</b>	<b>75.3</b>
<b>Non-Interview</b>	661	21.9	916	27.1	1,577	24.7
<b>Temporary dropout</b>	<b>563</b>	<b>18.7</b>	<b>868</b>	<b>25.7</b>	<b>1,431</b>	<b>22.4</b>
Non-contact	27	0.9	157	4.6	184	2.9
Temporary refusal	416	13.8	594	17.6	1,010	15.8
Temporarily physically or mentally unable/incompetent	48	1.6	27	0.8	75	1.2
Language problem	34	1.1	62	1.8	96	1.5
Other temp.	38	1.3	28	0.8	66	1.0
<b>Final Dropout</b>	<b>98</b>	<b>3.2</b>	<b>48</b>	<b>1.4</b>	<b>146</b>	<b>2.3</b>
Permanent Refusal	88	2.9	44	1.3	132	2.1
Permanently physically or mentally unable/incompetent	10	0.3	4	0.1	14	0.2

<sup>1</sup> All household members of participating households.

Table 17: **Response rate**

	M3	M4	Total
<b>Response Rate</b>	<b>78.1</b>	<b>72.9</b>	<b>75.3</b>

One major concern for all SOEP samples are the growing partial unit non-response (PUNR) rates, which are exceptionally high for the refugee samples. PUNR in sample M3 was at 53.7 percent in this year's first wave (**Table 18**) which is slightly lower than the PUNR in the sister sample M4 (57.3 percent). However, in the first wave of the two refugee samples the interviewer focused on interviewing the sampled key informant. From wave two on the SOEP standard procedural preference of interviewing all household members will become a major objective.



Table 18: Partial unit non response

	M3	M4
PUNR <sup>1</sup>	53.7	57.3

<sup>1</sup> Share of households (number of household members > 1) with at least one missing individual questionnaire.

## 6.3 Participation by types of questionnaires

**Table 19** presents the number of interviews and response rates for each of the two fielded questionnaires. Corresponding with the number of households in the net sample, 1,769 household questionnaires were produced in both samples, resulting in an adjusted response rate of 61.8 percent (sample M3) respectively 62.3 percent (sample M4). The response rates for the individual and life history questionnaire are 78.1 percent (sample M3) respectively 72.9 percent (sample M4).

Table 19: Number of interviews and response rate by questionnaire

	M3			M4		
	Gross sample <sup>1</sup>	Number of interviews	Response rate	Gross sample <sup>1</sup>	Number of interviews	Response rate
Household questionnaire	2,863	1,769	61.8	2,845	1,769	62.2
Individual questionnaire + life history questionnaire module	3,012	2,351	78.1	3,382	2,466	72.9

<sup>1</sup> Gross sample = target population in participating households, excluding deceased, moved abroad.

## 6.4 Interview length per questionnaire

The media interview length for key informants in both refugee samples was about 100 minutes (see **Table 20**). In many cases, the interview lasted three hours or more. The median interview length for other household members who completed the individual questionnaire without the household section was about 80 minutes. Further, one must add significant time for the contact phase. In many cases the interviewers had to answer questions after the interview was completed. Therefore, the survey was very demanding and time-consuming for both the interviewees and the interviewers.

Table 20: Median interview length in minutes

	M3	M4
Household questionnaire	18	20
Individual questionnaire + life history questionnaire module	84	76

## 6.5 Consent to Record Linkage

In recent years it has to a certain extent become standard to link respondents' survey data in the SOEP with registry data from the Integrated Employment Biographies Sample provided by the Institute for Employment Research (IAB). All interviewees were asked to give their written consent to the record linkage at the end of the individual interview. Table 21 shows the results for approval or rejection.

Table 21: **Consent to Record Linkage**

Individual level	M3		M4	
	Abs.	In %	Abs.	In %
Approved	1.828	77,8	2.006	81,3
Declined	197	8,4	227	9,2
Didn't understand the issue	326	13,9	233	9,4
Total	2.351	100,0	2.466	100,0

## 6.6 (App-)Follow Up

For this very specific SOEP top-up sample, it was assumed that the targeted population of refugees arriving in Germany from 2013 onwards would have significantly higher mobility compared to the general population. This pertains to both spatial mobility within Germany and their potential return to the foreign home country. This refugee survey also faces huge challenges with regard to panel management and respondents' commitment to the study "Living in Germany" for the medium and long term. The idea was to tackle these challenges with an innovative new approach: Upon completion of the survey, participants will be offered a smartphone application by the interviewers on site. The aim of this special instrument is to keep in touch with the panel members until the start of the next F2F survey wave in 2017.

The app primarily serves as a means of contact and information source for respondents to stay up to date on both the study and refugee and migration issues in Germany in general. Secondly, there is a function for conducting very short app surveys during the year. On the one hand, this should give researchers the opportunity to collect additional information. On the other hand, it should also increase the interviewees' commitment to "Living in Germany" as a whole. Thirdly, the respondents are offered a profile for updating and managing personal data, such as names and addresses, should changes in their living conditions arise. And last but not least, the decision was made to use the smartphone's GPS technology in order to be able to carry out passive technology-supported address management for consenting respondents. Of course, all the regulations of research ethics and data protection have been rigorously observed. Participants were provided with special clarifications, and all were asked individually for explicit consent regarding each stage of active and passive data usage. This piece of research is regarded as an innovative methodological pilot in the field of survey practice.

Table 22: Consent to app usage: Compliance rates

	M3		M4	
	Abs.	In %	Abs.	In %
Approved	783	44,1	634	35,9
Declined	542	30,6	723	41,0
Didn't understand the issue	176	9,9	229	13,0
No smartphone	205	11,6	147	8,3
Don't know	68	3,8	32	1,8
Total	1.774	100,0	1.765	100,0

Table 23: Consent to app usage: Installation of app

	M3		M4	
	Abs.	In %	Abs.	In %
Downloaded and installed	470	60,0	323	50,9
Downloaded but not installed	134	17,1	212	33,4
No download possible	179	22,9	99	15,6
Total	783	100,0	634	100,0

**Table 22 and 23** indicate that this innovation did not roll out as expected. Only around two of five key informants (M3: 44.1 percent, M4: 35.9 percent) agreed to download the app. While just around one-tenth did not own a smartphone (M3: 11.6 percent, M4: 8.3 percent) or did not understand the issue (M3: 9.9 percent / M4: 13.0 percent), a large portion of the respondents refused the download (M3: 30.6 percent, M4: 41.0 percent) altogether.

Further, there have been technical issues with downloading and installing the app. The interviewers were supplied with a UMTS device that allowed them to offer interviewees a Wi-Fi hotspot in case they did not have their own Wi-Fi access or the credit required for using their provider's Internet services. However, this was not sufficient to facilitate installation of the app, as there were further technical issues when trying to install it on different devices, especially older hardware and software versions that use rather uncommon configurations. Therefore, installation was technically only possible in up to three out of five cases (M3: 60.0 percent, M4: 50.9 percent). Consequently, only one of four key informants was able to successfully install the app on their smartphone (M3: 26.5 percent, M4: 18.3 percent).

## 7 Data preparation

In a large-scale panel study such as the SOEP, data processing and data preparation are fundamentally important processes of quality management. During the entire process, we must ensure that we generate and provide optimal data quality that is consistent over the years: Only then can the necessary longitudinal consistency be generated which is required by the SOEP users that conduct panel analyses. Data processing and data preparation are elements of complex and multistage work processes. Moreover, the procedures and techniques of the data checking and data cleaning are constantly being further developed: partly through learning processes that are “content-related”, partly through optimizations that occur as a result of IT innovations. However, over the years, the main features remain unchanged and have been presented in more detail in the field report 2011 by Kantar<sup>1</sup>. The DIW receives the (net) data in two forms, the adjusted and unadjusted data sets. If required, data inferences can thus be identified and traced any time on a case-by-case basis.

The major elements of data processing and data preparation in the SOEP are data collection, data checking, data cleaning and data enrichment:

- **Data collection** includes all activities of the production process that serve to provide the collected data for further processing and preparation of the data.
- **Data checking** is the most important element of quality management, because the checking criteria defined here specify to what extent and at which points checks are carried out, and consequently, to what extent the collected raw data is edited.
- **Data cleaning** includes all direct inferences into the collected data at the individual case level, which are usually corrections of data errors in order to generate cross-sectional and panel consistency.
- In the **data enrichment** process, new information (both at the individual and the aggregated level) is generated and added to the data set as an additional variable. This means e.g. encoding open answers or adding spatial indicators.

The various elements of the data processing procedures cannot be strictly separated. They are rather interdependent and continuous processes. For instance, certain checking steps are already integrated into the data collection, and a large part of the data cleaning is conducted as part of the data checking. In addition to regarding the individual elements, it is useful to outline the process stages that emerge during data processing.

### Preparatory activities:

- Creating or revising the data-entry screens and check programs
- Programming the data entry screens (scan programs)
- Testing the programs

---

<sup>1</sup> SOEP Methodenbericht 2011, Huber et al.

## **Collection of data and address protocols; first preliminary checks:**

- Collecting the address protocols (entry into the panel file)
- 1<sup>st</sup> checks on completeness of the households (correct quantity/type of completed questionnaires)
- First content checks (e.g. invalid multiple answers)
- Subsequent collection of certain information/missing questionnaires

## **Gross related basic checking:**

- Checks on completeness and consistency of the various data sets per household (household data, individual data)
- Comparing the gross information (panel file) with questionnaire and address protocol information (identity, marital status, household composition, address)
- Clarifying inconsistencies/irregularities with the target households

## **Net related checking:**

- Automatic checking of the single questionnaires according to certain processing rules and automatic setting of codes or completion/cleaning of the data
- Key checking elements: filtering, invalid multiple responses, total checks, value range, implausible answers
- Manual case-by-case checks for unclear cases
- Editing the interviewers' notes and respondents' comments (e.g. omitting names and addresses)

## **Data enrichment/coding:**

- Coding open plaintext answers on the job title and the industry as well as the educational level obtained based on the most recent classification schemes
- Adding regional indicators (postcode, official municipality key, etc.)
- Adding Microm data (including geocoded addresses, GPS codes)
- Geocoding of places of birth inside and outside Germany

## **Final checking:**

- Checking the entire prepared data stock (household questionnaire, individual questionnaire) for completeness and possible irregularities (including the correction of program and checking errors)

## **Data delivery:**

- Transfer of pre-checked intermediate data after half of the field time
- Transfer of the code book (description of the variables and documentation of the coding rules)
- Transfer of the final integrated and standardized gross and net data files (anonymized) in checked and unchecked form

It should be mentioned here that the individual process steps are not conducted in a fixed order, but mostly simultaneously, interdependently and in several loops. In order to be able to deliver the checked, cleansed and enriched data promptly, the major part of the described work steps is organized "on a weekly basis" (according to "field weeks"). This means that the collected data are not gathered and processed at once at the end of the survey phase, but continuously right from the beginning of fieldwork. This means that almost all process steps are conducted on an ongoing basis.

## 8 Delivered data

### Sample M3:

#### **Gross Data**

Codebook Panel data  
Gross data Household  
Gross data Individuals  
SOEP Individuals Sample M5  
Interviewer data  
Residential environment data

Codebuch 2016.pdf  
Hbru\_M3\_2016.sav  
Pbru\_M3\_2016.sav  
P\_M3\_2016.sav  
Intband\_M3\_2016.sav  
Wuma\_M3\_2016.sav

#### **Net Data**

Net data Household checked  
Net data Household unchecked  
Net data Individuals + Life history Refugees checked  
Net data Individuals + Life history Refugees unchecked  
Information on 7 and more children  
Consent to record linkage

H16\_M3.sav  
H17\_M3\_u.sav  
PB16\_M3.sav  
PB16\_M3\_u.sav  
7bis10Kind\_2016\_M3.sav  
EV16\_M3\_DIW.sav

#### **Other Data**

Professions, sectors, final coding  
Variable list  
Additional Codes for Individual + Life history Refugee

Berufe\_Branchen\_2016\_M3.sav  
Struktur\_2016.xlsx  
Nationen.xlsx

## Sample M4:

### Gross Data

Codebook Panel data  
Gross data Household  
Gross data Individuals  
SOEP Individuals Sample M5  
Interviewer data  
Residential environment data

Codebuch 2016.pdf  
Hbru\_M4\_2016.sav  
Pbru\_M4\_2016.sav  
P\_M4\_2016.sav  
Intband\_M4\_2016.sav  
Wuma\_M4\_2016.sav

### Net Data

Net data Household checked  
Net data Household unchecked  
Net data Individuals + Life history Refugees checked  
Net data Individuals + Life history Refugees unchecked  
Information on 7 and more children  
Consent to record linkage

H16\_M4.sav  
H17\_M4\_u.sav  
PB16\_M4.sav  
PB16\_M4\_u.sav  
7bis10Kind\_2016\_M4.sav  
EV16\_M4\_DIW.sav

### Other Data

Professions, sectors, final coding  
Variable list  
Additional Codes for Individual + Life history Refugee

Berufe\_Branchen\_2016\_M4.sav  
Struktur\_2016.xlsx  
Nationen.xlsx