SOEP The German Socio-Economic Panel study

291

SOEP Survey Papers

Series C – Data Documentation

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

The Request for Record Linkage in the IAB-SOEP Migration Sample

Philipp Simon Eisnecker, Klaudia Erhardt, Martin Kroh, Parvati Trübswetter

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:

Philipp Simon Eisnecker, Klaudia Erhardt, Martin Kroh, Parvati Trübswetter. 2017. The Request for Record Linkage in the IAB-SOEP Migration Sample. SOEP Survey Papers 291: 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 Request for Record Linkage in the IAB-SOEP Migration Sample

Philipp Simon Eisnecker, Klaudia Erhardt, Martin Kroh, Parvati Trübswetter

Content

Introduction 2

1 The Linked Data 3
1.1 The First IAB-SOEP-Migration Sample starting in 2013 3
1.2 The Integrated Employment Biographies (IEB) 4

2 Informed Consent to Record Linkage 4
2.1 Previous Research 4
2.2 Format of the Consent Request 10
2.2.1 Legal Background in Germany 10
2.2.2 The Fieldwork 11
2.2.3 Experimental Design and the Effect of Consent Requests on Panel Stability 12

3 Consent Rates and Correlates of Consent 14
3.1 Consent Rates at the Household Level 16
3.2 Consent Rates by Interviewers 18
3.3 Variance Components of Respondent- Household- and Interviewer-Level 19
3.4 Multivariate Analysis 20

4 Propensity Score Weights 29

5 Conclusion 32

6 Literature 34

Introduction

The collaborative research project of a joint sample of migrants conducted by the Institute for Employment Research (IAB) and the German Socio-Economic Panel (SOEP)—the IAB-SOEP Migration Sample—opens the opportunity to link the survey data with administrative data held by the Federal Employment Agency. To make this linkage, survey respondents are actively requested to give permission at the end of their personal interview. This paper covers how this procedure was completed, examines response rates, and addresses selectivity bias in the consent decision.

The IAB-SOEP Migration Sample was first conducted in 2013 focusing on migrants who entered Germany between 1995 and 2010 (N=2,723 participating households). An enlargement sample from 2015 adds respondents to the study who entered Germany between 2009 and 2013 (N=1,096 participating households). The present paper draws on data emanating from the first two waves – conducted in 2013 and 2014 – of the first migration sample.

This project is among the first in Germany to provide data on randomly sampled households of migrants and their descendants that not only contains detailed survey data on the family situation, social background, migration history, integration experiences, personality traits, and values, but also highly detailed register information on labor-market participation, incomes, and transfers (see, Brücker et al. 2014). In addition, both data sources have a longitudinal structure, thereby allowing researchers to study changes over time. Therefore the study closes an existing data-gap in social- and economic- integration research.

For data protection reasons, any linkage between survey and administrative social security data requires the informed and signed consent of respondents. In addition, in order to successfully identify individuals in the administrative records, consenting respondents often need to provide personal information like their full name or their social security number. Previous research suggests that not all respondents consent to linking their registry data, thus producing possible selectivity bias in the linked data (for an overview, see Sakshaug et al. 2012). The longitudinal nature of the IAB-SOEP Migration Sample provides the potential to maximize consent rates by choosing the optimal wave for implementing the request. However, the request may also undermine the motivation of respondents to participate in future waves, thus inflating attrition rates. To study the optimal timing and consequences of linkage requests, an experimental design was implemented in the first migration sample by randomly selecting the households into different groups. A control group was excluded from

the request, while the other groups received it in wave one or in a consecutive wave. By the second survey wave, in 2014, 43 percent (2,295 individuals) of the sample – 58 percent of the respondents who received the consent request – agreed to record linkage.

While results of the experimental design have been published by Eisnecker and Kroh (2016), this data documentation provides additional in-depth analyses as well more detailed supplementary information on the design as well as the findings. First, this paper gives a brief overview of both data sources, i.e. survey as well as register data. Second, a detailed description of the request procedure and its background is given. This includes a more in depth description of the survey experiment mentioned above and its main findings. Third, we analyze respondents' (non-)consent with special regard to the homogeneity of the decision within households and interviewers. We estimated multivariate models drawing on respondent-, household-, neighborhood-, and community- characteristics to identify correlates of consent. Based on these analyses, we fourth construct a propensity score weight that addresses the bias that results from selective consent decisions. Fifth and finally, the paper is briefly summarized.

1 The Linked Data

The following paragraphs provide a brief overview of the two data sources we aim to link. For a more general overview of the IAB-SOEP-Migration Sample see Brücker et al. (2014) and for a comprehensive description of the sampling design and non-response in the study see Kroh et al. (2015). Oberschachtsiek et al. (2009) provide an overview of the Integrated Employment Biographies (IEB), the work and benefits data set compiled by the Federal Employment Agency.

1.1 The First IAB-SOEP-Migration Sample starting in 2013

The 2013 IAB-SOEP-Migrant Survey was drawn from the population of persons migrating to Germany since 1995, as well as descendants of immigrants born in Germany after 1975. The study uses the Integrated Employment Biographies (IEB) - containing personal information on work and benefits - as a sampling frame. Households with migrants from Poland, Romania, the CIS countries, Turkey, the former Yugoslavia, and the countries of Southern-Europe (Italy, Spain, and Greece) as well as Arab or predominantly Muslim countries were deliberately oversampled.

The roughly 5,000 respondents across 2,700 households represent an enlargement sample of the ongoing Socio-Economic-Panel Study, which is conducted at a yearly basis. At the end of their personal interview all survey respondents were asked to give permission to link their survey information with the data held by the Federal Employment Agency. The questionnaire of the IAB-SOEP Migration Sample, targeting all adults in the household, parallels the standard SOEP questionnaire. Beyond this, the survey covers each respondent's complete immigration history, education and training completed in Germany and abroad, employment history in Germany and other countries, as well as numerous aspects of the cultural and living environments relevant to the social integration of migrants.

1.2 The Integrated Employment Biographies (IEB)

The IAB-SOEP Migration Sample will be linked with the corresponding records of the Integrated Employment Biographies, which also represented the sampling frame of the study. The IEB is a database compiled by the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung: IAB) that includes data on employment and benefits received on a daily basis reaching back until 1975. Data are collected on employees in Germany who are subject to social security contributions, which covers almost all private sector employment and a large part of public sector employment. Civil servants who are not covered by the social security insurance system, so called *Beamte*, are not contained, neither are freelancers and entrepreneurs. Employers are liable by law to submit information on starting and ending dates of all their workers' jobs as well as total earnings received (censored at the social security contribution ceiling) on an annual basis. In addition, the establishment identification number plus some job and personal characteristics are recorded. In total, the IEB V.11 contains almost 84 million individuals with roughly 1.9 billion spells from 1975 through the end of 2012. Furthermore, information on unemployment spells, benefit receipt, 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 individual labor market histories.

2 Informed Consent to Record Linkage

2.1 Previous Research

Table 1 presents an overview of the academic literature on requests for data linkage between surveys and various forms of records. The dates of the listed studies show that academic

interest in record linkage procedures arose around the beginning of the new millennium. A considerable number of the works stem from epidemiology and request the consent to link health surveys with various forms of health related datasets, like hospital and health-insurance records. Studies in the field of social sciences and economics are often interested in the linkage of surveys to records that provide information about employment histories, taxes, and various forms of income, such as pension and social benefits records.

Table 1**: Previous Research on Consent Requests**

Study	Survey	Record	Consent-Rate	Form of Consent	Significant Findings	Country
Antoni 2011	General topic survey	Employment and social security records	92%	unknown	Born in Eastern Germany (+), Active in high culture (-), Unemployed (-), Income refusal (-), Interviewer male (-), Age interviewer (+), Experience interviewer (+), Share of refused/ don't know answers (-), Problems during interview (-), Generally cooperative (+), Interviewer older than respondent (-)	Germany
Baker et al. 2000	Health survey	Health records	90%	Checking a box to refuse	N.S.	UK
Beste 2011	Panel survey with overrepresentation of poor households	Employment and social benefits records	80%	Written consent	Missing answers (-), Immigrant vs. native (-), third gen. immigrant vs. native (+), income (+), wrong information about social benefits reception (-), Panel refusal (-), other person in household consents (+), Interview not in German (-), Middle aged interviewer (-), Interviewer low education (+), Interviewer female (-)	Germany
Dunn et al. 2004	Health survey	Health records	70 - 90%	Written consent	Age (-), Male (+), Respondent has researched illness (+)	UK
Harris et al. 2005	Survey of older respondents	Health records	92%	unknown	Chronical illness (+), Chronic pain (+), Disabled (+), Anxious (+), Car owner (+)	UK
Hartmann & Krug 2009	Employment survey	Employment and social security records	74%	unknown	German vs. migrant (+), male (+), Eastern German vs. Western German (+), Income (+), No income information (-), Interviewer age (+), Interviewer education (+), History of unemployment (-), Social benefits recipient (+), Nonresponse sensible questions (-)	Germany
Hockley et al. 2007	Panel Survey of infants, parents answer questions	Birth registration Hospital records	87-93%	Written consent	Not researched	UK
Huang et al. 2007	Health survey	Health insurance records	88%	Written consent	Ethnic minority (+), Age (-), Illiterate (-), Suburban vs. urban area (-), Income (-)	Taiwan
Huber & Schmucker 2009	Panel survey on adult education	Employment and social security records	91%	unknown	Age (+), Income (+)	Germany

Study	Survey	Record	Consent-Rate	Form of Consent	Significant Findings	Country
Jenkins et al. 2006 (Linkage a)	Panel survey with overrepresentation of low income households	Social benefits and tax Records	77%, 89% of these provi-ded insurance number	Written consent + provision of insurance number	Age 25-39 (-), Couple vs. single (+), Social benefits (+), Problems during interview (-)	UK
Jenkins et al. 2006 (Linkage b)	Panel Survey with overrepresentation of low income households	Employer records	59% (of res-pondents in employment)	Written consent + employer contact details	Age 60+ (-), Number children in HH (-), Education (+), Health (+), Local unemployment rate (-)	UK
Klassen et al. 2005	Survey of caregivers of infants	Health records	67-73%	Written consent + Provision of personal health number (own and child) (had to be sent per mail)	Biological parent of child (+), No reminders for survey required (+), In parent support group (+), Income (+), Not working full-time (+), Health problem child (+)	Canada
Knies et al. 2012	General topic panel survey	Two types of health Records	41%	Written consent	From England vs. rest-UK (+), White (+), Age (-), Male (+), Education (+)	UK
Korbmacher & Schroeder 2013	Health panel survey of older respondents	Pension records	78%	Written consent (has to be sent per mail)	Age (+), Age^2 (-), Lives with partner (+), Ever divorced (-), Ever lived in GDR (+), Urban area (-), Income missing (-), Medium income (-), Respondent comprehension (+), Missing rate financial questions (-), High interviewer experience (-), Interviewer age (-), Interviewer age ^2 (+), Interviewer high education (+), Interviewer quality indicator (+)	Germany
Sakshaug & Kreuter 2014	Web survey on finances and employment	Employment records	60%	Check a box to answer	Wording that emphasizes time saving (+)	Germany
Sakshaug et al. 2012	Panel survey of older respondents	Social security records on earnings and benefits	68%	Written consent + Social security number (optional)	Years of education (+), Net worth (-), Social benefits (+), Confidentiality concerns (-), Refusal financial questions (-), Resistance against interview (-), Wave nonresponse (-)	USA

Study	Survey	Record	Consent-Rate	Form of Consent	Significant Findings	Country
Sala et al. 2012 (Linkage a)	General topic panel survey	Health records	41%	Written consent	British/Irish white vs. other white (+),British/Irish white vs. British Asian/black (+), Income refusal (-), Trust (+), Leftist/Liberal (+), Voluntary worker (+) Social benefits (+), Interview sequence position in household (-), Number of previous consents in household (+), Years within panel (-), Number of previous interviews by interviewer (-), Previous consents of interviewer (+)	UK
Sala et al. 2012 (Linkage b)	General topic Panel Survey	Social benefits Records	32%	Written consent	British/Irish white vs. other white (+), Income refusal (-), Trust (+), Leftist/Liberal (+), Health problems (+) Social benefits (+), Interview sequence position in household (-), Number of previous consents in household (+), Years within panel (-), Number of previous interviews by interviewer (-), Previous consents of interviewer (+), Interviewer belief: all can be persuaded (-)	UK
Silva et al. 2002	Health survey of women	Health records	34%	Written consent (has to be sent per mail)	N.S.	Australia
Tate et al. 2006	Panel survey of infants, parents answer questions	Birth registration Hospital records	92%	Written consent	England vs Scotland (+), England vs. N.Ireland (+), Mother racial/ethnic minority (-),Mother mid-level education (+), Lone parent (-), Teenage mother (+), Male translator (-)	UK
Woolf et al. 2000	Health survey	Health records	67%	Written consent	Age (+), Male (+), Health (-), Nonresponse (-)	USA
Young et al. 2001	Health panel survey of women	Health insurance records	37-59%	Written consent (has to be sent per mail)	Education (+), Private health insurance (+), Frequency doctor visits (-), Mortality in following time (-)	Australia

Notes: Consent rates are rates of respondents that were asked for their consent to record linkage and granted that consent. Persons who refused to take part in the survey or who did not reach the point of the consent request are not included. Written consent always includes a signing of the request. Findings are reported as significant if the 95% confidence interval does not include a certain critical value or have a p<=0.05 (dependent on what is reported). Insignificant findings are not reported. If available, the findings are taken from fully specified multivariate regression models. Country refers to the country in which the study was conducted.
Source: own compilation.

Consent rates to record linkage vary widely between 32 and 93 percent with an average of around 70 percent. The reason for this wide range of consent-rates may be seen in the vastly different settings in which the requests occurred, including varying target populations and countries. The general form of the (reported) requests is relatively uniform across studies and normally includes the signing of a written declaration. This formal procedure can be seen as a quite large barrier toward the agreement to data linkage. Requiring further information from the respondent that could be perceived as sensitive, like an insurance number, may present a further hurdle. Likewise, having to send back the form by mail could result in lower consent rates because more time must be spent and the pressure of an in-person request is missing. Although the general process of signing a consent form is quite similar in most studies, the placement within the interview and formulation of the question, as well as the provided information, appear to affect the outcomes. For example, presenting the record linkage as a way to shorten the interview increases consent rates (Sakshaug and Kreuter 2014).

It is not only the form of the request, but also the content of the administrative data which are to be linked that influences the rates of consent. For example, Sala et al. (2012) ask the same population for consent to link their health records as well as their benefits records to a single survey and obtain differing consent rates. In addition, the interplay between survey and administrative data might be important. In an epidemiological context, surveys aimed at health related issues often directly affect the respondent. The benefits of record linkage with health records to advance medical research, thus helping sufferers of these conditions are quite obvious in this context. The reasons to link general topic surveys and, for example, social security records may appear more obscure to many respondents. In addition, the perceived distance between the institution conducting the survey and institution holding the data may be relevant. For example, the surveys studied by Antoni 2011 and Huber and Schmucker 2009 (ALWA and WeLL, respectively) were conducted by the IAB, the research institute of the Federal Employment Agency, and persons consenting to participate in an IAB study might not strongly oppose against record linkage with Federal Employment Agency data. This may account for the extraordinarily high consent rates reported for those two studies.

Non-trivial levels of consent refusals and selection effects of the consent endanger the quality of the linked dataset. Table 1 also provides an overview of correlates of linkage consent found in previous studies. The main focus of the presented studies lies on characteristics of the individual respondents, though attributes of the household, the interview situation, the

interviewer, as well as the interactions between respondent- and interviewer-characteristics are also examined. The patterns of findings unsystematically differ across studies. In their review of 17 studies that requested access to medical records, Kho and colleagues conclude, "Across all outcomes there were differences between participants and non-participants, although there was a lack of consistency in the direction and the magnitude of effect" (Kho et al. 2009, p.5). There seem to be two exceptions from this general pattern of inconsistency. First, the refusal to answer questions that can be seen as sensitive, often concerning income, is correlated with the refusal of linkage queries (Woolf et al. 2000, Hartmann and Krug 2009, Antoni 2011, Beste 2011, Sakshaug et al. 2012, Sala et al. 2012, Korbmacher and Schroeder 2013). Second, migrants and ethnic minorities seem to have lower consent rates than the respective countries' majorities (Tate et al. 2006, Hartmann and Krug 2009, Beste 2011, Knies et al. 2012, Sala et al. 2012).[1] Beste (2011, p.13) supposes that difficulties in understanding the implications of the consent request due to language barriers as well as a lack of knowledge about the trustworthiness of the commissioning institutions may explain these differences.

2.2 Format of the Consent Request

2.2.1 Legal Background in Germany

In general, data protection issues in Germany are regulated under the *Bundesdatenschutzgesetz* (BDSG, German Federal Data Protection Law). The register data of the German Employment Agency belong to a special kind of data called *Sozialdaten*, which fall under the even stricter protection of the *Sozialgesetzbuch*. *Sozialdaten* are data that citizens are obliged to disclose in the course of certain interactions with administrative, executive, or state approved agencies. The use of register data of the German Employment Agency is regulated in Volume 10 of the *Sozialgesetzbuch* (SGB X).

Concerning data use and processing by third parties, both laws, BDSG and SGB X, require explicit and informed consent from the concerned persons. This consent must be in written form, except if important reasons (that have to be detailed) stand against it (BDSG § 4a (1) and SGB X § 67b (1) and (2)). While BDSG and SGB X are concurrent regarding the consent procedure, the differences between the regulations lie in the application process to obtain data from the data holding agency. In the case of *Sozialdaten* (SGB X) the use is restricted to research on certain topics, in the case of register data of the German Employment Agency to

[1] See Huang et al. (2007) for a counterexample.

research on labor and occupation. Moreover, the applicant must establish that no alternative way to achieve the research objective is available (§75 SGB X).[2]

2.2.2 The Fieldwork

The procedure to obtain consent from IAB-SOEP Migration Sample participants was the same in all households and panel waves. The request was presented to respondents verbally, with interviewers reading the following statement, translated from the original German:

> **This is the end of the individual questionnaire.**
>
> **Thank you for the interview!**
>
> **But we do have one last request.**
>
> **It is increasingly important for scientific research to have information about people's employment histories and to consider this data in statistical analyses.**
>
> **To this end, we would like to link the survey data from this interview with the social insurance data from the Institute for Employment Research (IAB).**
>
> **This is only possible under German data protection law with your express permission.**
>
> **Of course, the decision is entirely up to you.**
>
> **Please take a moment to read the form.**

The interviewers then handed over the printed consent declaration form. The form included an explanation of what the data that were to be linked are needed for. The form not only detailed what kind of information is contained in the administrative data, but also assures the respondent that all regulations concerning data protection are being followed (see TNS Infratest Sozialforschung 2014). Interviewers were instructed to treat the request for consent very sensitively and, if requested, to leave a copy of the form with the respondent.

If the respondent agreed to link the data, they were asked to fill in their family name, first name, and birth name (if deviating from their current family name). They then handed the signed and dated form back to the interviewer who forwarded it to TNS Infratest along with the other fieldwork-material.[3] In a final step, TNS Infratest sent the forms to the IAB.

---

2 The property *Sozialdaten* applies only to data which is attributable to individuals. If effectively anonymized, the data do not fall under special data protection laws any more, but - in case of the data of the German Employment Agency - the restriction of use concerning the research question still applies (so-called *Zweckbindung* of the data).

3 In 100 person-cases, verbal consent expressed during the interview to linkage did not match with submission of the signed form. In 79 of these cases, the questionnaire indicates that consent was given, but no signed form was received. The opposite is true in 21 cases. Regardless, the printed and signed consent declaration of the respondent is the only criterion determining if consent to record linkage was provided.

2.2.3 Experimental Design and the Effect of Consent Requests on Panel Stability

The IAB-SOEP Migration Sample is the first survey within the broader SOEP family that asks respondents for consent to link their answers to administrative records.[4] As such, it might serve as a point of reference for future linkage projects. One concern is that the request to link sensitive social security data may undermine the respondent's willingness to participate in further SOEP waves. To better understand the consequences of the consent request on the response behavior in the context of a longitudinal study, we experimentally varied the implementation of the request. In addition, the design allows us to study the influence of the request placement in later or earlier waves on consent rates (Eisnecker and Kroh 2016).

The experimental design randomly divided the households of the sample into five groups (see Table 2). This meant that all members of a household were selected into the same group. Treatment Group I was asked in the first wave (2013) and again in the third wave (2015), but not in the second wave (2014). The consent of Treatment Group II was requested in wave 2 (2014) and wave 3 (2015). The consent of Treatment Group III is asked for in all three waves, while Treatment Group IV is only queried in the third wave (2015). The respondents in the control group are not asked for their consent to record linkage in any panel wave. If an individual respondent has already consented, he or she is not asked again in later waves. In addition, members of Treatment Group III only receive a request in the third survey wave (2015) if they have not rejected twice already. This means that only respondents of Treatment Group III that have not participated in both wave 1 (2013) and wave 2 (2014) are subjected to the request in wave 3 (2015). Respondents starting to participate in the survey after wave 1 (2013) receive the experimental treatment of their household. If survey members moved to another household, this new SOEP household was allocated to the same experimental group as the household from which the original survey-member originated.

[4] However, this was not the only request for consent to data-linkage within the SOEP-family. The SOEP-LEE project (German Institute for Economic Research Berlin 2014b) linked employer and employee data. The SOEP-ECEC Quality project (German Institute for Economic Research Berlin 2014a) linked children with educational institutions. Finally, the EVA-MIN project (German Institute for Economic Research Berlin 2014c) connected selected low-income respondents with qualitative group discussions in which they had participated.

Table 2: **Experimental Design**

	Asked for consent in wave…			Realized Households (2013)	Realized Respondents (2013)
	2013	2014	2015	N	N
Control Group	-	-	-	377	691
Treatment Group I	X	-	X*	525	990
Treatment Group II	-	X	X*	362	657
Treatment Group III	X	X*	X*#	1,277	2,287
Treatment Group IV	-	-	X	182	339
Total N				2,723	4,964

** Only if individual respondent has not already agreed to record linkage.*

 # Only if individual respondent has not already rejected record linkage twice.

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

By the second survey wave in 2014, 3,767 of the original sample members (76 percent) were asked for their consent at least once. In addition, 213 respondents that were first questioned in wave 2 (2014) (60 percent of all newcomers) were asked for their consent[5].

In order to assess the influence of linkage requests on panel stability, Table 3 reports the number of persons across waves according to exposure to a consent query. We do not find any significant difference in the rate of non-continuing persons between exposed and unexposed respondents. Therefore, the experimental results do not support the hypothesis that the request to link sensitive administrative data may undermine the respondent's willingness to participate in further survey waves (see, Eisnecker and Kroh 2016).

[5] We received 60 signed consent forms in 2013 and additional 13 signed forms in 2014 from respondents who had not been allocated to a request condition in the respective wave. We treat these respondents as regular linkage-consenters.

Table 3: **Consent Request and Panel Participation**

	Panel Participation		
Consent Request 2013	Attrition 2014	Participation 2014	Total
Consent not requested (Control Group + Treatment Groups II & IV)	464	1,163	1,627
(percent)	(28.5)	(71.5)	(100)
Consent requested (Treatment Groups I + III)	1,023	2,314	3,337
(percent)	(30.7)	(69.3)	(100)
Total	1,487	3,477	4,964
(percent)	(30.0)	(70.0)	(100)

Chi2: 2.3817 Pr: 0.123

Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

3 Consent Rates and Correlates of Consent

The following paragraphs detail the outcome of the linkage request. Moreover, we also report correlates of the consent decision drawing on individual, household and neighborhood data. In general, we find a relatively low accumulative consent rate of about 58 percent; however, the analysis also suggests few systematic differences between those who consent to the linkage and those who do not.

Table 4: **Number and Rate of Consenting Respondents according to Treatment Groups**

	Year of Consent Request		
	2013	2014	Cumulative
Control Group	26*	7*	33*
(percent)	(.)	(.)	(.)
Treatment Group I	490	4*	494#
(percent)	(49.5)	(.)	(49.9)
Treatment Group II	25*	235	260#
(percent)	(.)	(49.3)	(51.8)
Treatment Group III	1,114	382	1,496
(percent)	(48.7)	(41.7)	(61.3)
Treatment Group IV	9*	3*	12*
(percent)	(.)	(.)	(.)
Total	1,664#	631#	2,295#
(percent requested)	(49.9)	(44.8)	(57.7)

** Signed consent forms from respondents that should not have been queried in the respective wave following the experimental design. As we do not know the total number of respondents that have been erroneously queried this way, no percentages can be provided.*
 # In these cumulated percentages, respondents who have been erroneously queried and provided written consent are counted as consenting respondents within the category.
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

The accumulative consent rate among all requested respondents amounts to 58 percent after the second survey wave in 2014, with a total number of 2,295 consenting respondents (Table 4).[6] This is lower than the average consent rate of around 70 percent found in other studies (see section 2.1 of this survey paper), especially since some respondents in Treatment Group III only consented after they were asked a second time in wave 2 (2014). One explanation for the lower turnout may be the target population of survey, as a number of studies show that migrants and ethnic minorities seem to have lower consent rates than the respective countries' majorities (Tate et al. 2006, Hartmann and Krug 2009, Beste 2011, Knies et al. 2012, Sala et al. 2012).

[6] As respondents in all treatment groups will be confronted with further consent requests, the cumulative consent rate can be expected to rise further in the future.

Following the experimental design, the placement of the first request for consent varied between Treatment Groups. However, the consent rates for the first confrontation with the request did not differ strongly between waves. It amounted to 49.5 percent for Treatment Group I and 48.7 percent for Treatment Group III (both being asked first in wave 1 (2013)) opposed to 49.3 percent for Treatment Group II (which was asked first in wave 2 (2014)). Therefore, a later placement of the first request didn't yield higher consent rates (for a more detailed report, see Eisnecker and Kroh 2016).

3.1 Consent Rates at the Household Level

The consent pattern is extremely homogeneous within households. Table 5 displays the homogeneity of the consent decision within households for the first consent request households were given (wave 1 for households in Treatment Groups I and III, wave 2 for households in Treatment Group II). Approximately 86 percent of all households with more than one respondent decided unanimously on record linkage, divided almost equally between those in which all respondents refused and those in which all respondents consented.

Table 5: **Consent Pattern within Households (2+ Adult Respondents)**

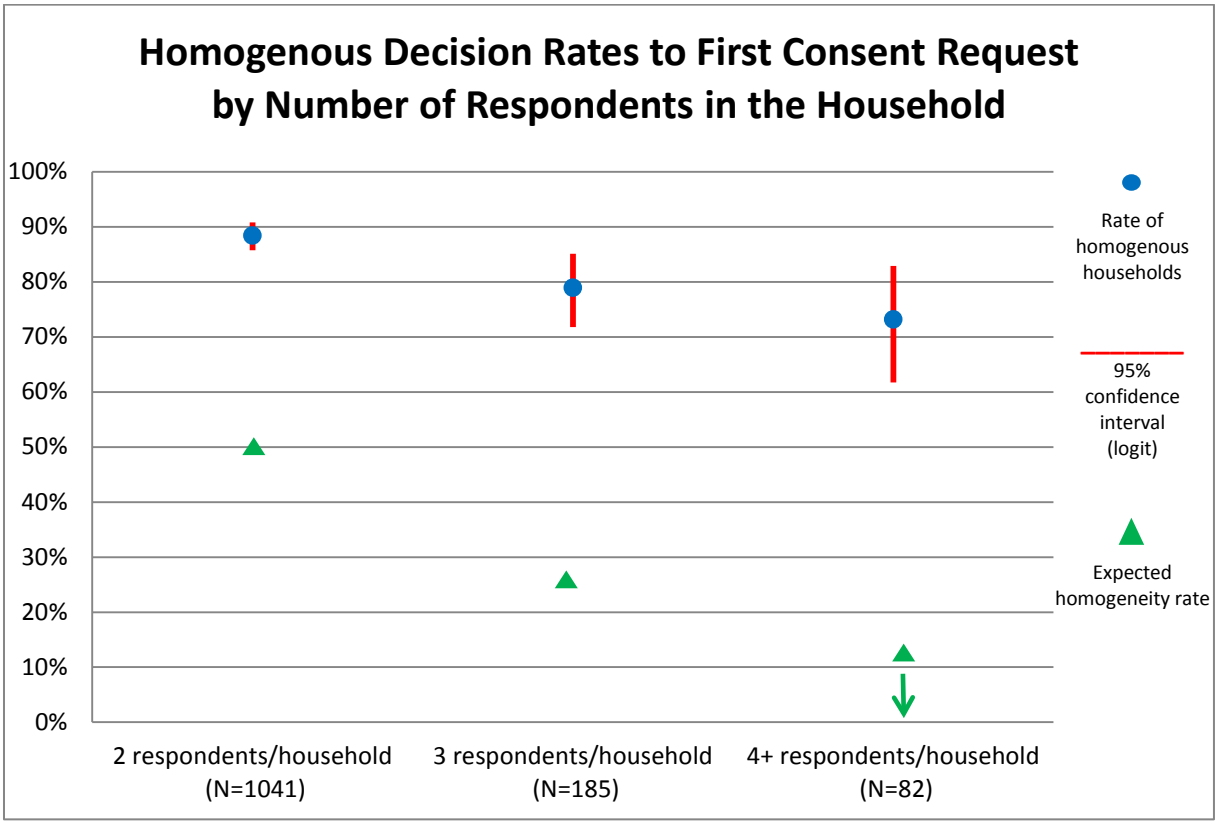
First Consent Decision in Households with more than one Respondent is…	N (households)	% (homogeneous HH)	% (all HH)
…homogeneous (all consented)	566	50.3	43.3
…homogeneous (all refused)	560	49.7	42.8
Subtotal	1126	100	86.1
…heterogeneous	182		13.9
Total	1308		100

Note: the number of respondents within a household is mostly less than the number of household members.
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

As expected, the homogeneity of the consent decision within households decreases with increasing numbers of respondents per household, but only slightly. The rate of households with a homogenous decision ranges from 88 percent for households with two respondents to 73 percent for households with four or more respondents (see Figure 1). Even if the consent-decision of household members was totally unrelated with each other, a number of

homogeneously deciding household would still be expected due to chance alone.[7] However, because there is no overlap between the confidence intervals of the observed homogeneity-rates and the expected rates at random, it can be inferred that the consent-decisions within households are interconnected.

Figure 1: **Homogenous Decision and Number of Respondents in Household**



Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

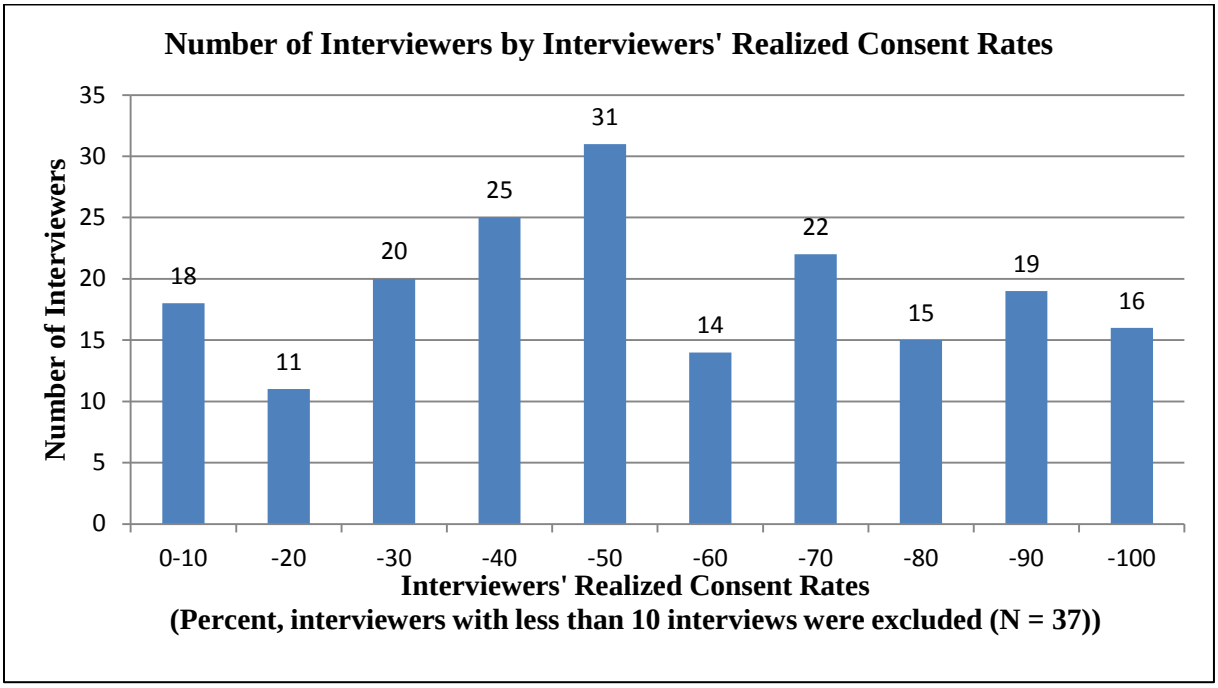
We can think of three underlying mechanisms leading to these patterns: first, some way of producing accordance within the respondents of the same household during the interview process is in action. However, the interviewer instructions state to interview members of the same household one by one and, if possible, alone. Therefore, discussions of the issue between the household members leading to unanimous decisions should occur only exceptionally. Second, the consent to record linkage might be the outcome of individual factors that tend to be rather similar among household members, such as shared attitudes and

7 As the probability to consent to the first linkage request is roughly 0.49 and the chance to reject is 0.51 among all respondents, the expected homogeneity rates at random can be calculated as follows: 0.49^2 + 0.51^2 = .50 for households with two respondents, 0.49^3 + 0.51^3 = .25 for households with three respondents and 0.49^4 + 0.51^4 = .125 or lower for households with four and more respondents.

sociodemographic characteristics that influence the consent decision independently for each respondent. Finally, household and/or interviewer characteristics may have the same impact on the decisions of all household members. Examples could be the personality of the interviewer or the household income.

3.2 Consent Rates by Interviewers

Figure 2: **Interviewers by Realized Consent Rates**



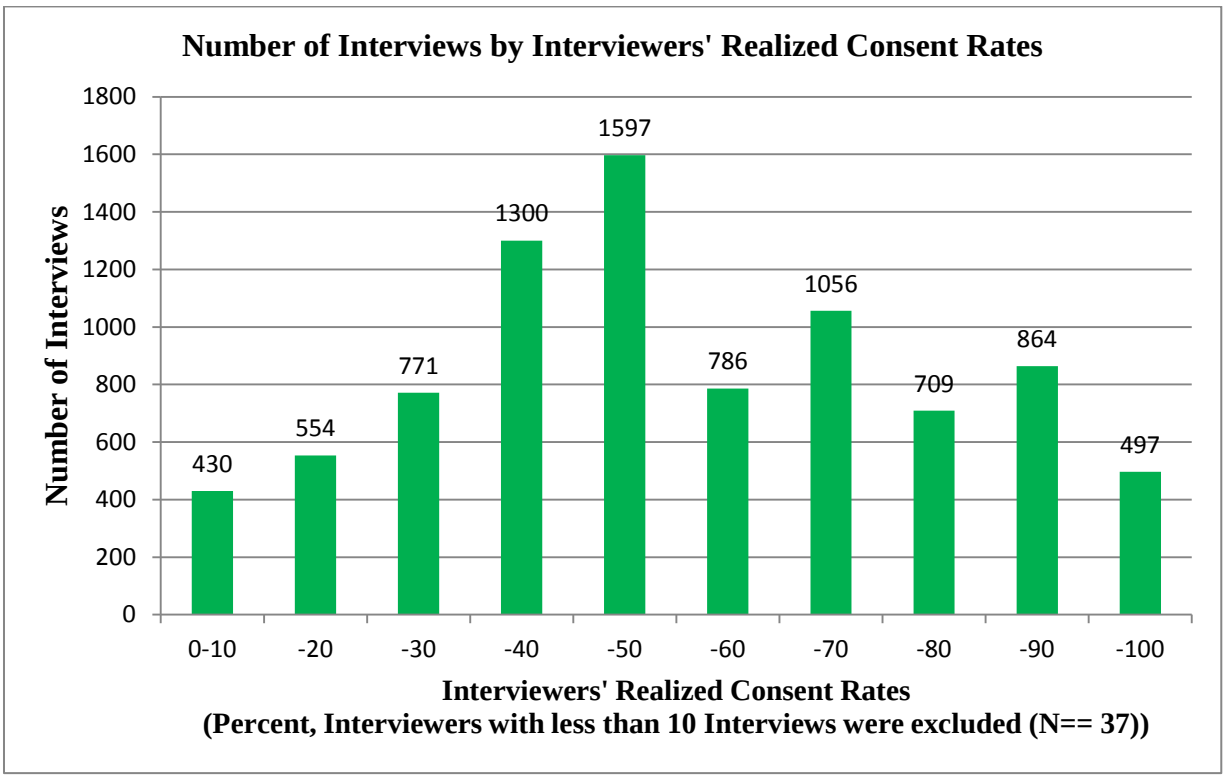
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

Figure 2 shows 10 classes of consent rates on the x-axis and the number of interviewers realizing the respective consent rate on the y-axis. The 37 interviewers conducting less than 10 interviews were excluded because calculating a consent rate on basis of less than 10 interviews seems inappropriate. The remaining 191 interviewers completed between 10 and 182 interviews in wave 1 (2013) and wave 2 (2014) of the survey.

The number of interviewers peaks in the 40-50 percent category, which roughly mirrors the average respondents' consent rate for the first consent request of less than 50 percent. However, the rather uniform distribution over consent categories indicates that the extent to which the interviewers obtained consent to record linkage varies greatly. Some interviewers received no agreements, while others attained the consent of every respondent. Note, however, that these differences may also reflect regional differences (like municipality size or

general “milieus”) instead of interviewer influence. Regional sample points in the IAB-SOEP Migration Sample and interviewer allocation match too closely to empirically disentangle both influences.

Figure 3: **Interviews by Interviewers’ Realized Consent Rates**



Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

Figure 3 shows the number of interviews that were conducted by the interviewers falling into the different consent rate classes. Again, the interviewers conducting less than 10 interviews are excluded. Most interviews were conducted by interviewers who achieved average consent rates between 40 and 50 percent. However, a relatively large number of interviews were also realized by greatly less or more successful interviewers. As interviewers’ average consent rates showed noticeable variation and a relatively large number of interviews were conducted by interviewers with extreme average consent rates, interviewer characteristics may have strongly influenced respondents’ consent decisions.[8]

3.3 Variance Components of Respondent- Household- and Interviewer-Level

Table 6: **Variance Components and Comparison of Hierarchical Regression Models**

[8] See Korbmacher and Schroeder (2013) for the role of interviewer effects in consent requests.

	Respondents nested in Households	Respondents nested in Households, Households nested in Interviewers
Variance Components (Percentages)		
Residual Variance	25.0	24.7
Household-level Variance	75.0	51.8
Interviewer-level Variance	.	23.5
Likelihood Ratio Test (Chi2, sig.)		
Vs. Naive Linear Model	1,478**	1,773**
Vs. Respondents in Household Model	.	295**

***p<0.01*
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

Consent and refusal to record linkage could be shown to be quite uniform within households. In addition, interviewers varied strongly in their success to obtain consent. To further study the influence of different hierarchical levels on respondents' consent decision; we compared the variance components on the level of interviewers, households, and respondents (Table 6). If respondents are modeled as hierarchically nested within households, 75 percent of the total variance of the linkage-consent can be attributed to the household-level. The Likelihood Ratio Test indicates that the two-level modeling fits the data significantly better than the naïve linear model. Introducing the third level of interviewers decreases the variance attributable to the household to about 52 percent, while the residual variance of roughly 25 percent remains unchanged. About 23 percent of the total variance lies on the interviewer level. The introduction of this third level significantly increases the model fit compared to the one- and two-level models. The variance decomposition further emphasizes the role of interviewers and to an even greater extent the role of the household context in respondents' consent decisions.

3.4 Multivariate Analysis

The following regression models of the individual consent decision are informed by previous findings in the field (see section 2.1). We consider a wide range of factors at the respondent-level, such as country of origin and citizenship, income and employment, German language skills as well as attitudes. Due to the high similarity of consent decisions within households, we also draw on a range of characteristics that are either directly related to the household or to

its surrounding neighborhood and community. Table 7 provides an overview of the employed variables. Note that we did not include interviewer characteristics in the models. Although we are aware that these are known to influence consent decisions (see for ex. Korbmacher and Schroeder 2013), the following analysis is only interested in selective characteristics directly or indirectly connected to respondents.

We use a variety of data sources. The individual-level characteristics are mainly drawn from the personal questionnaire of wave 1 (2013) and wave 2 (2014). Household characteristics are drawn from the household questionnaire of wave 1 (2013) and wave 2 (2014) as well as from the fieldwork organization. We exploited the MICROM-SOEP-dataset (Goebel et al. 2007) to obtain regional information on the households' neighborhoods. So-called 'microm segments' consist of at least five and an average of eight households, thereby representing the close neighborhood in which a respondent-household is located. If information on the 'microm segments' was not available, we used data on the level of 'street segments.' Information on the wider community is drawn from the fieldwork organization.

We enter all explanatory factors of consent in form of categorical variables (see Table 7 for an overview). Originally continuous variables were transformed into a reasonable number of categories, ensuring that not too much information was lost while enough cases remain in every cell. This procedure ensures that extreme outliers cannot distort findings. Moreover, this enables the detection of non-linear effects. One category serves generally as reference and is not contained in the models.

Table 7: **Variable Overview**

Variable Name	Remark	Values/Range	Level	Source
Basic Respondent Characteristics				
Gender: Male		0 = female 1 = male	respondent	personal quest.
Age		0(ref) = under 26 … 4 = Over 50	respondent	personal quest.
Highest educational degree	only general education considered	0(ref) = Compulsory school/Hauptschulab. 1 = No educ. degree/other degree 2 = Secondary education 3 = University degree 4 = Still in education 5 = Education missing	respondent	personal quest.
Employment and income				
Employment		0(ref) = Un/semiskilled position/helping family member 1 = Not working, in education, missing 2 = Skilled position 3 = Leading position 4 = Self-employed	respondent	personal quest.
Net income from employment	Monthly income	0(ref) = 0-600 EU 1 = No income from employment 2 = 601-1200 EU 3 = 1201-2000 EU 4 = Over 2000 EU 5 = Income missing	respondent	personal quest.
Origin and citizenship				
Country of origin	Country of birth if not Germany, Mother's country of birth if respondent was born in Germany, father's country of birth if mother and respondent were born in Germany	0(ref) = States of former USSR 1 = Germany/unclear 2 = Italy 3 = Greece/Spain 4 = Turkey 5 = Former Yugoslavia 6 = Poland 7 = Romania 8 = Arabic/Muslim countries 9 = Rest of the world	respondent	personal quest. parents' personal quest. (if parents live in household)
Born in Germany		0 = No 1 = Yes	respondent	personal quest.
German citizenship	Double German/other citizenships treated like German citizenship	0(ref) = Non-German citizen 1 = Naturalized German citizen 2 = German citizen since birth	respondent	personal quest.
Language and Integration				
Subjective German language skills		0(ref) = Very well/native speaker/missing … 3 = Badly/not at all	respondent	personal quest.
Feeling of being German		0(ref) = Completely/missing … 3 = Not at all/not very	respondent	personal quest.

Variable Name	Remark	Values/Range	Level	Source
Attitudes				
Interest in Politics		0(ref) = Disinterested 1 = Moderately interested 2 = (Very) interested	respondent	personal quest.
Readiness to take risks (0-10)	Originally measured on 11-point scale	0(ref) = Low (0-3) 1 = Medium (4-6) 2 = High (7-10)	respondent	personal quest.
Household characteristics				
Social housing		0 = No 1 = Yes	household	household quest.
State of dwelling		0(ref) = In good condition 1 = needs some renovation 2 = needs renovation/dilapidated	household	household quest.
Household size	Includes both respondents and nonrespondents in the household	One person(ref) … Five and more persons	household	fieldwork information
Equalized disposable net-income	The equalized disposable net-income was calculated by dividing the household net-income by the square root of the number of persons living in the household	0(ref) = 0-900 EU … 3 = Over 1700 EU 4 = Income missing	household	household quest. fieldwork information
Neighborhood characteristics				
Rate of non-Germans in direct neighborhood		0(ref) = Lowest-well below av./missing … 3 = highest	MICROM segment	MICROM
Rate of academics in direct neighborhood		0(ref) = 0-3%/missing … 3 = Over 10%	MICROM segment	MICROM
Spending power in street segment		0(ref) = Well below av.-below av./missing … 3 = (Well) above av.	Street segment	MICROM
Community characteristics				
Former East-Germany		0 = West 1 = East	household	fieldwork information
Community size		0(ref) = 500,000 and more … 4 = 20,000 and less	household	fieldwork information

Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

Two multivariate logistic regression models were estimated to investigate possible selectivity in the consent decision. Both models predict if consent to record-linkage was given (=1) or refused (=0). As some respondents were asked again if they refused the first request, we only consider the outcome of the last request.

All models include the respondents that were asked at least once for consent to record linkage during wave 1 (2013) or wave 2 (2014). For time varying predictors, the measurement point of the covariates is always the same wave in which the deciding consent request was answered.

Model 1 includes all covariates that were measured on the respondent- and household-level, including neighborhood- and community characteristics. Model 2 contains only those variables that were significant on the 5 percent level in Model 1. This procedure ensures that only the variables with the most robust effects under different model specifications were included in Model 2. Both models are calculated with clustered robust standard errors (Williams 2000) on the level of current household.

Table 8: **Fit Values for the Estimated Models**

	Full Model M1	Reduced Model M2
Pseudo R2	0.05	0.03
Rate of **rejecting** Respondents with false predictions	27.9	32.8
Rate of **consenting** Respondents with false predictions	10.0	6.9
Overall rate of Respondents with false predictions	37.8	39.7

Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014, own calculations.

Table 8 reports Pseudo R^2 values for both models as well as error rates of the models' predictions for both rejecting and consenting respondents. The models are assumed to predict rejection if the predicted consent probability is lower than 50 percent and consent if the probability is 50 percent or higher.

The fit values for both models indicate that their predictions are relatively unreliable, with incorrect predictions accounting for roughly 38 and 40 percent of all predictions in the model specifications. Therefore, the wide range of characteristics included in the models proved not to influence respondents' consent decisions substantially. This can be interpreted as a positive sign for data quality as the joint dataset can be expected to be relatively unbiased by selection effects.

Table 9 displays the regression coefficients of both calculated models in the form of average marginal effects. These can be interpreted as the average change in percentage points of the probability of consent if an independent variables' value is increased by one unit.

On the **respondent level**, persons with university degree were more likely to grant consent than those with only compulsory education. Contrary to this, respondents who did not provide their educational degree, who have roots in Arabic or Muslim countries, and/or who feel not very or not at all German are less likely to consent than the respective reference groups. Unlike other studies, respondents who did not provide their personal income were not more likely to reject the consent request. In general, we found no significant effects for respondent characteristics related to employment and income or to personal attitudes.

On the **household level,** we found respondents in very large households with five or more respondents to be less likely to consent than respondents living alone. Surprisingly, persons living in dwellings that need renovation are more likely to consent than those in houses in good condition. The strongest negative effect in the model is connected to non-provision of the household income in the household questionnaire. This finding contrasts sharply with the effectlessness of the non-provision of the personal income.

We found no selectivity connected to neighborhood- or street-characteristics. Respondents living in households located in the former East Germany are more likely to provide consent than those living in the former West Germany. Finally, persons residing in cities with 10,000 to 500,000 inhabitants are less likely to consent than those in cities with over 500,000 residents.

All coefficients in the reduced model remained significant at the 5 percent level and are relatively stable in their effect size, which indicates the robustness of the effects under different model specifications.

Table 9: **Multivariate Logistic Regression with the Dependent Variable "Consent to Record-Linkage" Granted (=1) or Rejected (=0)**

	Model 1 (Full Model) Average Marginal Effects (%)	Model 2 (Reduced Model) Average Marginal Effects (%)
Basic respondent characteristics		
Gender: Male	-0.20	
Age ***Ref: under 26***		
26-30	3.54	
31-40	-2.01	
41-50	-2.70	
over 50	-4.70	
Highest educational degree ***Ref: Compulsory school/Hauptschulab.***		
No educ. degree/other degree	-5.46	
Secondary education	-0.57	
University degree	6.10*	7.23**
Still in education	-5.92	
Education missing	-17.18**	-16.96**
Employment and income		
Employment ***Ref: Un/semiskilled position/helping***		
Not working/ in education/ missing	1.34	
Skilled position	-3.10	
Leading position	-7.03	
Self-employed	-5.52	
Net income from employment ***Ref: 0-600 EU***		
No income from employment	-4.61	
601-1200 EU	-3.51	
1201-2000 EU	-0.42	
Over 2000 EU	1.82	
Income missing	-6.60	
Origin and citizenship		
Country of origin ***Ref: States of former USSR***		
Germany/unclear	2.58	
Italy	-6.00	
Greece/Spain	1.45	
Turkey	-7.16	
Former Yugoslavia	-1.85	
Poland	0.34	
Romania	6.55	
Arabic/Muslim countries	-12.77**	-10.73*
Rest of the world	2.35	
Born in Germany	-0.34	
German citizenship ***Ref: Non-German citizen***		
Naturalized German citizen	-2.30	
German citizen since birth	-4.28	

Language and Integration		
Subjective German language skills ***Ref: Very well/native speaker/missing***		
well	-2.03	
okay	0.10	
badly/not at all	-0.40	
Feeling of being German ***Ref: Completely/missing***		
mostly	3.98	
in many respects	-4.56	
not at all/not very	-6.16*	-5.41**
Attitudes		
Interest in Politics ***Ref: disinterested***		
moderately interested	0.20	
(very) interested	1.41	
Readiness to take risks (0-10) ***Ref: low (0-3)***		
medium (4-6)	2.26	
high (7-10)	3.94	
Household characteristics		
Social housing	-2.14	
State of dwelling ***Ref: In good condition***		
needs some renovation	0.70	
needs renovation/dilapidated	12.07*	13.37*
Household size ***Ref: one person***		
two persons	-0.41	
three persons	-0.45	
four persons	-0.65	
five or more persons	-8.34*	-8.64**
Equalized disposable net-income ***Ref: 0-900 EU***		
901-1200 EU	-2.70	
1201-1700 EU	-1.29	
Over 1700 EU	-3.58	
Income missing	-24.63**	-24.53**
Neighborhood characteristics		
Rate of non-Germans in direct ***Ref: Lowest-well below av./missing***		
below av.-slightly below av.	3.40	
av.-above av.	1.79	
highest	-3.32	
Rate of academics in direct neighborhood ***Ref: 0-3%/missing***		
3-5 %	2.71	
5-10%	2.55	
Over 10%	4.75	
Spending power in street segment ***Ref: Well below av.-below av./missing***		
slightly below av.	-0.39	
av.-slightly above av.	-2.86	
(well) above av.	-2.99	

Community characteristics		
Former East-Germany	9.70**	10.60**
Community size *Ref: 500,000 and more*		
100,000 - 500,000	-8.13**	-5.93**
50,000 - 100,000	-5.34	
20,000 - 50,000	0.55	
20,000 and less	-6.35	
Total Number of respondents	3924	3924
Total Number of households	2113	2113
Pseudo R2	0.05	0.03

Note: Both models are calculated with clustered robust standard errors on the level of current household.
*** p<0.01 * p<0.05*
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

4 Propensity Score Weights

The previous analyses show that respondents' consent decisions are relatively independent from the wide range of variables included in our multivariate analysis. However, a relatively small number of significant effects are found. If statistical analyses are only restricted to respondents that consented to record linkage, inferences might be erroneous because some groups are more likely to provide consent than others. For this reason, a weighting variable is provided to compensate selection effects.

Weights were generated for all respondents who consented to record linkage in waves 2013 or 2014 on the basis of the Reduced Model. Contrary to the Full Model, the Reduced Model contains fewer insignificant variables that artificially increase the variation of estimated propensity scores and, thereby, also the non-consent weights. This is known to decrease the certainty of weighted estimations and should, therefore, be avoided if possible. The reduced model could be shown to fit the empirical data only slightly more poorly than the full model.

Table 10 displays percentiles, standard deviations as well as maximum- , minimum- and mean values of the assigned raw weights for all respondents and selected respondent subgroups. Respondents with group mean values that are higher than the group mean for all respondents are on average weighted up, while those with values lower than the group mean for all respondents are on average weighted down. For example, respondents in former East Germany are on average weighted down compared to those in the former West Germany. The percentiles and mean values of the subgroups, with the possible exception of respondents with Arabic/Muslim origin, do not differ to a great extent from the corresponding values from

all respondents taken together. This indicates that members of the chosen subgroups are not subjected to combinations of factors that lead to extreme up- or down-weighting.

Table 10: **Estimated Raw Non-Consent Weights for Selected Respondent-Groups**

	Min	10%	25%	50%	75%	90%	Max	Mean	SD
Total	1.22	1.44	1.57	1.60	1.78	2.10	7.47	1.73	0.38
Women	1.28	1.44	1.56	1.60	1.78	2.10	7.47	1.73	0.39
Age <= 30	1.28	1.44	1.60	1.64	1.88	2.13	7.47	1.78	0.42
Age >= 50	1.25	1.44	1.57	1.60	1.78	1.98	4.46	1.70	0.35
Educational degree – Compulsory education	1.34	1.48	1.60	1.76	1.88	2.13	5.08	1.79	0.39
Immigrant	1.22	1.44	1.56	1.60	1.78	2.12	7.47	1.73	0.38
Descendant of Immigrant	1.28	1.44	1.60	1.60	1.88	2.13	4.24	1.75	0.36
Originates from Arabic/Muslim country	1.45	1.70	1.78	1.96	2.24	2.39	7.47	2.13	0.65
Originates from country of the former USSR	1.25	1.41	1.56	1.60	1.78	1.98	4.17	1.70	0.35
Lives in former Eastern Germany	1.22	1.28	1.36	1.38	1.50	1.81	2.59	1.48	0.22

Note: Consent Weights are only calculated for respondents that actually consented. The 23 respondents that received weights of 0 are not included in the table.
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

The mean weights of all groups are higher than their corresponding median values (50 percent percentiles), showing that the distributions of their weights are positively skewed. This could be enforced by outliers with extremely high weights, as hinted by the maximum values of most groups, which are considerably higher than the 90 percent percentiles. Therefore, we used trimming to avoid extreme weights (a factor of 1.75 higher than the mean), which affects only 39 cases (for another example of weight-trimming, see Peytchev et al. 2011; for a critical view on trimming, see Van Goor and Stuiver 1998).

Consent weights were estimated to reduce the bias that occurs due to selection effects of the consent-decision. To test if this goal was accomplished, Table 11 displays in the first column the rate of various subgroups, such as women or persons of Arabic and Muslim origin, among all respondents that were confronted with the consent request (both consenters and rejecters). The second column presents the *unweighted* rates among consenters only, the third column the *weighted* rates among consenters only. Asterisks indicate if the group rates of consenters differ significantly from the group rates of consenters and non-consenters taken together. The 'Bias' is calculated as the difference between the rates of consenters as well as rejecters

and the rates of consenters only. Large percentage point values indicate selection effects for the particular (sub)group.

Table 11: **Bias Reduction**

Variable	Rate Consenter + Nonconsenters	Rate Consenters (unweighted)	Rate Consenters (weighted)	Bias# in Percentage Points (unweighted)	Bias# in Percentage Points (weighted)
Women	51.7%	51.9%	51.8%	0.2	0.1
Age <= 30	30.9%	31.4%	32.3%	0.5	1.3
Age >= 50	18.2%	17.5%	17.2%	0.7	1.0
Compulsory education	32.2%	31.6%	32.7%	0.6	0.5
Immigrant	75.0%	75.2%	75.3%	0.2	0.4
Descendant of Immigrant	16.8%	15.9%	16.2%	0.9	0.6
Originates from Arabic/Muslim country	6.5%	5.1%**	6.1%	1.4	0.4
Originates from country of the former USSR	29.0%	29.9%	29.6%	1.0	0.6
Lives in former Eastern Germany	10.1%	11.8%**	10.1%	1.6	0.1

Note: The 23 respondents that received weights of 0 are not included in the table.
 # Bias is calculated as the difference between the rates of consenters as well as rejecters and the rates of consenters only.
*** p<0.01*
Source: IAB-SOEP-Migration Sample, Waves 2013 and 2014. DOI: 10.5684/soep.iab-soep-mig.2013 and 10.5684/soep.iab-soep-mig.2014, own calculations.

For most groups, group means between consenters and consenters combined with non-consenters differ only slightly. In most cases, differences remain insignificant and bias does not reach one percentage point. Like the multivariate analyses, the univariate analyses indicate low selectivity of the consent decision. However, the table displays two divergences from this pattern: respondents from Arabic and Muslim countries are significantly underrepresented among consenters, while persons living in the former East Germany are significantly overrepresented. After weighting, both biases are reduced to insignificance. Weighting does not change other group-means in a significant way. Therefore, the generated

weights succeed in reducing bias for groups with selective consent decisions without disturbing non-selective groups.

5 Conclusion

The IAB-SOEP Migration Sample is the first empirical data source that provides linked survey and social security data on migrants in Germany, thus allowing for very detailed analyses of the integration processes in the German labor market. Such a linkage of data sources requires the signed consent of respondents, which we obtain from half of the respondents confronted with the request for the first time. The cumulative consent rate, including a second linkage request for some respondents who had refused the first request, amounts to about 58 percent.

About half of the variation in the decision to consent to the record linkage or not is attributable to the household level. Only smaller fractions of one quarter of the variation each can be located at the individual level or the interviewer / neighborhood level. This indicates next to a certain influences of interviewers a considerable importance of the household for individual consent decisions.

We tested for a wide range of effects from characteristics of respondents, households, neighborhoods and communities on the individual willingness to consent to record linkage. By and large, we found few systematic differences between consenters and non-consenters that may induce selectivity bias in applied analyses.

However, multivariate analysis revealed that a small number of characteristics differ significantly between consenters and non-consenters. For example, respondents with university degree were found to be more likely to consent, while respondents living in very large households were found less likely to do so. To counteract these selection effects, non-consent weights were estimated for every consenting respondent. The weights were shown to significantly improve univariate group-rates. Researchers are encouraged to apply the non-consent weights to their analyses.

As usual, non-consent weights can and should be combined with cross-sectional (design- and non-response-) as well as with longitudinal weighting variables. In practice, all combinations of weighting variables can be multiplied with each other to attain a combined weighting variable.

Finally, the implementation of the consent request in an experimental setting revealed that consent request did not inflate panel attrition nor did the late introduction of the consent request in wave 2 of the panel increase the propensity to consent to record linkages (Eisnecker and Kroh 2016).

6 Literature

Antoni, Manfred (2011). Linking Survey data with administrative employment data: The case of the German ALWA survey (Discussion Paper). Nuernberg, GER, Institute for Employment Research. http://doku.iab.de/fdz/events/2011/antoni.pdf.

Baker, Richard, Christopher Shiels, Keith Stevenson, Robin Fraser and Margaret Stone (2000). "What proportion of patients refuse consent to data collection from their records for research purposes?" British Journal of General Practice **50**: 655-656.

Beste, Jonas (2011). Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten. Nuernberg, GER, Institute for Employment Research. FDZ-Methodenreport 9/2011, http://doku.iab.de/fdz/reporte/2011/MR_09-11.pdf.

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

Dunn, Kate M., Kelvin Jordan, Rosie J. Lacey, Mark Shapley and Clare Jinks (2004). "Patterns of Consent in Epidemiologic Research: Evidence from Over 25,000 Responders." American Journal of Epidemiology **159**(11): 1087–1094.

Eisnecker, Philipp Simon and Martin Kroh (2016). "The Informed Consent to Record Linkage in Panel Study: Optimale Starting Wave, Consent Refusals, and Subsequent Panel Attrition." Public Opinion Quarterly, to be published [online first: 2016-10-12]. DOI: 10.1093/poq/nfw052.

German Institute for Economic Research Berlin. (2014a). "Early childhood education and care quality in the Socio-Economic Panel." Retrieved 11. November 2014, from http://www.diw.de/en/diw_01.c.430776.en.

German Institute for Economic Research Berlin. (2014b). "SOEP-LEE: The Linked Employer-Employee Survey of the Socio-Economic Panel Study." Retrieved 11. November 2014, from http://www.diw.de/en/diw_01.c.394066.en.

German Institute for Economic Research Berlin. (2014c). "SOEP and IAB create database to evaluate minimum wage." Retrieved 14. October 2015, from http://www.diw.de/en/diw_01.c.492395.en/topics_news/soep_and_iab_create_database_to_evaluate_minimum_wage.html.

Goebel, Jan, C. Katharina Spieß, Nils R. J. Witte and Susanne Gerstenberg (2007). "Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz." Data Documentation 26. DIW Berlin. http://www.diw.de/documents/publikationen/73/diw_01.c.78103.de/diw_datadoc_2007-026.pdf

Harris, T., D. G. Cook, C. Victor, C. Beighton, S. Dewilde and I. Carey (2005). "Linking questionnaires to primary care records: factors affecting consent in older people." J Epidemiol Community Health **59**(4): 336-338.

Hartmann, Josef and Gerhard Krug (2009). "Verknüpfung von personenbezogenen Prozess- und Befragungsdaten – Selektivität durch fehlende Zustimmung der Befragten?" Zeitschrift für Arbeitsmarkt Forschung **42**(2): 121-139.

Hockley, Christine, Maria Quigley, Jon Johnson, Rachel Rosenberg, Carol Dezateux and Heather Joshi (2007). Millennium Cohort Study: Birth Registration and Hospital Episode Statistics Linkage. A Guide to the Dataset. London, GB, Institute of Education. http://www.cls.ioe.ac.uk/library-media/documents/MCS_Birth_Registration_and_Hospital_Records_User_Guide_v4.pdf.

Huang, N., S. F. Shih, H. Y. Chang and Y. J. Chou (2007). "Record linkage research and informed consent: who consents?" BMC Health Serv Res **7**: 18.

Huber, Martina and Alexandra Schmucker (2009). "Identifying and Explaining Inconsistencies in Linked Administrative and Survey Data: The Case of German Employment Biographies." Historical Social Research **34**(3): 230-241.

Jenkins, Stephen P., Lorenzo Cappellari, Peter Lynn, Annette Jäckle and Emanuela Sala (2006). "Patterns of Consent: Evidence from a General Household Survey." Journal of the Royal Statistical Society. Series A (Statistics in Society) **169**(4): 701-722.

Kho, Michelle E., Mark Duffett, Donald J. Willison, Deborah J. Cook and Melissa C. Brouwers (2009). "Written informed consent and selection bias in observational studies using medical records: systematic review." British Medical Journal **338**: 1-8.

Klassen, Anne F, Shoo K Lee, Morris Barer and Parminder Raina (2005). "Linking Survey Data with Administrative Health Information." Canadian Journal of Public Health **96**(2): 151-154.

Knies, Gundi, Jonathan Burton and Emanuela Sala (2012). "Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population." BMC Health Services Research **12:52**.

Korbmacher, Julie M. and Mathis Schroeder (2013). "Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer." Survey Research Methods **7**(2): 115-131.

Kroh, Martin, Simon Kühne, Jan Goebel and Friedrike Preu (2015). The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment. Berlin, GER, German Institute for Economic Research. http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0271.pdf.

Oberschachtsiek, Dirk , Patrycja Scioch, Christian Seysen and Jörg Heining (2009). Integrated Employment Biographies Sample. Handbook for the IEBS in the 2008 version. Nürnberg, Forschungsdatenzentrum der Bundesagentur für Arbeit. 03/2009.

Peytchev, Andy, Lisa R. Carley-Baxter and Michele C. Black (2011). "Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse." Sociological Methods & Research **40**(1): 138-168.

Sakshaug, Joseph W., Mick P. Couper, Mary Beth Ofstedal and David R. Weir (2012). "Linking Survey and Administrative Records: Mechanisms of Consent." Sociological Methods & Research **41**(4): 535–569.

Sakshaug, Joseph W. and Frauke Kreuter (2014). "The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey." Public Opinion Quarterly **78**(1): 166–176.

Sala, Emanuela, Jonathan Burton and Gundi Knies (2012). "Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics." Sociological Methods & Research **41**(3): 414–439.

Silva, Marluce S, Wayne T Smith and Gabriele Bammer (2002). "The Effect of Timing When Seeking Persmission to Access Personal Health Services Utilization Records." Annals of Epidemiology **12**: 326–330.

Tate, A. Rosemary , Lisa Calderwood, Carol Dezateux and Heather Joshi (2006). "Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study." International Journal of Epidemiology **35**: 294–298.

TNS Infratest Sozialforschung (2014). Methodenbericht zum IAB-SOEP Migrationssample 2013. SOEP Survey Papers 217: Series B. Berlin, DIW/SOEP. http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0217.pdf.

Van Goor, H. and 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." Social Science Research **27**(4).

Wagner, Gert G., Joachim R. Frick and Jürgen Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements." Schmollers Jahrbuch **127**(1): 139-169.

Williams, Rick L. (2000). "A Note on Robust Variance Estimation for Cluster-Correlated Data." Biometrics **56**: 645-646.

Woolf, Steven H., Stephen F. Rothemich, Robert E. Johnson and David W. Marsland (2000). "Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research." Arch Fam Med **9**: 1111-1118.

Young, Anne F, Anette J Dobson and Julie E Byles (2001). "Health services research using linked records: who consents and what is the gain." Australian and New Zealand Journal of Public Health **25**(5): 417-420.