SOEP The German Socio-Economic Panel study

473

SOEP Survey Papers

Series C – Data Documentation

SOEP — The German Socio-Economic Panel study at DIW Berlin 2017

The 2015 IAB-SOEP Migration Study M2: Sampling Design, Nonresponse, and Weighting Adjustment

Simon Kühne, Martin Kroh

DIW SOEP

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 Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

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

Please cite this paper as follows:

Simon Kühne, Martin Kroh. 2017. The 2015 IAB-SOEP Migration Study M2: Sampling Design, Nonresponse, and Weighting Adjustment. SOEP Survey Papers 473: Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)
DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany
soeppapers@diw.de

The 2015 IAB-SOEP Migration Study M2:

Sampling Design, Nonresponse, and Weighting Adjustment

Simon Kühne Martin Kroh

October 7, 2017

DIW Berlin – Deutsches Institut für Wirtschaftsforschung e.V.

Sozio-oekonomisches Panel (SOEP)

Mohrenstraße 58

10117 Berlin

Contents

List of Tables

List of Figures

1 Introduction

The 2015 IAB-SOEP Migration Study M2 is a survey of private households of immigrants conducted jointly by the Institute for Employment Research (IAB) in Nuremberg and the Socio-Economic Panel Study (SOEP) at DIW Berlin. The 2015 IAB-SOEP Migration Study M2 is the follow-up sample of the 2013 IAB-SOEP Migration Study M1 (see Brücker et al. 2014 and Kroh et al. 2015).

With the second Migration Sample, M2, the IAB and the SOEP aim to further improve the data basis for research on migration to Germany by adding a refresher sample of households with members who migrated to Germany between 2009 and 2013. The sample consists of 1,096 households containing 1,689 adult respondents and close to 1,000 children. Integrating the sample into the SOEP allows for in-depth analysis of recent immigration to Germany, and generally increases the statistical power of analyses of integration.

About IAB and SOEP:

The **Institute for Employment Research ("Institut für Arbeitsmarkt- und Berufsforschung", IAB)** is an independent institute of the Federal Employment Agency in Nuremberg, Germany. Its work focuses on German labor market and occupational research in areas such as labor market policy and social inequality. The IAB also does research on statistical methods and survey methodology. The Research Data Centre (FDZ) of the Federal Employment Agency is based at the IAB and offers a wide variety of data for research purposes. For more information about the IAB, see: `http://www.iab.de/en/ueberblick.aspx`.

The **German Socio-Economic Panel (SOEP)** is a longitudinal survey of private households in Germany based at the German Institute for Economic Research (DIW Berlin) that has been conducted annually since 1984. Since 2002, the SOEP has been receiving federal and state funding through the Joint Science Conference (GWK). Before that, its funding came mainly from the German Research Foundation (DFG). The survey provides information on various topics such as household composition, employment, health, and attitudes. The 2015 SOEP data provide information on 19,236 households from 37,315 individual questionnaires (see Kroh et al. 2016).

2 Target Population and Sampling Frame

With the 2015 Migration Study, M2, IAB and SOEP are focusing on recent immigrants who arrived in Germany between 2009 and 2013. Surveying immigrants to Germany is beset by a number of difficulties (see also Kroh et al. 2015). While data on this group is available from the “Central Register of Foreigners” (Ausländerzentralregister) of the Federal Office for Migration and Refugees (Bundesamt für Migration und Flüchtlinge, BAMF), access is only provided to the Federal Institute of Population Research at the BAMF. Moreover, although municipal registries (Melderegister) contain information on residents’ nationalities, sampling naturalized citizens with a migration background and targeted sampling of certain immigrant cohorts is impossible in practice (Salentin 1999).[1] Alternative sampling strategies used in the SOEP have employed large numbers of screening interviews, for instance by telephone (Sample D), or onomastic procedures that rely on address information (Sample F, H, J and M1). However, due to the comparatively small number of members of the target population within the overall population, screening interviews would be highly inefficient and expensive. The same holds for onomastic procedures (Humpert/Schneiderheinze 2013) based on selection by family names visible next to doorbells, but here it is impossible to determine the number of years since immigration. Finally, telephone-based screening to construct a migrant sample has become increasingly problematic in recent years, given the increasing number of households that either do not have a land line or are not listed in the telephone directory. These issues tend to affect immigrant populations in particular (Lipps/Kissau 2012).

As an innovative alternative to the aforementioned approaches – and similarly to our strategy in the first IAB-SOEP Migration Study M1 in 2013 – we use register data from the Federal Employment Agency (FEA) as a sampling frame. The Integrated Employment Biographies (IEB) data set 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 onwards. The IEB is available since 2004 and provides information from as far back as 1990 (Oberschachtsiek et al. 2009). The IEB can be understood as the result of a merging procedure of different process-produced databases (Oberschachtsiek et al. 2009, Jacobebbinghaus/Seth 2007): 1) Employment: IAB Employee History (BeH), 2) Unemployment: Benefit Recipient History (LeH), 3) Active labor market policies: Participants-in-Measures History File (MTH), and 4) Job search: Applicant Pool

[1] One exception are immigrants living at their first registered address in Germany (Diehl 2007).

Data (BewA).

Employment data exist for all employees in jobs that are subject to social security contributions, which describes almost all private sector employment in Germany. 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. Data is stored in spells that are linked to individuals. The spells are accompanied by a variety of socio-demographic variables such as gender, age, and nationality as well as geographic information such as address, municipality, and regional classification. In total, the IEB contains 83,521,672 individuals with 1,894,018,836 spells.[2] Furthermore, information on unemployment spells, receipt of benefits, participation in active labor market policies, and job-search status are directly matched from the different sources within the social security system to form a complete picture of individuals' labor market history.

IEB spell information on individuals' current and past nationalities enabled us to construct an appropriate sampling frame for our study purposes. Using the IEB also provides further benefits. First, the IEB is a centralized sampling frame, whereas register offices in Germany work at the local level, and a national sampling frame would require working with each of the sampled municipalities individually. Second, the wealth of data on individual labor market participation and wages as well as employer information allows researchers to model non-response processes more fully than many alternative sampling frames. The (model-based) weighting of the data in the IAB-SOEP Migration Study M2 – for instance, the data on employment status – thus corrects for any deviation in registered employment status in the IEB between the gross and the net sample. Third, the IEB sampling frame allows for a linkage of survey and register data in subsequent research projects (record linkage).

Nonetheless, using the IEB as a sampling frame has some disadvantages as well. Although the database represents a great share of the target population, some groups are not covered (on undercoverage in the IEB, see Jacobebbinghaus/Seth 2007). 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 (Beamte) are not. Moreover, self-employed people who have never held a job that is subject to social security contributions and have never received unemployment benefits or taken part in an active labor market

[2] Cut-off date: December 31, 2013.

policy measure are not covered in the IEB. The IEB covers about 80 percent of the cross-section of the German labor force (Jacobebbinghaus/Seth 2007: 336). An estimation based on the SOEP and the German Microcensus in the context of the IAB-SOEP Migration Study M1 has shown that by choosing the IEB as a sampling frame, 5 to 8 percent of the target population is excluded. However, by considering not only individuals' current but all of their previous employment spells, we only exclude individuals from sampling who never were in contact with the FEA. Finally, excluded groups, such as individuals who were always self-employed, may enter the survey as members of the same households as sample "anchor" respondents (see below), but possibly less than proportionally.

3 Overview of the Sampling Design

This section describes the multi-stage sampling design used in the study M2. The different steps of the sampling procedure are shown in Figure 1. First, the target population of recent migrants to Germany was preliminarily identified on the basis of the year in which they entered the IEB as well as former and current citizenship. Next, the target population was regionally clustered into 3,288 sampling points with a minimum of 160 persons from the target population. These clusters represent combinations of postal codes and municipalities. Next, a sample of 125 geographically clustered Primary Sampling Units (PSUs) was drawn. Finally, using a disproportional sampling design, we sampled 80 addresses for fieldwork – the Secondary Sampling Units (SSUs) – in each of the selected PSUs.

Below, we describe each of the steps in detail.

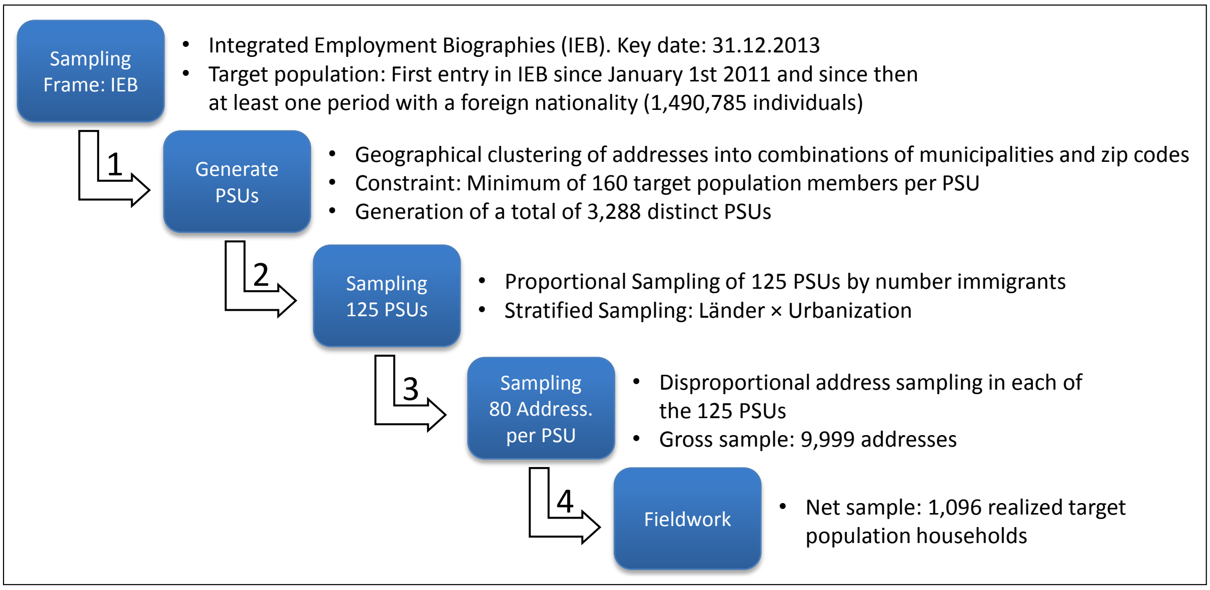


Figure 1: Sampling Design of the IAB-SOEP Migration Study M2

3.1 Identification of Target Population Members

The first step in the sampling process was the identification of target population members on the basis of the IEB register information. The selection of individuals as eligible was based upon three conditions. First, individuals had to be part of the IEB database as of December 31, 2013 (version V11.01.01). This gave us 37,587,723 individuals. Second, we used only those individuals who were first registered with the IEB register as of January 1, 2011, and third, individuals who had since had at least one period (at least one spell) of holding a foreign (non-German) nationality (1,490,785 individuals were identified).

3.2 Geographic Clustering

After excluding addresses with missing values as well as invalid addresses such as post office box and business addresses, the sampling frame of target population members amounted to 1,292,618 individuals.

In the next step, individual addresses were clustered using combinations of official municipality keys and postal codes. First, each address was assigned to the smallest geographic area possible. In large cities, this is usually a postal code, while in rural regions, the municipality is usually smaller than the area covered by a postal code. This clustering resulted in a total of 13,467 distinct geographic areas. Second, each point was supposed to consist of at least 160 target population members. Points that did not meet that condition were combined with other points based on a set of rules: On the one hand, points had to be in the same municipality or postal code. On the other hand, points had to be combined with other small points. And finally, points had to be combined with nearby points.

The final gross sample of primary sampling units consisted of 3,288 distinct and disjoint sample points.

3.3 Sampling of Primary Sampling Units

We used a stratified proportional sampling design to select 125 primary sampling units (PSUs) out of the total of 3,288 sample points. PSUs were assigned to 19 different strata in accordance to (1) federal states (*Bundesländer*), (2) county *(Kreis)* type, either rural or urban, and (3) the proportion of migrants in the PSU. Out of the total of 3,288 PSUs across Germany, 125 PSUs were randomly selected from the various strata considering the number of both PSUs and target population members per strata. The number of points to be drawn in each stratum relates to the number of target population members in each. Thus, sampling is based on a *probability proportional to size* approach (PPS design). Selecting primary sampling units proportional to size and sampling a fixed number of secondary sampling units in each point results in a self-weighted sampling design. The 125 sampled PSUs covered 414 distinct postal code/municipality combinations comprising a total of 71,353 target population members.

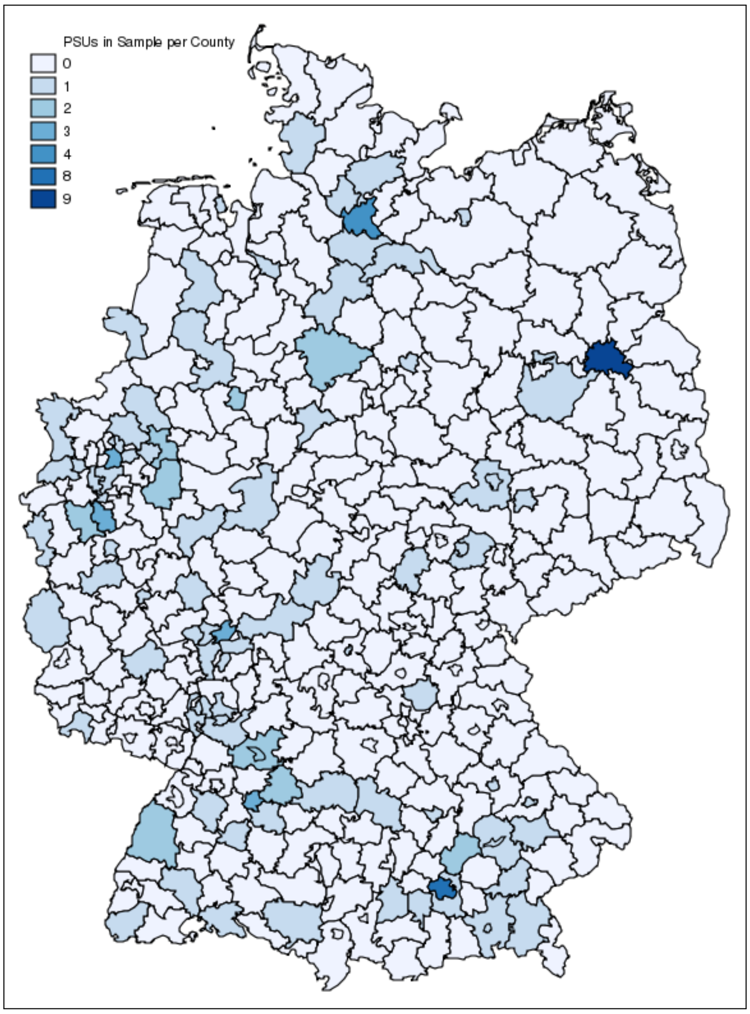


Figure 2: Number of Sampled PSUs across Counties

Figure 2 displays the geographical distribution of the selected 125 PSUs at the county level. For instance, 9 sample points were selected in the Berlin area.

3.4 Sampling of Secondary Sampling Units

Next, we sampled a total of 80 individuals in each of the 125 PSUs, resulting in a gross sample of 10,000 individuals. We applied a disproportional sampling procedure to ensure that specific subpopulations are represented sufficiently in the final sample. Therefore, different sampling probabilities were assigned in accordance with the individuals' country of origin. Table 1 displays the variation in sampling probabilities.

The highest sampling probabilities were assigned to migrants from Bulgaria (2.11) and Romania (1.66). Additionally, migrants from Italy, Portugal, Spain, and Greece were

Table 1: Sampling Probability of SSU by Migration Background

Country of Origin	Sampling Probability
Poland, Hungary, Arabic Countries, Western Europe (Rest), Eastern Europe (Rest)	1.00
Bulgaria	2.11
Romania	1.66
Italy, Portugal, Spain, Greece	1.33
Turkey	0.44
Rest of World	0.88

slightly oversampled as well. In contrast, lower sampling probabilities were assigned to migrants from Turkey (0.44) and from the rest of the world (0.88).

Table 2 displays the frequency and percentages of individuals with different nationalities in the sampling frame as well as the frequency and percentages in the final gross sample consisting of 9,999 individuals.

Table 2: Composition of Nationalities in the Sampling Frame (IEB) and in the Gross Sample

	Sampling Frame		Gross Sample	
Current Nationality	Freq.	%	Freq.	%
Germany	2,249	3.15	80	0.80
Poland	10,924	15.31	1,601	16.01
Romania	4,886	6.85	901	9.01
Italy	3,795	5.32	537	5.37
Bulgaria	3,210	4.50	512	5.12
Hungary	3,255	4.56	352	3.52
Portugal/Spain	3,458	4.85	470	4.70
Greece	2,834	3.97	297	2.97
Turkey	7,138	10.00	273	2.73
Islamic Countries	6,475	9.07	868	8.68
Western Europe (Rest)	5,200	7.29	968	9.68
Eastern Europe (Rest)	9,375	13.14	1,698	16.98
Rest of the World (Rest)	8,554	11.99	1,442	14.42
Total	71,353	100.00	9,999	100.00

4 Results from Fieldwork and Response Rates

Fieldwork for the IAB-SOEP Migration Study M2 was carried out from May to December of 2015. Target sample households were sent a letter with information about the survey prior to the actual interview. All data was collected in computer-assisted personal interviews (CAPI). A total of 129 interviewers interviewed between 1 and 42 households each, with a mean of 8.5 households per interviewer.

Table 3: Final Results from Fieldwork (Households)

	n	%	n	%
Quality Neutral Drop-Out				
Moved abroad	327	3.3	-	-
Deceased	8	0.1	-	-
Non-existent/invalid address	2424	24.2	-	-
Other	427	4.3	-	-
Subtotal	*3,186*	*31.9*	-	-
Response				
Full/Partial	1,096	11.0	1,096	16.1
Screenout	863	8.6	863	12.7
Subtotal	*1,959*	*19.6*	*1,959*	*28.8*
Nonresponse				
Refusal	1,532	15.3	1,532	22.5
Non-contact	1,992	19.9	1,992	29.2
Language Problems	239	2.4	239	3.5
Other Reason	1,091	10.9	1,091	16.0
Subtotal	*4,854*	*48.5*	*4,854*	*71.2*
Total	9,999	100.0	6,813	100.0

Table 3 displays the final results of the fieldwork. From the 9,999 initially sampled households ("anchor" respondents), 3,186 households were classified as "quality-neutral drop-outs" either because the address was invalid/non-existent (2,424), the anchor respondent had already moved abroad (or back to the country of origin) (327), the anchor respondent was deceased (8), or due to other reasons, including that not all addresses were used in the fieldwork (427). From the reduced gross sample of 6,813 households, 3,868 households participated in a screening interview. Thus, the overall mean response rate for Sample M2 was 28.8%.[3]

Figure 3 displays response rates by federal states (left) and counties (right). As can be seen from the maps, response rates show notable regional variation. For instance, according to the map on the left, households in Northern Germany seem to be more likely to participate

[3]AAPOR Non-Response Definition RR2, see AAPOR (2011).

in the survey. At a county level, the variation in response rates is even greater, but there is no distinct geographical pattern. These differences may in part also reflect the performance of interviewers allocated to these specific counties.

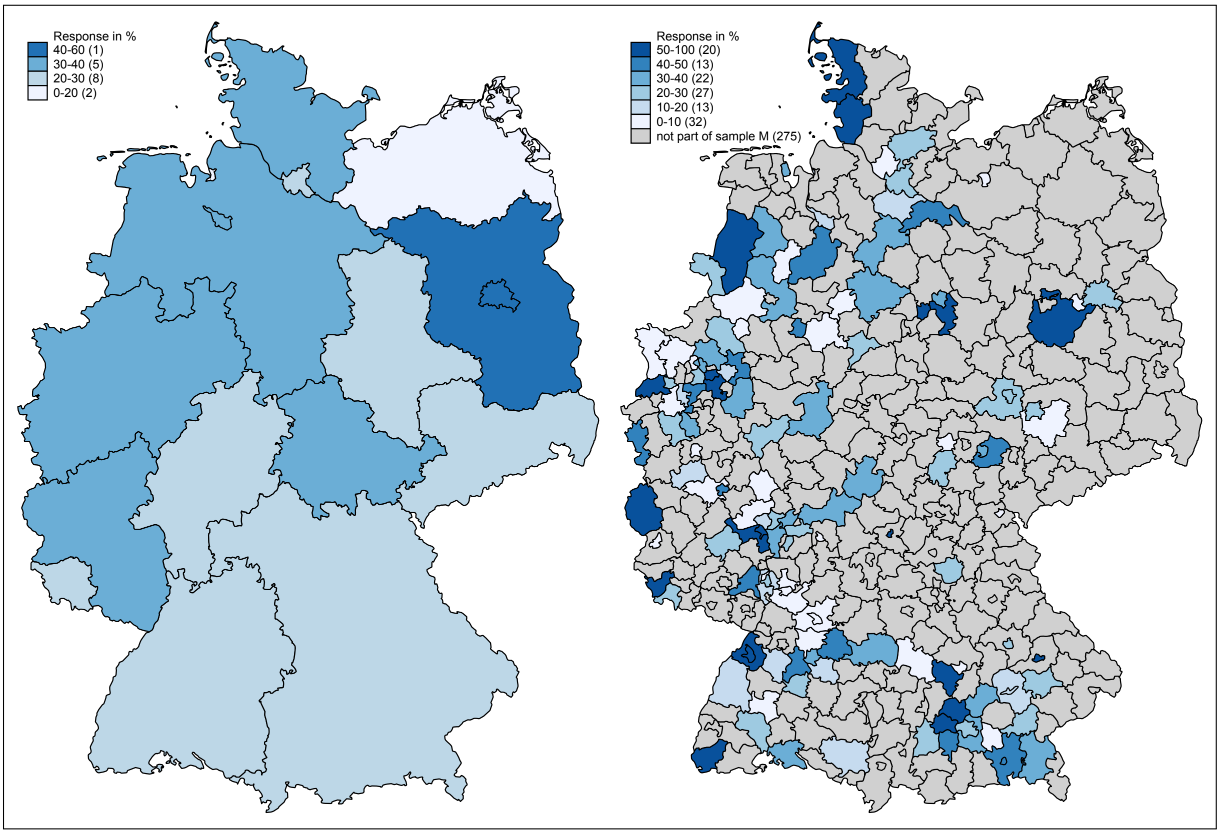


Figure 3: Response Rates by *Länder* and Counties

Furthermore, response rates also vary across migrant groups, as can be seen from Table 4. Subgroup participation rates range from 20.9% in the group of migrants from Hungary to 46.2% for migrants from Turkey. Explaining the variation in these response rates is one of the main tasks in the nonresponse weighting adjustment discussed in section 5.2.

Table 4: Response Rates by Migrant Groups

	Gross Sample	Gross Sample w/o. QND*	Response	Response Rate (%)	Net Sample
Background	n	n	n	%	n
Germany	80	74	33	44.6	9
Poland	1601	1001	218	21.8	132
Romania	901	598	147	24.6	109
Italy	537	341	91	26.7	44
Bulgaria	512	380	94	24.7	69
Hungary	352	187	39	20.9	37
Portugal/Spain	470	290	67	23.1	45
Greece	297	222	57	25.7	23
Turkey	273	251	116	46.2	9
Islamic Countries	868	717	280	39.1	147
Western Europe (Rest)	968	580	162	27.9	100
Eastern Europe (Rest)	1698	1183	363	30.7	220
Rest of the World (Rest)	1442	989	292	29.5	152
Total	9999	6813	1959	28.8	1096

*Gross Sample without *Quality-Neutral Drop-Outs* such as non-existent addresses and business addresses.

Screening of Target Population Members

So far, the processing of addresses was based on our preliminary estimate of the number of target population members in our sampling frame. The first set of questions in the interview validated these estimates. Interviews were only continued if the anchor respondent passed this screening. Screening was based on a few questions about the respondents' birthplace, their duration of residence, as well as their year of immigration. Figure 4 displays the applied screening scheme. In order to reduce the number of screen-outs, anchor persons who reported having immigrated in 2014 and 2015 were not screened out, even though sample M2 generally targeted immigrants between 2009 and 2013.

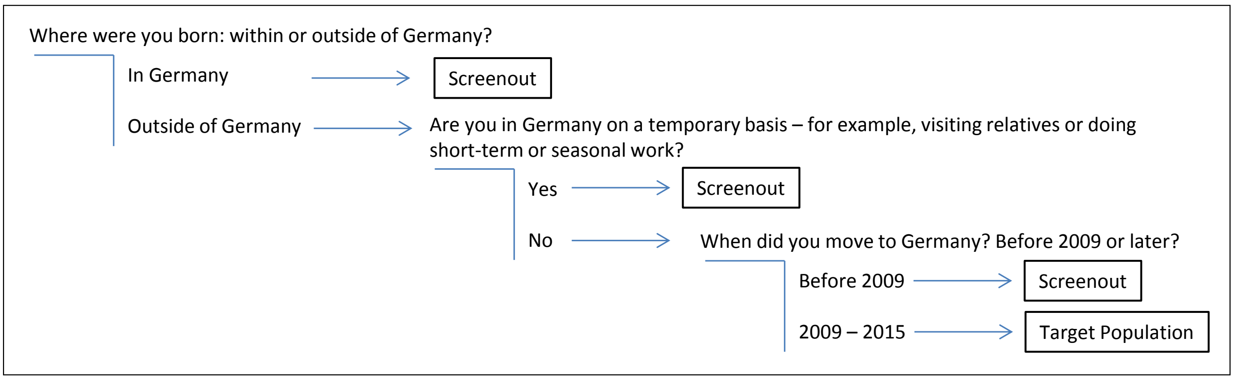


Figure 4: Screening and Identification of Target Population Anchors

Table 5 displays frequencies and total percentages for the different reasons for screen-out grouped by country of origin. Altogether, 863 sampled households were screened out during fieldwork as it turned out that they did not comprise part of the target population. As can be seen from the totals, the two main reasons for screen-outs were that the anchor respondents were born in Germany (414) or had migrated before 2009 (360). Only 89 anchor respondents reported living in Germany on a temporary basis.

Table 5: Household Screen-Outs by Migrant Groups

	Reasons for Screenout			Total Screenout	
Nationality	Seasonal Laborer	Born in Germany	Migration before 2009		
	n	n	n	n	%
Germany	0	17	7	24	34.4
Poland	37	2	47	86	8.6
Romania	24	0	14	38	6.4
Italy	0	38	9	47	13.8
Bulgaria	1	1	23	25	6.6
Hungary	0	1	1	2	1.1
Portugal/Spain	2	13	7	22	7.6
Greece	0	27	7	34	15.3
Turkey	0	95	12	107	42.6
Islamic Countries	3	65	65	133	18.5
Western Europe (Rest)	3	38	21	62	10.7
Eastern Europe (Rest)	10	54	79	143	12.1
Rest of the World	9	63	68	140	14.2
Total	89	414	360	863	12.7

The 863 screened-out households are not of interest in the following analysis of nonresponse and are therefore not taken into further consideration. All in all, the final, reduced gross sample for the nonresponse analysis consists of 5,950 households, of which a total of 1,096 households participated in the survey.

5 Cross-Sectional Weighting of Study M2

In almost all surveys in the social sciences, members of the target population vary in their probability of being interviewed. One the one hand, this may be due to decisions made by the researcher, for example, to use a complex sampling design that assigns different selection probabilities to subgroups of the target population (*selection by design*). On the other hand, these unequal probabilities may result from nonresponse in a subsample of individuals, who, for instance, declined to take part in the survey (*self-selection*).

There are different strategies for dealing with selective samples. Besides specialized model-based strategies (such as the Heckman selection model, see Heckman 1979) and the imputation of data (that is, replacing missing data), the (ex-post) weighting of survey data is the most common way to handle selective samples. Different weighting procedures and techniques have been proposed over the last few decades (Kalton/Flores-Cervantes 2003). One of the most common methods is referred to as "propensity score weighting" (Rosenbaum/Rubin 1983). Propensity score weighting approaches assign sample elements (such as households) a greater "importance" if they have characteristics associated with lower sample probabilities and lower response probabilities and vice versa. In this process, sample elements are weighted by the combination of their inverse sampling probability and their inverse response probability (conditional to being sampled in the first place). The combination allows for an unbiased estimation of population parameters. An example is the mean estimator developed by Horvitz/Thompson 1952:

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{N} \frac{s_i}{\pi_i \cdot P(x_i h \in S)} \tag{1}$$

In equation 1, π_i denotes the response probability of household i and $P(x_i h \in S)$ refers to the sampling probability of household i in stratum h. The variable s_i denotes a binary indicator, taking the value 1 if household i was observed, and 0 if household i was not observed.

In the following sections, we describe the three stages in obtaining a cross-sectional weight for Sample M2. In the first stage, *design-based weighting* corrects for unequal but known probabilities of sampling by the researcher. In the second stage, (model-based) *nonresponse weighting adjustment* corrects for unequal response probabilities. These response probabilities are unknown and are therefore to be estimated. Finally, we compare the net sample with the known margins of the underlying target population (*post-stratification* or *raking*).

5.1 Design Weighting

Design weights account for unequal sampling probabilities of households within sample points. As Section 3.3 describes, the two-stage sampling procedure uses equal selection probabilities. Only in the subsequent step of randomly selecting 80 persons per sample point did we introduce unequal sampling factors across countries of origin:

$$w^{(1)}_{\text{design}} \begin{cases} 1.00 & \text{if Origin} = \text{Germany} \\ 1.00 & \text{if Origin} = \text{Poland} \\ 1.00 & \text{if Origin} = \text{Hungary} \\ 1.00 & \text{if Origin} = \text{Islamic/Arabic Countries} \\ 1.00 & \text{if Origin} = \text{Western Europe (Rest)} \\ 1.00 & \text{if Origin} = \text{Eastern Europe (Rest)} \\ 2.11 & \text{if Origin} = \text{Bulgaria} \\ 1.66 & \text{if Origin} = \text{Romania} \\ 1.33 & \text{if Origin} = \text{Italy} \\ 1.33 & \text{if Origin} = \text{Portugal} \\ 1.33 & \text{if Origin} = \text{Spain} \\ 1.33 & \text{if Origin} = \text{Greece} \\ 0.44 & \text{if Origin} = \text{Turkey} \\ 0.88 & \text{if Origin} = \text{Rest of World} \end{cases}$$

Please note that the SOEP is a household panel survey, and that the IEB sampling frame lists persons. Hence, a household with, for instance, two members of our target population would have twice the sampling probability of a household with a single person from the underlying population. In a second step, we thus correct the design weights by the number of persons in the household that match the definition of the target population[4] by multiplying $w^{(1)}_{\text{design}}$ by the inverse of the number of target population members per

[4]Target population members were identified using information obtained in the interviews based on whether they fit into one of the following groups: (i) the sampled anchors, (ii) those non-anchors who were born outside of Germany and immigrated after 2009, who were not living in Germany on a solely temporary basis (e.g., for seasonal work), and who were unemployed, employed (full-time, part-time), or in an apprenticeship between 2011 and 2013, and who were aged between 15 and 65 during that period of time.

household n_{target} (see equation 2). Correction factors for households with more than three target population members were truncated at the bottom (max = 3) in order to limit the variance in weights:[5]

$$w^{(2)}_{\text{design}} \begin{cases} w^{(1)}_{\text{design}} \times \frac{1}{n_{target}}, & \text{if } n_{target} <= 3 \\ w^{(1)}_{\text{design}} \times \frac{1}{3}, & \text{if } n_{target} > 3 \end{cases} \tag{2}$$

With a final scalar multiplication, we construct design weights at the household level so that their sum over all units of the reduced gross sample of 5,950 equals the number of households in the target population, namely 979,099 (Microcensus 2015). The final household design weights are used as a basis for further weighting procedures.

$$w^{(3)}_{\text{design}} = w^{(2)}_{\text{design}} * \frac{979,099}{\sum_{n=1}^{5,950} w^{(2)}_{\text{design}}} \tag{3}$$

5.2 Nonresponse Weighting Adjustment

In addition to the selectivity resulting from complex sample designs, further selectivity in the observed data may be introduced by choices made by the selected units of analysis to participate or not participate in the survey (*(self)selection into surveys*).

Several theoretical explanations for unit nonresponse have been proposed to account for self-selection into surveys. Most of them aim to model the individual's process of deciding whether to agree to or decline to participate (e.g., Groves et al. 1992). According to the rational choice approach (e.g., Coleman/Fararo 1992), this decision can be regarded as a result of a cognitive evaluation of perceived costs and benefits of participation. Unfortunately, due to the non-participation itself, very little is known about nonrespondents (Giraldo/Zuanna 2006: 296), and their individual evaluations of costs and benefits are usually not measurable. Therefore, survey researchers aim to identify other variables from alternative data sources that might influence the individual's perception of costs and benefits of participation. For the nonresponse weighting adjustment in Sample M2, we use information on sampled households provided by the interviewer (see also, Olson 2006, Keeter et al. 2006, Abraham et al. 2006) and geo-coded information on the regional context (neighborhood, municipality, counties) of sampled households (see also, Johnson

[5]The actual maximum number of target population members per household was 5 (4 households); 13 households contained 4 target population members.

et al. 2006). Additionally, we make us of the rich IEB data available at the individual level.

Over recent decades, response rates in social science surveys in Germany have been constantly decreasing. To date, very few studies have achieved response rates above 40% (Schnell 2012: 164). For Study M2, the response rate – as the share of interviews completed in the reduced gross sample (see AAPOR 2011) – amounts to 28.8% and is therefore only marginally lower than that of the most recent SOEP samples, e.g., Sample J (33%), Sample K (35%) or Sample M1 (32%, see Kroh et al. 2014, Kroh et al. 2015). This minor decrease is probably due to the specific target population. Recent research has shown that migrant groups have lower response rates than non-migrants (e.g., Bethlehem et al. 2011: 64, Deding et al. 2008). From a rational choice perspective, it seems reasonable to assume that the average perceived costs of participation are higher in migrant groups. For example, difficulties in understanding the receiving country's language may result in higher nonresponse rates, as survey materials are not always provided in different languages. Furthermore, interviewers' language skills may be limited.[6] Additionally, a lack of detailed knowledge about local institutions (e.g., federal agencies) may lead to a sense of suspicion and lower valuations of interviewers' (and sponsors') reliability, which would also result in lower response rates. Another possible explanation may be confounding effects of various socio-demographic characteristics such as education and social status, and those associated with both migration and participation probabilities.

If respondents and nonrespondents show systematic differences in specific survey variables, estimates may be biased. From a statistical perspective, nonresponse bias is a function of the response rate and of differences between respondents and nonrespondents. This notion is based on the assumption of a "real" or "true" population parameter – a mean, for example, which consists of the mean value of participants $\bar{x}_r$ and non-participants $\bar{x}_n$. Depending on the prevalence of unit nonresponse and the amount of variation between the two groups, estimation results in a certain bias b (Bethlehem et al. 2011: 42):

$$b_{\bar{x}} = (\bar{x}_{response} - \bar{x}_{nonresponse}) \cdot \frac{n_{nonresponse}}{n_{total}} \tag{4}$$

The size of a nonresponse bias increases with decreasing response rates and/or increasing differences between participants and non-participants.[7]

[6] Therefore, additional materials might not be used even when they are necessary.

[7] Due to the fact that bias depends upon both the nonresponse rate and the difference between groups,

Addressing Nonresponse

Various statistical techniques addressing nonresponse adjustment have been proposed over the last few decades. Their common aim is to correct for differences between respondents and non-respondents that cause biased estimates of population parameters. Model-based weighting methods such as *propensity score adjustment* use observable differences between the selected gross sample and the surveyed net sample. For this purpose, information on both respondents *and* nonrespondents is used to calculate so-called "nonresponse weights". Using these nonresponse weights, the net sample distributions for various variables are weighted to meet the gross sample distributions for these variables (Kalton/-Flores-Cervantes 2003: 83). In other words, nonresponse weights attach more "importance" to observed units of analysis (e.g., households) if they have characteristics that are associated with nonresponse. In many cases, weighting procedures are accompanied by other techniques that focus on differences between marginal distributions of the net sample and known marginal distributions of the underlying target population (e.g., post-stratification, raking, GREG).[8]

For the IAB-SOEP Migration Study M2, the unknown participation probabilities were estimated using logistic regression and then transformed into propensity weights (Kim/Kim 2007: 501f). A binary variable for participation was constructed and used as the dependent variable. The calculated nonresponse weights can be combined with design weights to correct the parameter estimation of an underlying target population, namely, private households of recent immigrants in Germany in 2015.

Correlates of Nonresponse: Data Sources

Usually, individual information on nonrespondents is rare in social surveys, and obtaining such information requires a sophisticated and expensive nonresponse study. This is why

relatively small response rates do not automatically cause estimators to be biased: If respondents and nonrespondents do not systematically differ (with regard to specific survey variables), a bias will be small or even nonexistent. Hence, the response rate itself holds no information about the presence or amount of bias. Groves/Peytcheva (2008) provide a meta-analysis of the impact of nonresponse rates on nonresponse bias.

[8] Regardless of which technique is used to correct for self-selection into the survey, assumptions about the nature of nonresponse need to be made. Usually, scholars differentiate three mechanisms of missingness, or types of missing data: Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR) (see Rubin 1976). The majority of weighting techniques assume that data is at least MAR, i.e., nonresponse on a variable is related to other (observed) variables but not to values of the variable itself. In other words, some subgroups of the target population (with specific characteristics) are more or less likely to participate in a survey than others. However, when those differences in observed characteristics are controlled for, no systematic differences between respondents and nonrespondents (within groups) remain (Schafer 1997).

most studies rely on information on aggregated data levels such as (non)respondents' county, municipality, and neighborhood.

One of the key advantages of register-based samples – such as the sample of migrants in the IEB – is the availability of information on nonrespondents at an individual level. This allows us to rely on both individual and aggregated-level covariates in order to model nonresponse in the IAB-SOEP Migration Study M2.[9] A variety of variables from different data sources and at different levels of aggregation were used to model nonresponse in Study M2: Fieldwork Information (Address), Microm Data (Neighborhood), INKAR Database (County), Regional Information from the Federal Statistical Office (Municipality), as well as IEB data at an individual level. A table with all variables used in modeling nonresponse is presented at the end of this section.

Individual: IEB Using the IEB as a sampling frame allows us to rely on a variety of information at the individual level to model nonresponse. This includes an individual's country of origin, their age, the year in which they entered the IEB, their current/past employment status and vocational education, as well as whether a valid telephone number is available or not.

Addresses: Fieldwork Information During the address sampling process, SOEP interviewers collected information about sampled households and their environments. Those variables contain useful information about the living standard of sampled households. For instance, residents of more expensive single-family homes (possibly homeowners) are assumed to be more likely to participate than people living in rental units (Durrant/Steele 2009: 376), as wealthier individuals have been reported to show higher participation probabilities (see. Abraham et al. 2006: 693f). To model nonresponse, we used information on the type of home or building (e.g., rental unit in multi-story building vs. single-family home) as well as information about its condition.

Neighborhood: Microm Data Besides the information collected during fieldwork, we used a dataset provided by the private enterprise Microm GmbH. Microm data can

[9] Please note that the focus of our analysis is on the consistent estimation of response propensities. Hence, we are not primarily interested in a theoretical interpretation of effects (Spieß 2010: 123). A distinction also has to be made between variables available for individual households on the one hand, and aggregated spatial data linked to the sampled addresses of households on the other hand. Particularly in the latter case, causal interpretations of significant relationships are hampered by the problem of ecological fallacies (McGaw/Watson 1976: 134f). Therefore, caution is needed in interpretation.

be linked to the SOEP data and is available for use by guests and staff of the SOEP (see Goebel et al. 2007). It contains detailed local and regional information about the social structure and environment/neighborhoods of households in Germany. Variables are available at different levels of aggregation, ranging from the household-cell level (a few households grouped together), to market-cells (approx. 470 households per cell), to eight-digit postal code districts (approx. 500 households per district). Microm data therefore provide highly granular regional data for analysis. The variables used here mainly measure the social structures of households (e.g., family structure, educational status) as well as the economic situation of households (e.g., purchasing power).

County: INKAR Database The Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR) in Germany provides the database "Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa" (INKAR). INKAR contains useful information on regional economic issues (e.g., prices for building grounds, household income, welfare benefits) as well as inhabitant characteristics (e.g., educational data) of different regional entities. Variables are available at the county level and for NUTS-2[10] regions[11].

Municipality: Regional Information from the Federal Statistical Office As a joint project of the Federal Statistical Office with its state-(*Länder*)-level counterparts, the "Regionaldatenbank Deutschland" (regional database on Germany) provides useful data on different levels of aggregation. For the analysis of nonresponse in M2, we used variables compiled at the county level as well as at municipality level. The variables fall into three groups: data from the 2009 general parliamentary election (turnout, shares for different parties), age structure, and dwelling type.[12].

Additional Data Sources We used two additional variables on membership in the various political parties and level of civil engagement at the county level as provided in the "Deutscher Lernatlas", a comparative research publication on learning conditions and educational quality at the county level.[13]

[10]NUTS-2 is a statistical region used in cross-country comparison by European Union Statisticians.
[11]For additional information on variables and technical issues, see INKAR (2011) in the bibliography.
[12]For further information, see the web link under REGIONAL (2012) in the bibliography.
[13]Data can be downloaded freely without registration. See Lernatlas (2011) and `http://www.deutscher-lernatlas.de/deutscher-lernatlas/ergebnisse/` [visited the 30th January 2017].

Table 6: List of Variables used in Analysis of Nonresponse of Sample M2 – I

Variable	Source	Type	Values/ Range	level	Year
type of house	field information	ordinal (4 steps)	1= individual 4= high multi-story	household	2015
municipality size	field information	ordinal (6 steps)	1= <2k inh. 7= > 500k inh.	municipality	2015
business intensity (street)	Microm	ordinal (6 steps)	1 = accommodation only 6 = business only	street level	2015
household structure	Microm	ordinal (9 steps)	1= mainly single persons 9= mainly families with children	house cells	2015
children per household	Microm	ordinal (9 steps)	1= lowestvalue 9= highest value 6= average	house cells	2015
status (socio-economic)	Microm	ordinal (9 steps)	1= lowest status 9= highest status 5= average	house cells	2015
exclusive housing environment	Microm	binary	1=yes 0=no	house cells	2015
share of Turkish immigrants	Microm	metric	-	market cells	2015
share of eastern European immigrants	Microm	metric	-	market cells	2015
unemployment	Microm	ordinal (7 steps)	1= lowest 7= highest 4= national average	8-digit postal codes	2015
unemployment index (FRG = 100%)	Microm	metric	-	8-digit postal codes	2015
household purchasing power index (FRG = 100%)	Microm	metric	-	8-digit postal codes	2015
number of commercial operations	Microm	metric	-	house cells	2015
relocation balance	Microm	ordinal (9 steps)	1= strongly neg. balance 9= strongly pos. balance 5= balanced	house cells	2015
number of relocations	Microm	ordinal (9 steps)	1= strongly neg. balance 9= strongly pos. balance 5= balanced	house cells	2015
share of foreign households	Microm	ordinal (9 steps)	1= lowest share 9= highest share 7= average	house cells	2015
composition of households	Microm	ordinal (9 steps)	1= mostly single househ. 9= mostly families w. children 5= mixed	house cells	2015
share of households refusing admail	Microm	ordinal (9 steps)		house cells	2015

Table 7: List of Variables used in Analysis of Nonresponse of Sample M2 – II

Variable	Source	Type	Values/ Range	level	Year
prices for building grounds	Inkar	metric	in EUR/m^2	county	2013
share of migrants	Inkar	metric	-	county	2013
GDP/capita	Inkar	metric	in 1000 Euros	county	2013
unemployed migrants	Inkar	metric	-	county	2015
unemployment rate	Inkar	metric	-	county	2015
employment rate	Inkar	metric	-	county	2015
unemployment rate among migrants	Inkar	metric	-	county	2015
employment rate among migrants	Inkar	metric	-	county	2015
share of females among migrants	Inkar	metric	-	county	2015
share of long-term among unemployed	Inkar	metric	-	county	2015
average unemployment benefits	Inkar	metric	in EUR	county	2015
average unemployment benefits (males)	Inkar	metric	in EUR	county	2015
average unemployment benefits (females)	Inkar	metric	in EUR	county	2015
share of single- and two familie houses	Inkar	metric	-	county	2015
share of employees without formal qualification	Inkar	metric	-	county	2015
demographic development (2006-2011, change in inhabitants)	Inkar	metric	in %	county	2015
share of school students	Inkar	metric	-	county	2015
share of university students	Inkar	metric	-	county	2015
share of students leaving school with higher education entrance qualification	Inkar	metric	-	county	2015
average compensation of employees	Inkar	metric	in EUR	county	2015
settlement density	Inkar	metric	-	county	2015
business tax per capita	Inkar	metric	-	county	2015
spatial planning region	Inkar	ordinal	-	county	2015
share of unemployment and social benefits (SGB II)	Inkar	metric	-	county	2015
share of employees in the tertiary sector	Inkar	metric	-	county	2015
share of employees in the primary sector	Inkar	metric	-	county	2015
type of county	Inkar	binary	1=county 2=non-county municipality	county	2015

Table 8: List of Variables used in Analysis of Nonresponse of Sample M2 – III

Variable	Source	Type	Values/ Range	level	Year
electoral turnout in 2009 general election	Statistics Office	metric	-	municipality	2009
vote share for SPD	Statistics Office	metric	-	municipality	2009
vote share for CDU/CSU	Statistics Office	metric	-	municipality	2009
vote share for FDP	Statistics Office	metric	-	municipality	2009
vote share for *Alliance '90/ The Greens*	Statistics Office	metric	-	municipality	2009
vote share for *The Left*	Statistics Office	metric	-	municipality	2009
vote share for small parties	Statistics Office	metric	-	municipality	2009
share of small flats (1-2 rooms)	Statistical Office	metric	-	municipality	2011
share of big flats (6+ rooms)	Statistical Office	metric	-	municipality	2011
share of age bracket 18-25	Statistical Office	metric	-	municipality	2011
share of age bracket 25-35	Statistical Office	metric	-	municipality	2011
share of age bracket 35-45	Statistical Office	metric	-	municipality	2011
share of age bracket 45-55	Statistical Office	metric	-	municipality	2011
share of age bracket 55-65	Statistical Office	metric	-	municipality	2011
share of elderly (65+)	Statistical Office	metric	-	municipality	2011
share of people active in non-profit org.	Lernatlas	metric	-	county	2008
quota of party members	Lernatlas	metric	-	county	2009
highest professional training	IEB	categorical	no professional training (1), professional training (2), currently in training (3), university degree (4), rest/missing (5)	individual	2013
professional training spell exists	IEB	categorical	0 = no, 1 = yes	individual	2013
country of origin	IEB	categorical	Turkish, Italian, Spanish, Greek, (...)	household	2013
completeness of telephone number	IEB	categorical	0 = complete, 1 = incomplete	individual	2013
year of first IEB data entry	IEB	metric	-	individual	2013
age category	IEB	categorical	≤25y, 26-40y, >40y	individual	2013
first employment status	IEB	ordinal	employed w/o. welfare benefits (1), unemployed w. welfare benefits (2), employed w. welfare benefits (3), unemployed w/o. welfare benefits, e.g. FEA measure (4)	individual	2013
current employment status	IEB	ordinal	(same as first employm. stat.)	individual	2013

Multiple Imputation and Data Coding

For analysis of nonresponse and the generation of weights, it is necessary to have complete observations: otherwise, observations will be omitted from the regression model, and weights cannot be estimated for those observations. Some of the selected variables contained missing values. In the majority of cases, all values for all variables for one source were missing for a spatial unit (county, municipality). However, none of the households yield complete missings. In other words, missing values do not cluster for one particular set of households. Furthermore, the overall share of missings was very low. No missing values were found in the IEB data, the INKAR data, or the Federal Statistical Office data. Only a few missing values were present in the fieldwork data (less than 0.4%) and the Microm data (less than 0.5%).

Missing observations in the dataset were imputed by means of the "multiple imputation by chained equations" method (Royston 2009). To account for the imputation uncertainty that this procedure implies, ten different predictions for the missing values were calculated (White et al. 2011: 378). Furthermore, the entire statistical procedure was implemented ten times with different starting values (Horton/Lipsitz 2001: 248). As a result, ten different complete datasets were available for analysis, thereby making it possible to take the uncertainty of the imputation into account through appropriate statistical procedures (White et al. 2011: 377).

Some of the imputed variables were then transformed for the remainder of the analysis. Continuous variables were categorized, resulting in three distinct categories for most of the variables; in general, the middle category served as a reference group in the regression analysis. Ordinal indicators with several categories (e.g., socio-economic status) were recoded to two or three categories in order to produce more qualitatively distinct groups. Using categorized variables and their respective binary indicators in regression analysis has several advantages. Non-linear effects are controlled for because individual parameters are estimated for each group. Also, this categorization prevents the estimation of extreme probabilities very close to zero or one because of single outliers on a variable. This is necessary in order not to inflate the estimated weights inappropriately (Spieß 2010: 122; Valliant/Dever 2011: 116). Finally, this makes interpretation and comparison of coefficients more straightforward.

Modeling Nonresponse and M2 Nonresponse Weights

In order to model the households' nonresponse propensities, logit regressions were performed for different combinations of covariates using statistical routines that accounted for imputation uncertainty. Additionally, we used robust standard errors in order to account for possible heteroscedasticity and non-independent observations within sample points (see White 1980, Spieß 2010). A sample point identifier variable was used as a cluster variable in each of the estimated models.[14]

All 5,950 households in the reduced gross sample sample were included in each model. The initial full model included all the available covariates displayed in Tables 6, 7 and 8. The second reduced model was estimated using only those variables that exert a significant effect ($\alpha = 5\%$ level).

Figure 5 displays coefficients and their respective 95% confidence intervals in the reduced model. Only a small fraction of the initial covariates used in the full model reaches statistical significance. Hence, the reduced model is much more parsimonious. Overall, these results can be understood in a positive sense, especially regarding the quality of sampling: A wide variety of variables from various data sources have been tested for their influence on response propensities, and only a small fraction of them reach significance. The final model explains only about 7% of the overall variance in participation propensities. The results suggest that participation across groups is indeed determined largely by chance, and that in many respects, respondents and nonrespondents may not differ from each other systematically.

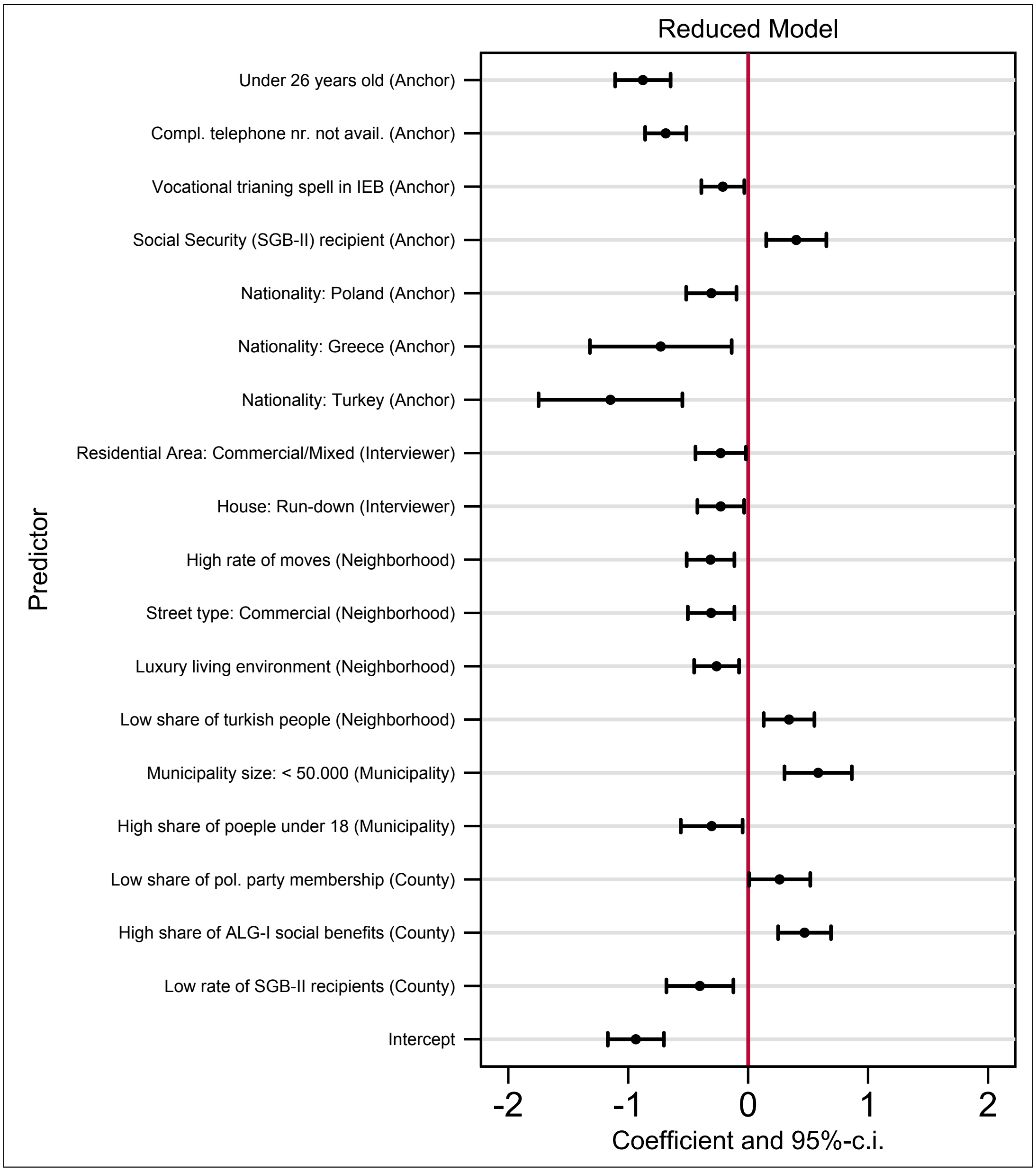
However, turning to the individual coefficients displayed in Figure 5, some systematic differences remain.

Age: In line with past research (e.g., Keeter et al. 2006), households in which the anchor respondent was relatively young were less likely to participate. Accordingly, households in municipalities with a comparatively high share of residents under 18 were less likely to participate.

Migration: While the majority of covariates capturing migration background do not show significant effects, immigrants from Poland, Greece, and Turkey were less likely to participate in the survey than other groups of migrants. In line with this, higher

[14] Not doing so would run the risk of estimating standard errors too high or too low, the latter being potentially more detrimental to valid inferences.

Figure 5: Coefficients and Confidence Intervals for the Estimated Reduced Model



Note: The dependent variable was coded 1 for participation and 0 for non-participation. Number of observations in the model n = 1,096

response rates were observed in neighborhoods associated with a low share of Turkish people. In conclusion, the results are quite positive, as only a few migrant subgroups show statistically significant differences in their response propensities.

Interviewer Observations: Turning to the interviewer observations collected during fieldwork, households in commercial/mixed residential areas were less likely to participate than those in areas with mainly private households. Moreover, and in line with past analyses in the context of the SOEP (e.g., Kroh et al. 2015), response rates were lower for households in buildings that appear somewhat run-down.

Employment: Households whose anchor respondents had a vocational training spell in the IEB were less likely to participate. In contrast, households whose anchors currently receive social security benefits (SGB II) were more likely to participate. At the county level, higher response rates were obtained in areas with a comparatively high share of ALG-I social benefit recipients. In line with this, response rates were lower in areas with a low share of SGB-II recipients.

Neighborhood: Households in neighborhoods with high housing market or occupancy turnover, with higher numbers of commercial streets, and in neighborhoods with luxury housing were less likely to participate.

Municipality: A municipality's size is associated with response propensities, with higher participation rates being found in smaller municipalities and cities.

Further variables: Not surprisingly, households (anchors) for whom no complete telephone number was available had a lower probability to participate – most likely because it was simply more difficult to contact the households. Contrary to our expectations, the share of party membership in an area is negatively related to participation probabilities. However, comparable effects were observed in the nonresponse adjustment of Sample K data (Kroh et al. 2014) and Sample M1 (Kroh et al. 2015), in which areas with a high election turnout showed lower participation rates as well.

The reduced model (as displayed in Figure 5) was used to predict household participation probabilities. Table 9 displays a comparison between actual response rates and mean estimated response probabilities by sample point. The correlation between the two variables ($\rho_{(obs;est)} = 0.55$) reveals that the prediction of response rates at the sample point level is reasonably good.

The nonresponse weights for further analysis are calculated as inverse response probabilities.

Table 10 displays characteristics of the raw estimated nonresponse weights. Estimated weights still cover a wide range, even after having been reduced in variation through regression design, and relatively extreme differences are found between the value for the 90th-percentile (9.36) and the corresponding maximum (47.37), indicating the presence of

Table 9: Comparison of Estimated and Actual Response Rates by Sample Point

Response Rates	Min.	*p25*	Median	*p75*	Max.
observed (x_{obs})	00.00	11.11	18.18	23.91	47.77
estimated (x_{est})	08.88	14.54	18.26	21.96	33.55

Note: At the sample point level the two variables correlate with $\rho_{(obs;est)} = 0.55$

Table 10: Characteristics of Raw Estimated Nonresponse Weights

Min.	*p10*	*p25*	Median	*p75*	*p90*	Max.	Mean	SD
1.56	2.50	3.21	4.29	6.19	9.36	47.37	5.35	3.50

outliers in the data. As mentioned before, wide variation in weights can be harmful as variation in estimates may increase substantially (van Goor/Stuiver 1998). The trimming of weights (e.g., Peytchev et al. 2011) to be more equally distributed seems reasonable to counteract possibly undesired results of the use of weights in estimation procedures. However, trimming may result in a slight loss of efficiency. There are various possible trimming approaches, and the selection of a specific procedure is best based on the ratio of estimated weights.[15] For the given sample, we aimed for household weights not to exceed 2.0 times the weight's mean:

$$w_{trimmed} \begin{cases} 2.0 * \bar{x}_w + 0.5 * ln(w - 2.0 * \bar{x}_w + 1), & \text{if } w > 2.0 * \bar{x}_w \\ w, & \text{otherwise} \end{cases}$$

Table 11: Characteristics of Trimmed Estimated Nonresponse Weights

Min.	*p10*	*p25*	Median	*p75*	*p90*	Max.	Mean	SD
1.56	2.50	3.21	4.29	6.19	9.36	12.51	5.11	2.59

The trimming procedure affected a total of 86 households with a weight above approximately 10.7. Hence, trimming succeeded, as variation in weights was decreased while only a few weights were adjusted, therefore keeping the loss of efficiency at a minimum level. Table 11

[15] Theoretical guidance on appropriate figures is scarce; for an example showing the difficulties of trimming, see van Goor/Stuiver (1998).

displays characteristics of the final trimmed nonresponse weights for all 1,096 participating households in Study M2.

5.3 Post-stratification and Raking

In addition to the reported nonresponse weighting adjustment, household weights were corrected using *post-stratification* and *raking* so that the sample M2 meets known cell distributions or marginal totals. In standard post-stratification or cell weighting procedures, weights are adjusted so that given sample totals fit the known cell distributions from the underlying target population "on a cell-by-cell basis" (Kalton/Flores-Cervantes 2003). Raking – also referred to as "iterative proportional fitting" (Deming/Stephan 1940) – is a special case of post-stratification and is used "when poststrata are formed using more than one variable, but only the marginal population totals are known" (Lohr 2010: 344).

The post-stratification procedure completes the three-step process of calculating what we refer to as "first-wave weights" for new SOEP samples, thus constituting a combination of design weights, nonresponse weights, and post-stratification procedures. First-wave weights are available for all SOEP subsamples A to M2. They are of special importance as they serve as base weights for the calculation of both longitudinal weights and cross-sectional weights from wave 2 onwards.

Final Study M2 first-wave household weights are stored in the variable `BFHHRFM2`. First-wave weights at individual (person) level are stored in the variable `BFPHRFM2` respectively.[16]

Total marginal values derived from the German Microcensus were used in the raking procedure in which weights were corrected to meet benchmarks of the underlying target population in 2015. The Microcensus is conducted by the Federal Statistical Office of Germany (FSO) and is a one-percent sample of the German resident population.

For post-stratification at the household level, target population households were identified both in Sample M2 as well as in the 2015 Microcensus. Households were defined as belonging to the target population if at least one household member had immigrated since 2009. Sampled households needed to be clearly classified in terms of immigration year and country of origin. As households often contain multiple target population members, decisions had to be made on how to unambiguously classify households. Here, we used the

[16]Researchers interested in using design weights alone are recommended to use the wave-specific variable `design` stored in the dataset `design`. The variable only contains the inverse probability of sample selection.

concept of household representatives.[17] The household representative selection strategy was performed both for M2 data and Microcensus data.

Table 12: Population Characteristics Used in the SOEP Raking Procedure at Household Level

Variable	**Values**
Year of Immigration & Nationality	2009-2011, Germany 2009-2011, Poland 2009-2011, Romania/Bulgaria 2009-2011, Italy/Portugal/Spain/Greece 2009-2011, Western Europe (Rest) 2009-2011, Eastern Europe (Rest) 2009-2011, Islamic/Arabic Countries 2009-2011, World (Rest) 2012-2013, Germany 2012-2013, Poland 2012-2013, Romania/Bulgaria 2012-2013, Italy/Portugal/Spain/Greece 2012-2013, Western Europe (Rest) 2012-2013, Eastern Europe (Rest) 2012-2013, Islamic/Arabic Countries 2012-2013, World (Rest)
Household Size	1 / 2 / 3 / 4 / 5+
Federal States *(Länder)*	Baden-Württemberg, Berlin/Brandenburg, Bremen/Lower Saxony, Mecklenburg-Western Pomerania/Thuringia/Saxony/Saxony-Anhalt, North Rhine-Westphalia, Hesse, Rhineland-Palatinate/Saarland, Hamburg/Schleswig-Holstein
Municipality Size	$<$ 20.000 / 20.000-100.000 / 100.000-500.000 / $>$ 500.000
Region in Germany	North, East, South, West

Note: Population characteristics were derived from the Microcensus 2015.

Table 12 lists the characteristics at the household level that were used in the raking process in Sample M2. Weights were adjusted with respect to a household's year of immigration and country of origin as well as household size, federal states (*Länder*), municipality size, and region in Germany.

Subsequent to the raking at household level, Study M2 data were additionally post-stratified

[17]Household representatives pass their individual characteristics on to the household as a whole. Rules were established to select an individual as a representative. First, IEB anchors were given preference over other household members. Second, if there were multiple immigrants in a household, we chose the household member with the most recent year of immigration. In the case of duplicate immigration years, female household members were given preference over male members. In the next stage, we considered the current employment status, giving preference to full-time employment over part-time employment, and part-time employment over unemployment. Finally, in the rare case of multiple potential household representatives at this stage of the selection procedure, a household member was randomly selected.

at the individual level to generate individual first-wave weights. Again, information on population totals was derived from the 2015 German Microcensus. Raking at the individual level is based on the combined first-wave household weights (the result of the raking procedure at the household level) and uses information about an individual's year of immigration & country of origin as well as gender and age (see table 13).

Table 13: Population Characteristics Used in the SOEP Raking Procedure at the Individual Level

Variable	**Values**
Year of Immigration & Nationality	2009-2011, Germany 2009-2011, Poland 2009-2011, Romania/Bulgaria 2009-2011, Italy/Portugal/Spain/Greece 2009-2011, Western Europe (Rest) 2009-2011, Eastern Europe (Rest) 2009-2011, Islamic/Arabic Countries 2009-2011, World (Rest) 2012-2013, Germany 2012-2013, Poland 2012-2013, Romania/Bulgaria 2012-2013, Italy/Portugal/Spain/Greece 2012-2013, Western Europe (Rest) 2012-2013, Eastern Europe (Rest) 2012-2013, Islamic/Arabic Countries 2012-2013, World (Rest)
Gender	Male / Female
Age	0-4 / 5-9 / 10-14 / 15-19 / 20-24 / 25-29 / 30-34 / 35-39 / 40-44 / 45-49 / 50-54 / 55-59 / 60-64 / 65+

Note: Population characteristics were derived from the Microcensus 2015.

6 Characteristics of Cross-Sectional Weights

As described above, cross-sectional weighting of first-wave SOEP data is a three-stage process: design weighting (1), nonresponse adjustment (2), and post-stratification (3). A combination of all three stages results in what we refer to as “first-wave weights”, which are available for all SOEP subsamples A to M2.

Table 14 displays characteristics of household weights in each of the three weighting stages in Study M2. The initial design weights of M2 were grossed up so that the number of households in the sample meets the total number of 979,099 households in the underlying target population (see section 5.1).

Table 14: Characteristics of Weights During the Weighting Process

		Quantiles								
	Min.	10%	25%	50%	75%	90%	Max.	Mean	SD	n
Design Weight	50	85	150	169	179	282	357	175	67	1,096
Design*Nonresponse	129	339	486	714	1,097	1,810	3,964	894	601	1,096
Combined First-Wave Weight	51	202	367	681	1,162	2,014	3,390	899	744	1,096

Variance in design weights is due to the disproportional sampling of anchor respondents in accordance to their country of origin (see chapter 5.1). In the second stage, design weights were combined (multiplied) with nonresponse weights. As can be seen from the quantiles and standard deviations, variation in weights increased as expected. Finally, the product of design and nonresponse weights was post-stratified, resulting in the final combined first-wave weights for Sample M2. The raking procedure introduced additional variation into the weights, but only to a limited extent.

The distribution of the weights at all three steps is also displayed in Figure 6. As can be seen, weights are fairly right-skewed, both pre-raking (orange line) and post-raking (blue line).

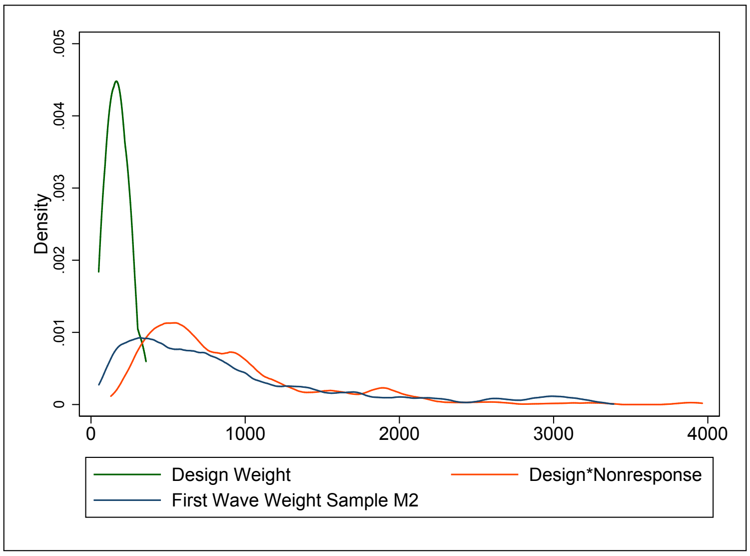


Figure 6: Distribution of Study M2 Weights at all Three Weighting-Steps

A Description of Variables and Expected Effects

Microm Data

Business intensity (street): Similar to the variable from the fieldwork information but on another level, this variable captures whether a street appears overall to be primarily residential or dominated by services, manufacturing, or business. It has six ascending steps.

Mean age of heads of households: This variable captures the mean age of household heads in eight steps, ranging from "under 35 yrs." to "over 65 yrs.". It is compiled at the house cell level. Older people are known to participate less frequently in surveys (Johnson et al. 2006: 711); therefore a negative association with response rates is expected.

Share of families/household structure: This variable displays the dominant structure of households based on the number of households members on the house cell level. Nine steps range from "mainly single-person households" to "almost exclusively families w/ children". Families with children are expected to be much easier to reach, and therefore, participation probabilities should be higher in cells dominated by families (Keeter et al. 2006: 768).

Children per household ratio: Similar to the last variable, the children per household ratio in nine steps on the house cell level should be associated with higher response probabilities, since families with children are more likely to participate (Olson 2006: 746).

Socio-Economic Status: This variable is a composite index aggregating education and income of a house cell's inhabitants. It is coded in nine ascending steps. People with a higher status are often believed to be more likely to participate because of having more often experienced positive outcomes from other processes of social exchange (education, career, etc.) and therefore perceiving more potential benefit from participation (Durrant/Steele 2009: 375). Some researchers speak of a "middle class bias" (e.g., Goyder et al. 2002). However, "elites" are sometimes believed to have lower response probabilities. Either way, accounting for status seems necessary.

Share of college graduates: Nine steps ranging from "below 2%" to "over 35%" cover the share of college graduates in the population of a street. Streets with higher share of college graduates should display higher participation probabilities (Abraham et al. 2006: 694; Singer et al. 1999: 258).

Purchasing power: This index captures a market cell's (approximately 470 households) purchasing power in relation to the national average (=100). Purchasing power serves as a proxy for wealth; therefore, a positive effect on response is expected (see above).

Turnover in housing (mobility): This variable captures the turnover in accommodation/housing in nine ascending steps, with 5 being the national average. High turnover in a market cell should be associated with higher mobility. This might make people less likely to participate in multi-wave surveys, especially since giving notice of an upcoming address change may increase the perceived cost of participation. Furthermore, existing literature has shown that moving may make household members feel less integrated into new communities and reduce their participation (Durrant/Steele 2009: 377). While the findings are mixed so far, controlling for mobility seems reasonable.

Balance in housing turnover (mobility): On the market cell level, this variable indicates whether turnover in housing results in negative balances (low values) or positive balances (high values). People in potentially less attractive cells with negative balances (that are probably perceived as "disadvantaged areas") can be expected to be more likely to plan to move away, and therefore less likely to participate in repeated surveys (see above).

Unemployment: Using the finer distinction provided by the German eight-digit postal codes (approx. 500 households per cell), unemployment is captured in this variable in seven steps in relation to the national average. It serves as a proxy for wealth and should therefore be negatively related to response probabilities. Unemployed people are often less integrated into economic life, and may be what Johnson et al. (2006) label "disadvantaged" and others call "isolated" (Durrant/Steele 2009: 375) However, non-contact may be less of a problem with unemployed people, as is the case with older people.

INKAR

Prices of Undeveloped Land: Average prices for property per square meter on the county level are included as an additional indicator for the wealth and attractiveness of a region and should be positively related to response probabilities.

Average household income: The average disposable household income per person at the county level as an indicator for wealth is included. Positive effects on probabilities are expected.

GDP per capita: Similarly, the GDP per capita ratio is included. It is coded in thousand euros.

Welfare benefits to cover rent: County-level data on the monetary value of welfare benefits provided to cover rent and housing expenses capture the cost of living. Higher values per person should indicate high costs of living, and wealthier or more attractive areas. Thus, higher values should be accompanied by higher probabilities of response.

Share of residents entitled to welfare benefits: The higher the share of people entitled to welfare benefits in a county, the less wealthier this county should be. Therefore, a negative effect on response probabilities is assumed.

Medical doctors per inhabitants ratio: Drawing on the concept of "disadvantaged areas" (Johnson et al. 2006: 707f), higher numbers of doctors in a (more privileged) county should have a positive association with response probabilities. It is coded in doctors per 100k inhabitants.

Share of high school and college graduates: It has been shown several times in the literature that highly educated individuals have higher response probabilities, as education is often accompanied by a greater sense of civic obligation (e.g. Abraham et al. 2006: 694; Durrant/Steele 2009: 372). High school education in particular should play a vital role, and should produce the greatest differences between individuals because of its gatekeeper function for college education. The variables were compiled at the NUTS2-level of EU regions.

Regional Information from the Federal Statistical Office

Electoral turnout: Electoral turnout for the 2009 general parliamentary election was calculated on the community level to capture general participation tendencies. High rates of turnout should relate positively to the affinity to participate in surveys. As has been shown by Keeter et al. (2006: 768) using split-ballot surveys in the U.S., people who are harder to sample are less frequently registered to vote. On a theoretical level, participation in elections may be related to the same construct as participation in surveys (for instance, civic obligation).

Percentage of votes for dominant and smaller parties: As Keeter et al. (2006: 768) demonstrate, people with lower response probabilities tend to vote for more ideologically moderate parties. The percentage of votes for the two largest German parties (SPD; CDU/CSU), which are relatively centrist, is included to test for this effect. Moreover, the share of votes for small parties (falling below the 5% threshold) is included to account for the possible opposite effect. In addition, the modeling of nonresponse in Study M2 relies on information on vote shares for other parties generally represented in the German *Bundestag*, such as The Left and Alliance '90/The Greens. For example, we expect higher nonresponse rates in areas with a high share of votes for The Greens, as such households tend to be more difficult to contact (Schnell 2012: 161).

Share of small and very large rental housing units: The share of differently sized rental housing units in current overall number of rental housing units available are included in the model as a proxy for household structures. Small units are commonly inhabited by single-person households (which are more difficult to reach) and larger units are more often inhabited by families, which are easier to reach and more prone to participation (Durrant/Steele 2009: 372).

Age structure: Shares of different age groups (7 variables from "18 to 25" in ten-year steps to "elderly (65+)") as a percentage for 2010 (community level) are included as controls. In past research, older people have been shown to be less likely to participate Keeter et al. (2006: 765). Furthermore, age is an important predictor in many social sciences research settings, and controlling for it in weight generation seems reasonable.

Integrated Employment Biographies - IEB

Country of origin: Generally, it has been shown that migrant groups differ in their responses rates (e.g., Babka von Gostomski/Pupeter 2008).

Years since first IEB data entry: We expected individuals who only recently immigrated to be less likely to participate for two main reasons: First, recent immigrants may move more frequently as it usually takes some time to find a relatively permanent residence in Germany. Second, recent immigrants may be more skeptical about the subject matter and institution conducting the survey, as they usually have less information about German institutions and agencies.

Year of birth and gender: Past research has shown that older people are less likely to participate in surveys (Johnson et al. 2006). Furthermore, it has been shown that males are more likely to decline participation (Smith 1983).

Household size: Single households were expected to be less likely to participate than multiple-person households (Groves/Couper 1998).

Further Variables

Share of people active in non-profit organizations: The share of people active in non-profit organizations (sports clubs, churches, community service, etc.) captures the general affinity for participation and possibly a sense of civic obligation. People with a high affinity for participation may not need high incentives to participate (Durrant/Steele 2009: 378), therefore participation in a survey should be more likely in this group. This variable was compiled at the county level in 2009[18] (Groves et al. 2000: 302f).

Percentage of political party members in population: The percentage of members of any political party in the total adult population should also be related to general participation affinity. Therefore, positive effects are anticipated.

[18] Data was missing for some counties, but for unknown reasons. However, it was extracted manually from Engagementatlas (2009).

References

AAPOR (2011). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. Revised 2011.*

Abraham, K. G./A. Maitland/S. M. Bianchi (2006). “Nonresponse in the American Time Use Survey”. In: *Public Opinion Quarterly* 70(5), pp. 676–703.

Babka von Gostomski, C./M. Pupeter (2008). “Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters”. In: *Methoden – Daten – Analysen* 2(2), pp. 149–177.

Bethlehem, J./F. Cobben/B. Schouten (2011). *Handbook of Nonresponse in Household Surveys*. Wiley Handbooks in Survey Methodology. New York: John Wiley & Sons.

Brücker, H./M. Kroh/S. Bartsch/J. Goebel/S. Kühne/E. Liebau/P. Trübswetter/I. Tucci/-J. Schupp (2014). *The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Contents.* SOEP Survey Papers No. 216, Series C – Data Documentations. DIW Berlin.

Coleman, J./T. Fararo (1992). *Rational choice theory.* New York: Sage.

Deding, M./T. Fridberg/V. Jakobsen (2008). “Non-response in a survey among immigrants in Denmark”. In: *Survey Research Methods* 2(3), pp. 107–121.

Deming, W./F. Stephan (1940). “On a least squares Adjustment of a ssample frequency table when the expected marginal ttotal are known”. In: *Journal of the American Statistical Association* 35, pp. 615–630.

Diehl, C. (2007). *Materialband und Endbericht zur Neuzuwandererbefragung - Pilotstudie: 1. und 2. Welle.* Materialien zur Bevölkerungswissenschaft, No. 122. Wiesbaden: Bundesinstitut für Bevölkerungsforschung (BIB).

Durrant, G. B./F. Steele (2009). “Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence from Six UK Government Surveys”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 172(2), pp. 361–381.

Engagementatlas (2009). *Engagement Atlas 09 - Daten. Hintergründe. Volkswirtschaftlicher Nutzen.* http://zukunftsfonds.generali-deutschland.de/online/CMSDeliveryPortlets/servlet/contentblob/314342/309602/3/data/Engagementatlas%25202009_PDF_Version.pdf. (Visited on 05/22/2012).

Giraldo, A./G. D. Zuanna (2006). "Investigation of a Unit Non-Response Adjustment Procedure: The Case of the Urban Fertility Survey, Italy, 2001-2002". In: *Population (English Edition)* 61(3), pp. 295–307.

Goebel, J./C. K. Spieß/N. R. Witte/S. Gerstenberg (2007). *Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz.* SOEP Data Documentation No. 26. DIW Berlin.

Goyder, J./K. Warriner/S. Miller (2002). "Evaluation Socio-economic Status (SES) Bias in Survey Nonresponse". In: *Journal of Official Statistics* 18(1), pp. 1–11.

Groves, R. M./M. P. Couper (1998). *Nonresponse in Household Interview Surveys.* New York: John Wiley & Sons.

Groves, R. M./E. Peytcheva (2008). "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis". In: *Public Opinion Quarterly* 72(2), pp. 167–189.

Groves, R. M./R. B. Cialdini/M. P. Couper (1992). "Understanding The Decision to Participate in a Survey". In: *The Public Opinion Quarterly* 56(4), pp. 475–495.

Groves, R. M./E. Singer/A. Corning (2000). "Leverage-Saliency Theory of Survey Participation: Description and an Illustration". In: *The Public Opinion Quarterly* 64(3), pp. 299–308.

Heckman, J. (1979). "Sample Selection Bias as a Specification Error". In: *Econometrica* 47(1), pp. 153–161.

Horton, N. J./S. R. Lipsitz (2001). "Multiple Imputation in Practice". In: *The American Statistician* 55(3), pp. 244–254.

Horvitz, D. G./D. J. Thompson (1952). "A Generalization of Sampling Without Replacement From a Finite Universe". In: *Journal of the American Statistical Association* 47(260), pp. 663–685.

Humpert, A./K. Schneiderheinze (2013). *Linguistic analysis of personal names based on th "Onomastic process". www.stichproben.de/herunterladen/onomastic_process_HS_GbR.pdf Letzter Abruf: 20.02.2014.*

INKAR (2011). *Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa.* URL: http://www.bbr.bund.de/nn_21272/BBSR/DE/Veroeffentlichungen/INKAR/inkar__node.html (visited on 05/22/2012).

Jacobebbinghaus, P./S. Seth (2007). "The German Integrated Employment Biographies Sample IEBS". In: *Schmollers Jahrbuch* 127(2), pp. 335–342.

Johnson, T. P./Y. I. Cho/R. T. Campbell/A. L. Holbrook (2006). "Using Community-Level Correlates to Evaluate Nonresponse Effects in a Telephone Survey". In: *Public Opinion Quarterly* 70(5), pp. 704–719.

Kalton, G./I. Flores-Cervantes (2003). "Weighting Methods". In: *Journal of Official Statistics* 19(2), pp. 81–97.

Keeter, S./C. Kennedy/M. Dimock/J. Best/P. Craighill (2006). "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey". In: *Public Opinion Quarterly* 70(5), pp. 759–779.

Kim, J. K./J. J. Kim (2007). "Nonresponse weighting adjustment using estimated response probability". In: *Canadian Journal of Statistics* 35(4), pp. 501–514.

Kroh, M./K. Käppner/S. Kühne (2014). *Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Berlin* SOEP Survey Papers No. 260, Series C – Data Documentations. DIW Berlin.

Kroh, M./S. Kühne/J. Goebel/F. Preu (2015). *The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment.* SOEP Survey Papers No. 271, Series C – Data Documentations. DIW Berlin.

Kroh, M./S. Kühne/R. Siegers (2016). *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2015).* SOEP Survey Papers No. 408, Berlin C - Data Documentation. DIW Berlin.

Lernatlas (2011). *Deutscher Lernatlas.* URL: http://www.deutscher-lernatlas.de/de/ergebnisse/profile.html (visited on 05/22/2012).

Lipps, O./K. Kissau (2012). "Nonresponse in an Individual Register Sample Telephone Survey in Lucerne/Switzerland". In: *Telephone Surveys in Europe: Research and Practice.* Ed. by M. Häder/S. Häder/M. Kühne. Heidelberg: Springer.

Lohr, S. L. (2010). *Sampling: Design and Analysis*. 2nd ed. Boston, MA: Brooks/Cole.

McGaw, D./G. Watson (1976). *Political and social inquiry*. New York: Wiley.

Oberschachtsiek, D./P. Scioch/C. Seysen/J. Heining (2009). *Integrated Employment Biographies Sample (IEBS). Handbook for the IEBS in the 2008 version*. FDZ Datenreport, Documentation of labour market data, No. 03/2009 (EN). Bundesagentur für Arbeit. Nürnberg.

Olson, K. (2006). "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias". In: *Public Opinion Quarterly* 70(5), pp. 737–758.

Peytchev, A./L. R. Carley-Baxter/M. C. Black (2011). "Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse". In: *Sociological Methods & Research* 40(1), pp. 138–168.

REGIONAL (2012). *Regionaldatenbank Deutschland*. URL: `https://www.regionalstatistik.de/genesis/online;jsessionid=3798563B386EC36254C9997A78F78D57?Menu=Impressum` (visited on 05/22/2012).

Rosenbaum, P./D. Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70, pp. 41–55.

Royston, P. (2009). "Multiple imputation of missing values: Further update of ice, wich an emphasis on categorial variables". In: *The Stata Journal* 9(3), pp. 466–477.

Rubin, D. B. (1976). "Inference and Missing Data". In: *Biometrika* 63(3), pp. 581–592.

Salentin, K. (1999). "Die Stichprobenziehung bei Zuwandererbefragungen". In: *ZUMA-Nachrichten* 45(23), pp. 115–135.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Schnell, R. (2012). *Survey-Interviews. Methoden standardisierter Befragungen*. Berlin: Springer.

Singer, E./R. M. Groves/A. D. Corning (1999). "Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation". English. In: *The Public Opinion Quarterly* 63(2), pp. 251–260.

Smith, F. (1983). “On the Validity of IInference from Non-Random Samples”. In: *Journal of the Royal Statistical Society, Series A* 146, pp. 394–403.

Spieß, M. (2010). “Der Umgang mit fehlenden Werten”. In: *Handbuch der sozialwissenschaftlichen Datenanalyse*. Ed. by C. Wolf/H. Best. VS Verlag für Sozialwissenschaften, pp. 117–142.

Valliant, R./J. A. Dever (2011). “Estimating Propensity Adjustments for Volunteer Web Surveys”. In: *Sociological Methods & Research* 40(1), pp. 105–137.

Van Goor, H./B. Stuiver (1998). “Can Weighting Compensate for Nonresponse Bias in a Dependent Variable? An Evaluation of Weighting Methods to Correct for Substantive Bias in a Mail Survey among Dutch Municipalities”. In: *Social Science Research* 27(4), pp. 481 –499.

White, H. (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”. In: *Econometrica* 48(4), pp. 817–838.

White, I. R./P. Royston/A. M. Wood (2011). “Multiple imputation using chained equations: Issues and guidance for practice”. In: *Statistics in Medicine* 30(4), pp. 377–399.