

# 1334<sup>2024</sup>

**SOEP** Survey Papers

Series C - Data Documentations (Datendokumentationen)

## Data quality control in the 6th wave of the IAB-BAMF-SOEP Survey of Refugees, samples M3-M6

Silvia Schwanhäuser and Lukas Olbrich

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

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

The SOEP Survey Papers is comprised of the following series:

**Series A** – Survey Instruments (Erhebungsinstrumente)

**Series B** – Survey Reports (Methodenberichte)

**Series C** – Data Documentation (Datendokumentationen)

**Series D** – Variable Descriptions and Coding

**Series E** – SOEPmonitors

**Series F** – SOEP Newsletters

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

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

#### **Editors:**

Dr. Carina Cornesse, DIW Berlin and University of Bremen

Dr. Jan Goebel, DIW Berlin

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

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

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

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

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

Please cite this paper as follows:

Silvia Schwanhäuser and Lukas Olbrich. 2024. Data quality control in the 6th wave of the IAB-BAMF-SOEP Survey of Refugees, samples M3-M6. SOEP Survey Papers 1334. Series C. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2024 by SOEP

ISSN: 2193-5580 (online)

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

soeppapers@diw.de

# Data quality control in the 6th wave of the IAB-BAMF-SOEP Survey of Refugees, samples M3-M6

**Silvia Schwanhäuser and Lukas Olbrich**

*Nuremberg, 2024*

# Data quality control in the 6<sup>th</sup> wave of the IAB-BAMF-SOEP Survey of Refugees, samples M3-M6

Silvia Schwanhäuser<sup>1</sup> and Lukas Olbrich<sup>1</sup>

<sup>1</sup> Institute for Employment Research (IAB)

## Introduction

In face-to-face surveys, interviewers take on many key tasks. They contact households and target persons, encourage them to participate in the survey and conduct standardised interviews. In some cases, however, interviewers may deviate from specified guidelines when performing this task, which in turn can affect the quality of the data. In order to minimise the risk of deviant interviewer behaviour in interviews, data quality controls are implemented throughout the entire survey period of the IAB-BAMF-SOEP Survey of Refugees (Brücker et al. 2017). This enables follow-up interviewer training and feedback discussions, and to identify anomalies at an early stage. Such controls have been in place since the first survey wave (see Kosyakova et al. 2019, Olbrich et al. 2020). Through the use of these quality controls, it was found that some interviews did not meet the high-quality standards of the survey. Several analyses indicated that a total of three interviewers had failed to carry out individual interviews properly. It was therefore decided to remove these interviews from the dataset. This decision was taken in close cooperation with the project partners (IAB, BAMF, and SOEP) as well as the survey organisation (infas). The following report documents this decision and presents the findings that led to this decision being taken.

## Analyses performed

Various analyses and control measures were used to monitor the quality of the interviews. These were implemented by both infas and the Institute for Employment Research (IAB), and include the following control measures:

### *(1) Analyses of statistical key figures from paradata*

- Entire duration per instrument and/or interview
- Duration of individual modules or time spent on screens
- (Screen) duration of vignette

## (2) Results of interviewer monitoring

- Evaluation of audio recordings of interviews, to identify deviant interviewer behaviour
- Results from postal interview control sheets
- Results from feedback discussions with field operations managers
- Number of interviews and/or surveyed households per interviewer
- Number of interviews and/or surveyed households per day
- Time of day/time of interviews
- Comparison of sample information with statements from the interview
- Review of delivered interviews and contacting process

## Findings

The above-mentioned monitoring procedures revealed suspicious results for three interviewers in particular (referred to in the following as interviewers A, B and C). Table 1 provides a list of the conspicuous control measures for these interviewers.

Table 1: Summary of conspicuous control measures

Interviewer	A	B	C
Number of person interviews	46	63	46
Number of household interviews	30	38	35
Number of interviews (all instruments)	83	135	103
Conspicuous overall duration	Yes	Yes	Yes
Conspicuous duration of individual modules	Yes	Yes	Yes
Conspicuous results overall monitoring by infas	Yes	Yes	-
Conspicuous control sheets/recordings	Yes	Yes	Yes
Many interviews in one day	-	Yes	-
Negative feedback from respondents	-	Yes	-

*Note:* The total number of interviews for all instruments includes instruments for individuals, households, children and young people.

Interviewer A showed conspicuous results in particular under the combined monitoring procedures performed by the survey organisation, including the assessment of the total duration of the interviews, the amount of time spent on each screen, number of interviews conducted per day, audio recordings, and feedback from the interview control sheets. This applies all the more when considering the relative share of person interviews that showed suspicious results. Overall, interviewer A conducted 30 household interviews and 46 person interviews. The

analysis of interview durations over the course of the fieldwork shows that durations became unrealistically short, particularly from the 14<sup>th</sup> interview onwards, with significant deviations from the average duration of other interviewers. The analysis of the audio recordings by infas also revealed that the interviewer deviated significantly from the standardised interviewing process. Hence, interviews were not completely fabricated, but the interviewer did not follow proper interviewing procedures. As a result, all the interviews conducted by this interviewer after the 14<sup>th</sup> person interview were removed from the dataset.

Interviewer B showed the most conspicuous results and suspicious values for all the control measures. Analyses of the respective durations of interviews during the fieldwork showed that many interviews were unrealistically short. Particularly from the 13<sup>th</sup> interview onwards, the number of conspicuous durations steadily increased. In addition, there were only very few audio recordings available for in-depth monitoring, which can also be interpreted as an indication of problems with the interviews. It was also conspicuous that the interviewer conducted many interviews in one day or several interviews in areas with different district codes (Gemeindekennziffer, GKZ). Furthermore, respondents gave negative feedback in the control sheets. As a result, all the interviews conducted by this interviewer after the 13<sup>th</sup> person interview were removed from the dataset. Overall, this interviewer conducted 38 household interviews and 63 person interviews.

Interviewer C initially showed no conspicuous results in the first analyses of durations. From the 15<sup>th</sup> person interview, the duration per interview became similarly unrealistic, as with interviewer B. There were also very few audio recordings available for more detailed examinations. Overall, this interviewer conducted 35 household interviews and 46 person interviews. There were no conspicuous results when it came to the number of interviews in a day, in several areas or negative feedback from respondents. However, due to the unrealistic duration of the interviews, all interviews conducted by this interviewer after the 15<sup>th</sup> person interview were removed from the dataset.

A total of 87 households with 160 respondents were therefore subsequently excluded from the survey. Table 2 shows the number and shares of excluded interviews for the respective instruments of the 6<sup>th</sup> wave of the IAB-BAMF-SOEP Survey of Refugees.

Table 2: Number and share of excluded interviews by survey instrument

<b>Instruments</b>	<b>Excluded interviews</b>		
	N	Share in SOEP	Share in IAB-BAMF-SOEP
<i>Household</i>	87	0.6 %	2.2 %
<i>Person</i>	160	0.8 %	6.2 %
<i>Young people</i>	6	1.6 %	7.5 %
<i>Children</i>	168	1.9 %	5.3 %

## References

Brücker, Herbert, Nina Rother & Jürgen Schupp (2017): IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. (IAB-Forschungsbericht 13/2017), Nürnberg.

Kosyakova, Yuliya, Lukas Olbrich, Joseph Sakshaug & Silvia Schwanhäuser (2019): Identification of interviewer falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany. (FDZ-Methodenreport 02/2019 (en)), Nürnberg.

Olbrich, Lukas, Yuliya Kosyakova, Joseph Sakshaug & Silvia Schwanhäuser (2020): Interviewerkontrolle in der 4. Welle der IAB-BAMF-SOEP-Befragung von Geflüchteten, Stichproben M3-M5. (SOEP Survey Papers. Series C, No. 901), Berlin.