RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

FDZ-METHODENREPORT

Methodological aspects of labour market data

02|2019 EN Identification of Interviewer Falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany

Kosyakova Y., Olbrich L., Sakshaug J.W., Schwanhäuser S.

Bundesagentur für Arbeit

Identification of Interviewer Falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany

Kosyakova Y. (Institute for Employment Research, Nuremberg)
Olbrich L. (Institute for Employment Research, Nuremberg)
Sakshaug J.W. (Institute for Employment Research, Nuremberg and University of Mannheim, Germany)
Schwanhäuser S. (Institute for Employment Research, Nuremberg and University of Mannheim, Germany)[1]

[1] All authors contributed equally and are listed in alphabetical order. Corresponding author: Yuliya Kosyakova. E-Mail for correspondence: Yuliya.Kosyakova@iab.de; Lukas.Olbrich@iab.de; Joe.Sakshaug@iab.de; Silvia.Schwanhaeuser2@iab.de

Contents

Abstract

Interviewer-administered surveys are susceptible to intentional deviant behavior by interviewers and this type of behavior is a potential source of survey error. One example of such deviant behaviour is the falsification of entire interviews, which can negatively impact data quality if such cases are not identified. Therefore, the development and application of falsification detection methods is important to ensure high quality data. However, methods of detecting falsification are usually evaluated using simulated or laboratory data instead of actual falsified data. This report examines the effectiveness of statistical identification methods for detecting falsifiers in the IAB-BAMF-SOEP Survey of Refugees in Germany, in which there was an actual case of interviewer fraud.

Zusammenfassung

Interviewer-geschützte Befragungen sind potentiell anfällig für bewusstes Fehlverhalten durch den Interviewer. Dies stellt eine mögliche Quelle für Verzerrungen in Befragungsdaten dar. Ein Beispiel für diese Art von Verhalten ist die Fälschung von Interviews durch den Interviewer. Werden solche Fälle nicht identifiziert, kann sich dies negativ auf die Datenqualität auswirken. Daher ist die Entwicklung und Anwendung von Methoden zur Identifikation von Fälschungen wichtig, um qualitativ hochwertige Daten zu gewährleisten. Bisher wurden derartige Methoden meist unter Verwendung von simulierten oder experimentellen Daten anstelle von tatsächlich gefälschten Daten evaluiert. In diesem Bericht wird das Potential statistischer Identifikationsmethoden anhand der Daten der IAB-BAMF-SOEP Befragung Geflüchteter untersucht, für welche ein Fall von Interviewfälschungen durch das Befragungsinstitut aufgedeckt wurde.

Keywords: interviewer falsification, statistical identification methods, data quality.

1 Introduction

Deviant interviewer behavior is a well-known problem in survey research. Such behavior can take various forms. For example, interviewers may deviate from strictly standardized interviewing by failing to read the survey questions as worded or by providing unscripted feedback to respondents (Blasius and Friedrichs 2012; Bredl, Storfinger, and Menold 2013; Harrison and Krauss 2002). Another example is when interviewers deliberately manipulate responses to filter questions in order to avoid follow-up questions and, thus, reduce the length of the interview (e.g. Kosyakova, Skopek, and Eckman 2015; Kreuter et al. 2011). While both of these examples of interviewer deviance can adversely affect the quality of the collected survey data, neither of them does this while generating as much publicity in the mass media as interviewer falsification of entire interviews. We refer to interviewer falsification as the intentional deviation from the intended data collection guidelines or instructions, which remains unreported by the interviewer (see AAPOR 2003); specifically, we focus on the fabrication of complete interviews. The empirical evidence suggests that deliberate interviewer falsification is a rare occurrence in survey research (see Blasius and Friedrichs 2012; Bredl et al. 2013). Nevertheless, even small amounts of unidentified cases can, to some extent, have a negative impact on data quality and introduce bias in survey estimates (see Schnell 1991; Schräpler and Wagner 2005). Therefore, detecting fraudulent interviews is an important task for ensuring high data quality.

A recent example of interviewer falsification was reported in the IAB-BAMF-SOEP Survey of Refugees in Germany (IAB 2017; DIW 2018). The partners conducting the survey – the German Institute for Employment Research (IAB), the Federal Office for Migration and Refugees (BAMF), and the German Socio-Economic Panel (SOEP) – were informed in December 2017 by the survey institute that one of their interviewers falsified 289 person interviews in 217 reported households during the baseline wave of the study (conducted during the period from June to December 2016). The fabricated interviews from the first wave were discovered after the field period for the second wave began. Meanwhile, the baseline data had been released (V33, on 29.11.2017) and preliminary results had been published and presented to the public. Revision of the data followed on 30.01.2018 (V33.1). The discovery of interviewer falsifications and its consequences were prominently featured in a news article published in Der Spiegel (Seibt 2018a) alongside a series of related articles on the topic of “fake surveys” and the manipulation of opinion polls especially in market research (e.g. Kwasniewski et al. 2018; Seibt 2018b).

In light of this high-profile case, it is evident that survey organizations are in need of improving their tools to identify potential interviewer falsification. This report contributes to the development of such tools through a case study aimed at ensuring high data quality. Specifically, this case study aims to: 1) identify potential predictors of interviewer falsification; 2) extend existing methods for detecting possible interviewer falsification; and 3) develop and test new methods for detecting interviewer falsification. These aims are achieved by using the known fraudulent data from the IAB-BAMF-SOEP Survey of Refugees in Germany, for which the authors have access to as part of their employment at the IAB[2]. This case study allows for the retrospective identification of fake interviews and testing of various fraud detection tools. These tools, as we

[2] External researchers may apply for access to these data by submitting a user-contract application.

will show later, were successful in identifying the known falsifier and identifying additional suspicious cases of potential interviewer fraud, which were previously undiscovered in the IAB-BAMF-SOEP Survey of Refugees in Germany.

2 IAB-BAMF-SOEP Survey of Refugees in Germany

The IAB-BAMF-SOEP-Refugee Survey is a recently-established longitudinal household survey in Germany (Brücker et al. 2016; Brücker, Rother and Schupp 2017; Brücker et al. 2019) in which a single case of interviewer fraud was identified by the survey institute almost one year after the data was collected. The first wave was launched in 2016. The gross sample size amounted to 7,101 households of which 3,554 participated in the survey (Kroh et al. 2017). With a response rate of about 48.7 percent, the first data release included 4,816 respondents (in 3,554 households), interviewed by 98 interviewers. The median number of completed interviews per interviewer was 85 (mean = 108 interviews). There were three interviewers who each interviewed more than 200 respondents; these interviewers conducted 17.3 percent of all interviews. The lone fraudulent interviewer reported by the survey institute was one of these interviewers. This interviewer conducted 289 person interviews (6.0 percent of all person interviews) and 217 household interviews (6.1 percent of all household interviews).

The second wave of the survey was carried out in 2017. The re-interview rate was 67% among the first wave participants (Brücker et al. 2019). In addition, a refreshment sample was added in 2017. In total, these two samples yielded data for 2,747 panel respondents (in 2,166 households) and 2,974 new respondents (in 2,134 households). 48 interviewers conducted interviews in the second wave: the median number of interviews per interviewer was 147 (mean = 180 interviews). There were six interviewers who each completed over 200 interviews, two of which completed over 300 interviews. In total, these two interviewers conducted 38.18 percent of all interviews.

Interviews were conducted using computer-assisted personal interviewing (CAPI) with audio computer-assisted-self-interviewing (ACASI) used for a subset of questions. The questionnaires were available in seven languages (Arabic, English, Farsi/Dari, German, Kurmanji, Pashtu and Urdu). Auditory instruments (spoken audio files) were implemented to facilitate interviews with individuals with poor reading skills. If needed, other persons (such as family members) and, to a lesser extent, professional translators and interpreters were also used as language mediators.

The core topics of the (person and household) questionnaires included migration, education and employment biographies, history and process of refugee migration, registration in Germany, asylum procedure, accommodation, human capital, language proficiency, cognitive and non-cognitive competences, personality traits and attitudes, health and life satisfaction, participation in integration measures and labor market integration in Germany.

3 Identification strategy

3.1 Indicator-based identification strategies

In the first step of our analyses, we identified potential indicators of interviewer falsification guided by the survey literature (Stokes and Jones 1989; Hood and Bushery 1997; Murphy et al. 2004; Bredl et al. 2013; Blasius and Friedrichs 2012; Menold et al. 2013; de Haas and Winker 2016; Murphy et al. 2016). These indicators attempt to distinguish between honest and deviant interviewing behaviors as well as distinguish between the responding styles of real and fake respondents. In addition, we include new indicators, which are partly specific to the survey (such as the proportion of record-linkage consent and interviewers' evaluation of the interview situation) and potentially available in other surveys (such as the relative duration of the interview). Table 1 lists all of the indicators analyzed in this study. Existing indicators include citations where they have been used in previous research, while newly considered indicators are correspondingly marked as such. The choice of indicators is largely motivated from the notion that interviewers fabricate interview data in a way that minimizes their time and effort. The main assumption is that in the case of fraudulent interviewer behavior the indicators will point in a specific direction. For instance, the falsifier can influence the number of questions presented by strategically answering filter questions in a way such that follow-up questions are not triggered (Brüderl et al. 2013). Employing this strategy may significantly reduce time and effort (Kosyakova et al. 2015). Another example is the frequency of extreme responses in rating scales. Since extreme answers make it difficult to answer all questions in a consistent manner and can attract attention, a falsifier is likely to avoid extreme responses (Porras and English 2004). Table 1 summarizes the underlying assumptions of the falsification direction for each indicator.

For the empirical analysis, the indicator values resulting from single interviews were aggregated to the interviewer-level and then standardized. In this way, we ensure the comparability of the values resulting from different indicators. This transformation can be realized via the following formula:

$$z = \frac{x - \mu}{\sigma}, \qquad (1)$$

where x denotes the original interviewer-level indicator, μ and σ the mean and the standard deviation of x, respectively, and z the resulting standardized indicator. In total, 32 standardized interviewer-level indicators were created. These indicators were then introduced in three different types of analyses as described in the following.

First, we calculated the share (percentage) of standardized indicator values that lay in a suspicious direction for each interviewer. Based on assumptions about the rational behavior of falsifiers, we marked one direction of the standardized values for each indicator as suspicious. As an example, we expect falsifiers to show less item non-response. Therefore, a relatively large standardized negative value for the item non-response indicator is regarded as suspicious. More examples can be found in Table 1.

Table 1. List of interviewer-level indicators for identifying interviewer falsification

Indicator	Description	Assumed direction of falsifiers	References
Acquiescent Responding Style	Share of positive connotation ("Agree/Strongly Agree") independent of question content	Lower share of positive connotation independent of question content for falsifiers	Messick (1966), Menold et al. (2013)
Benford's Law	Decreasing distribution of leading digit for numeric quantities	Poor fit to Benford's distribution to first digits for falsifiers	Swanson et al. (2003)
E-Mail	Share of E-Mail address provision	Lower share of provided e-mail addresses for falsifiers	NEW
Extreme responses	Share of extreme responses to rating scales	Lower share of extreme responses to rating scales for falsifiers	Schäfer et al.(2005)
Filter questions	Share of responses which lead to follow-up questions	Lower share of responses which lead to follow-up questions for falsifiers	Hood and Bushery (1997), Kosyakova et al. (2015)
Interview duration	Duration of completed interviews	Shorter duration of completed interviews for falsifiers	Hood and Bushery (1997)
Interviewer evaluation	Interviewer's evaluation of the interview situation	Very positive evaluation of the interview situation for falsifiers	NEW
Item nonresponse	The item nonresponse rate within an interviewer's workload	Lower item nonresponse rate for falsifiers	Schäfer et al. (2005)
Middle category responses	Share of the middle response in rating scales	Higher share of middle responses on rating scales for falsifiers	Schäfer et al. (2005)
Non-Differentiation	Standard deviation across item scales	Lower standard deviation across item scales for falsifiers	Reuband (1990)
Primacy effects	Share of choosing the first two categories in non-ordered answer option lists	Higher share of choosing the first two categories in non-ordered answer option lists for falsifiers	Krosnick and Alwin (1987), Menold et al. (2013)
Recency effects	Share of choosing the last two categories in non-ordered answer option lists	Lower share of choosing the last two categories in non-ordered answer option lists for falsifiers	Krosnick and Alwin (1987), Menold et al. (2013)
Record linkage consent	Share of consent to record linkage	Higher share of consent to record linkage for falsifiers	NEW
Relative interview duration	Duration of completed interviews relative to the triggered questions	Shorter duration of completed interviews relative to the triggered questions for falsifiers	NEW
Rounding	Share of rounding numbers in numerical open-ended questions	Lower share of rounded numbers in numerical open-ended questions for falsifiers	Tourangeau et al. (1997), Menold et al. (2013)
Semi-Open responses	Share of responses to "other" in semi-open-ended question	Lower share of responses to "other" in semi-open-ended question for falsifiers	Hood and Bushery (1997)
Stereotyping	Strength of stereotypical response to attitudinal items	Higher strength of stereotypical response to attitudinal items for falsifiers	Reuband (1990)
Telephone number	Share of telephone number provision	Lower share of provided telephone numbers for falsifiers	Stokes and Jones (1989)
Response variance	Standard deviation of responses between interviews	Lower standard deviation of responses between interviews for falsifiers	Porras and English (2005)

Source: Own literature research.

Second, we reverse coded the standardized indicator values, in order to guarantee that all positive values lie in the suspicious direction. Then we summed up the values across all indicators for each interviewer in order to create one overall interviewer-level indicator, which we denote as the meta-indicator. By reverse coding the individual indicators, we ensure that the suspicious values cannot offset each other when they are summed up. Values of the meta-indicator that lie at the right tail-end of its distribution are therefore deemed suspicious.

Third, we used the standardized indicator values in a cluster analysis. The basic idea of cluster analysis is to identify distinct subgroups (or clusters) of available elements that share similarities on a set of multivariate measurements. In the context of interviewer falsification, the goal is to identify distinct clusters – honest interviewers and dishonest interviewers – based on the presented multiple indicators of falsification (e.g. see Table 1). The idea of using cluster analysis for identification of interviewer falsification was first introduced by Bredl et al. (2012) and successfully used in applications involving simulations, laboratory data, and small data sets (e.g. Bredl et al. 2012; de Haas and Winker 2016). Applications involving actual falsifications in large-scale survey settings are rare. For our cluster analyses, we first employ a dissimilarity measure, namely, the Euclidean distance, to assess the dissimilarity distance between interviewers. Using this measure, the elements (interviewers) are grouped on the basis of the resulting distance matrix. We use hierarchical-agglomerative methods, including Single-Linkage and Ward's-Linkage (see also, Menold et al. 2013; Storfinger and Winker 2013) to carry out the grouping process.

3.2 Variation of indicators over interviewers' field experience

In the second step of our analyses, we conduct a rather novel examination of patterns in the falsification indicators over the course of the field period. This analytical strategy permits a more detailed analysis of how potential fraudulent behavior unfolds over time (i.e. from interview-to-interview). Here, we expect the falsifiers to set the optimal falsification effort for each interview based on their perception of the survey organization's data control processes. The higher this perception, the higher the effort invested in falsifications. Over the field period, the perception can change since a falsifier receives new information about the control processes after each undetected falsified interview, namely, that their deviant behavior has gone unnoticed by the survey institute. This information signals that the falsification effort was sufficient to avoid detection which indicates that the control processes were overestimated. We assume that the falsifier has two options to react to this signal. First, he or she can ignore it and stick with their initial perception. In this case, the falsifier trusts the prior information and assumes that the signal will only bias his or her perception. As a result, the optimal falsification effort is likely to remain constant over the field period. Moreover, this type of falsifier will avoid presumably suspicious deviations from the optimal strategy to avoid detection, which leads to a reduced variance in the falsification effort. The second option is putting value on the signal received after each falsified interview. Hence, this type of falsifier continuously updates his or her perception of the control processes and adapts the optimal falsification effort. In this case, the perception is decreasing over the field period since successful falsifications signal limited control processes by the survey organization. Consequently, the optimal falsification effort is likely to decrease after each falsified interview.

To summarize, falsifiers are expected to show either (a) reduced variance in their falsification effort or (b) decreasing falsification effort over the field period. By developing measures of falsification effort, this allows for distinguishing falsifiers from normal interviewers whose values for the effort measure are expected to vary randomly.

3.3 Content-related patterns

In the third and final step of our analyses, interviewers with the most suspicious patterns revealed by the above analyses are further examined with respect to (1) content-specific abnormalities in the responses of the interviewees and (2) the panel response rate (i.e. re-interview rate in the second wave).[3]

In the case of content abnormalities, we compare the mean values of responses for a variety of survey questions between all suspicious and unsuspicious interviewers.[4] We focus on specific topics for which it is less likely that an interviewer would be able to fabricate responses that would resemble actual distributions in the heterogeneous population of refugees, such as vocational education or university studies, participation in integration courses, and specific skills. We expect serious deviations from the mean responses for the dishonest interviewers compared to all other interviewers.

In the case of the panel response rate, we make a close examination of the response rate for the respondents potentially affected by the interviewer fraud in the second wave of the data. We expect higher rates of nonresponse in the follow-up wave for respondents interviewed in the first wave by the suspicious interviewers compared to all other respondents. Our rationale is that respondents interviewed by the suspicious interviewers in the first wave may claim that they were not interviewed in that wave when approached for the second wave and may simply refuse to participate in light of the misunderstanding.

4 Empirical Results

4.1 Indicator values

In order to identify falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany, we considered the indicators listed in Table 1: acquiescent responding style, Benford's law, contact information (e-mail and phone number), extreme responses, filter questions, interview du-

---

[3] In both analyses, we focus mainly on the comparison of suspicious and unsuspicious interviewers identified in the previous analyses .

[4] Generally, it is quite likely that some individual characteristics depend on, for instance, regional differences. Since interviewers are bound to regional clusters (PSUs), deviations could also be due to regional differences and not due to fraudulent interviewer behavior. However, in Germany refugees are subject to national dispersal policies and are bound to the place of their first allocation, at least until their refugee status is recognized. Given this exogenous residence assignment and the fact that a high share of respondents were still waiting for the decision on their asylum claim, it is less likely that the data disparities observed between suspicious and unsuspicious interviewers are due to regional (self-)selection of the respondents.

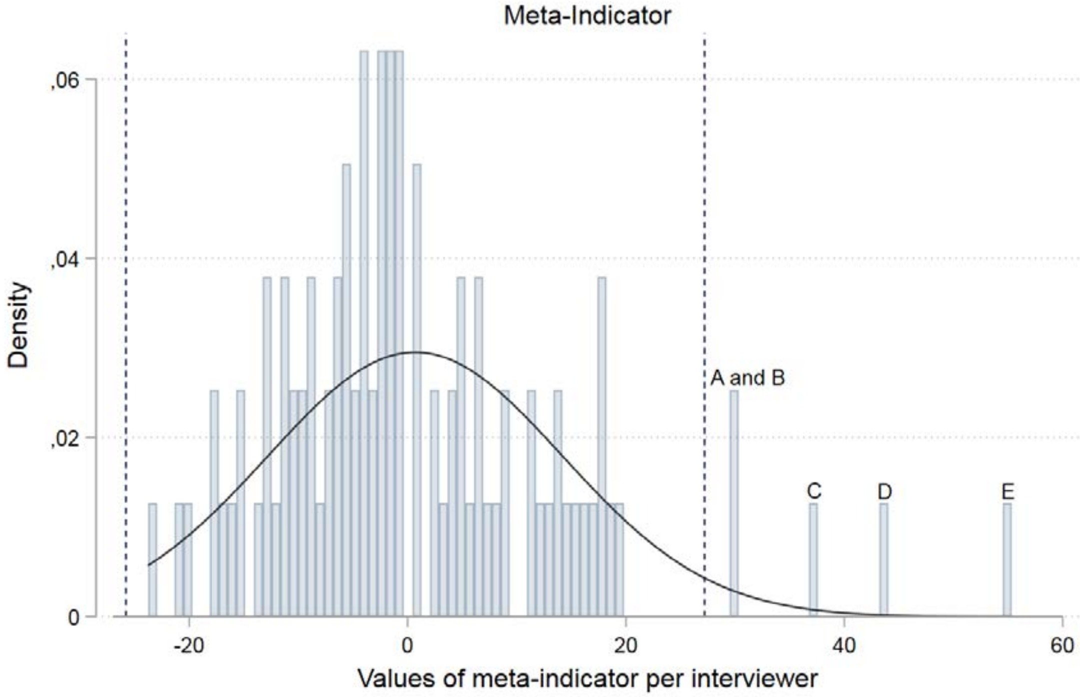
ration, interviewer evaluation, item nonresponse, middle responses, non-differentiation, primacy effects, recency effects, record-linkage consent, relative interview duration, rounding, stereotyping, semi-open responses, and response variance.

For the first analysis, we investigated the extent of suspicious indicator values per interviewer. As described above, interviewers with a very high share of suspicious indicator values were flagged as such. For the confirmed falsifier (hereafter referred to as interviewer A), 87.5% of all indicators (n=32) were suspicious (i.e. the standardized values pointed in the expected direction). Hence, we may conclude that this method is useful for identifying a known falsifier. Only one interviewer (hereafter referred to as interviewer C) shows a higher share (90.7%).

4.2 Meta-indicator

For the meta-indicator analysis, we aggregated the standardized indicator values for each interviewer. The aggregated values of all indicators resemble a right-skewed normal distribution (Figure 1). We consider the extreme right-most values, above the 97.5th percentile, to be outliers and thus as potential falsifiers. Only interviewers with a very small number of cases may be seen as exceptions as these values have greater uncertainty attached to them.

Figure 1. Interviewer-level meta-indicator with aggregated values of all indicators.



Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

Figure 1 shows that the aggregated value of all indicators for the confirmed falsifier (interviewer A) lies within the suspicious range and has the 5th highest meta-indicator value (29.53). Further suspicious interviewers have meta-indicator values of 30.08 (interviewer B), 37.08 (interviewer C), 43.82 (interviewer D), and 55.33 (interviewer E). However, any outlying value should be interpreted with the interviewer’s workload in mind: for instance, interviewer D and interviewer

E conducted only 1 and 2 interviews,[5] respectively. Due to the small number of interviews and great uncertainty regarding potential fraudulent behavior, both interviewers are excluded from the pool of suspicious interviewers. Since indicator values are first measured at the interview level and afterwards aggregated to the interviewer level, they have high uncertainty for interviewers with a low workload. Suspicious values can therefore easily result from single outlying respondents.

4.3 Cluster analyses

In the next step, we tested all indicators for systematic deviations using cluster analysis. Two different clustering algorithms were used: Ward's-Linkage and Single-Linkage. For both algorithms, the optimal clustering solution was determined by considering dendrograms as well as the Duda-Hart Je(2)/Je(1) index. Following this, we evaluated the created cluster groups for suspicious indicator values.

The Single-linkage algorithm, in particular, is useful to separate the most deviant observations from the rest of the data, since most similar observations are fused first (cf. Backhaus et al. 2016: 480). Table 2 shows the list of identified clusters according to the Single-Linkage clustering method, with cluster 1 containing the group of unsuspicious interviewers and clusters 2-7 each containing a single outlying (suspicious) interviewer.

Table 2. List of clusters identified by the Single-Linkage cluster analysis[6]

Clusters	Number of interviewers per cluster	Interviewer ID	Number of conducted person interviews	Number of conducted household interviews
Cluster 1	92	-	-	-
Cluster 2	1	B	46	34
Cluster 3	1	C	16	13
Cluster 4	1	A	289	218
Cluster 5	1	E	2	2
Cluster 6	1	D	1	1
Cluster 7	1	F	1	1

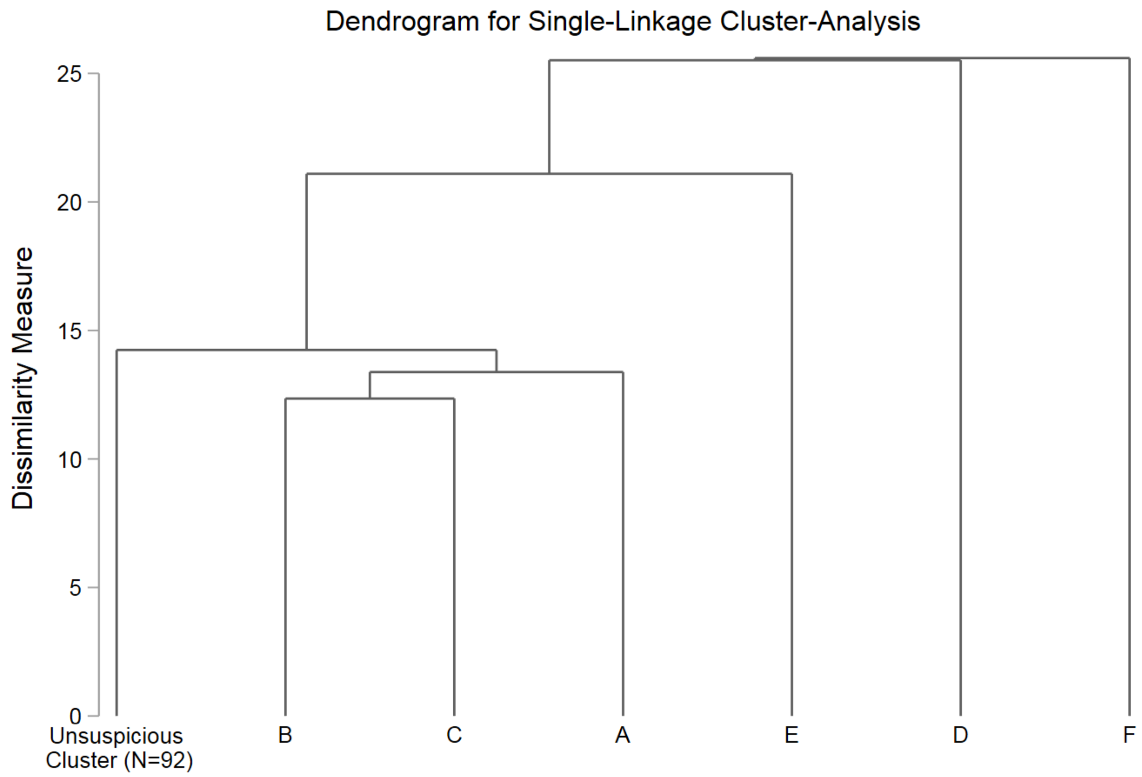
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

The set of interviewers identified as suspicious in the cluster analysis are similar to the set of suspicious interviewers identified by the meta-indicator. As can be seen in Figure 2, interviewers A, B, and C are similarly distinguishable from the majority of interviewers. As a reminder, interviewer A was the confirmed falsifier – which indicates that the clustering method successfully distinguished the behavior of this known case of interviewer fraud. The cluster results for the second and third suspicious interviewers (B and C) also support the findings from the meta-indicator analysis and provide further support for potentially fraudulent behavior. As previously mentioned, the interviewer’s workload should be kept in mind while interpreting the results. Therefore, the remaining three suspicious interviewers (D, E, and F) should not be considered as suspicious due to their small number of interviews conducted ($N \leq 2$).

[5] Table A1 in the Appendix provides the number of person and household interviews for each interviewer referred to in the analysis.

[6] For visualization purposes, only results of the single-linkage method are presented.

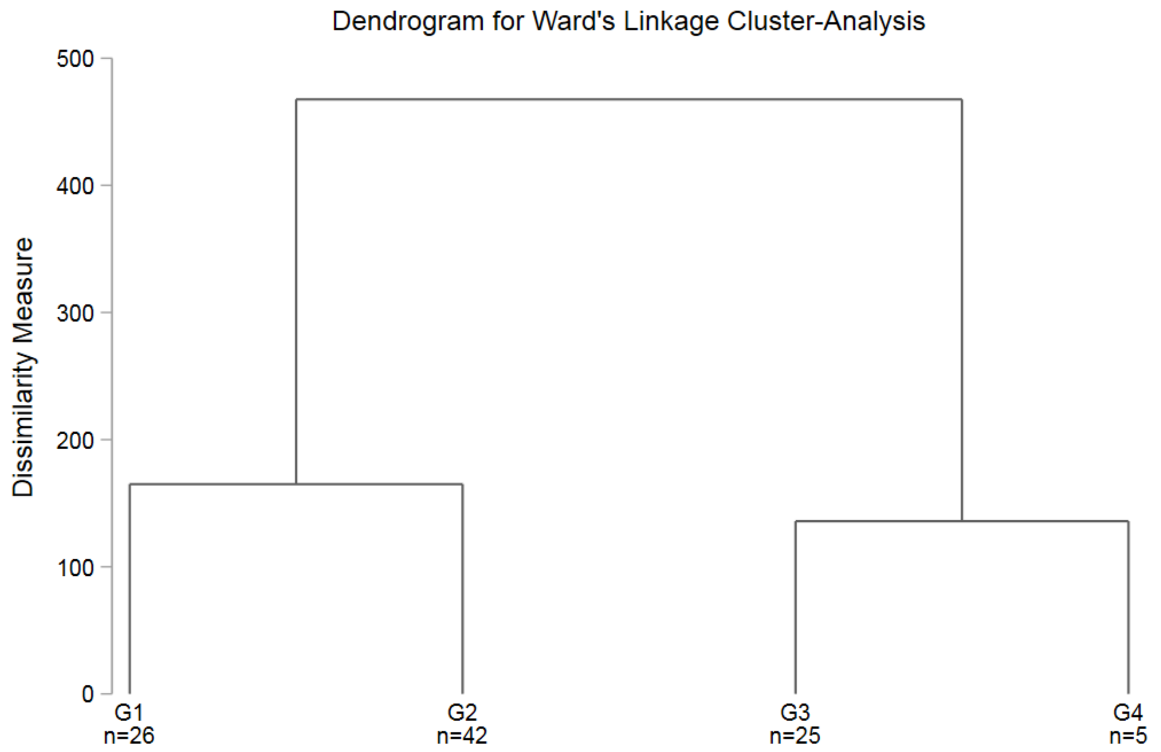
Figure 2. Dendrogram for optimal cluster solution (Single-Linkage)



Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

Ward's Linkage, in turn, fuses those cases/clusters that increase the variance within a group as little as possible, so that homogeneous clusters are formed. Since dishonest interviewers are likely to show relatively similar patterns, Ward’s Linkage should result in at least one small cluster, which is different to the others. Figure 3 reveals such a case with Cluster 4 which contains five interviewers. Interviewers included in this Cluster are the same interviewers identified in all previous methods: interviewers A, B, C, D and E. Given that all conducted analyses point to abnormalities for interviewers B and C and are very similar to the abnormalities of the confirmed falsifier, this can be regarded as strong evidence for further cases of interviewer fraud.

Figure 3. Dendrogram for optimal cluster solution (Ward's Linkage)



Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

4.4 Development of the indicators over the field period

Following the examination of aggregated indicator values, we now turn to their development over the course of the interviewers' field period. For these analyses, we focus on indicators that are associated with a falsifier's effort to fake a single interview. Therefore, indicators based on the interviewer's overall workload (e.g., E-Mail or telephone number provision, record linkage consent) or response patterns which are relatively independent of effort (e.g., acquiescent responding style, primacy effects, recency effects) are excluded from this part of the analysis. Hereinafter, the used indicators and their presumed influence on the falsifier's effort are explained as follows:

i.) The frequency of choosing extreme-value response categories for scale questions: extreme responses make it difficult to answer all remaining questions in a consistent fashion and therefore suggest suspicious behavior.

ii.) The frequency of choosing middle-response categories for scale questions: the choice of the middle response option facilitates consistent and rapid answering of the questionnaire and thus attracts less attention.

iii.) Triggered follow-up questions: the falsifier can affect the number of questions posed by strategically answering filter questions in a way that the follow-up questions are avoided. The corresponding indicator indicates the share of follow-up questions that have been answered.

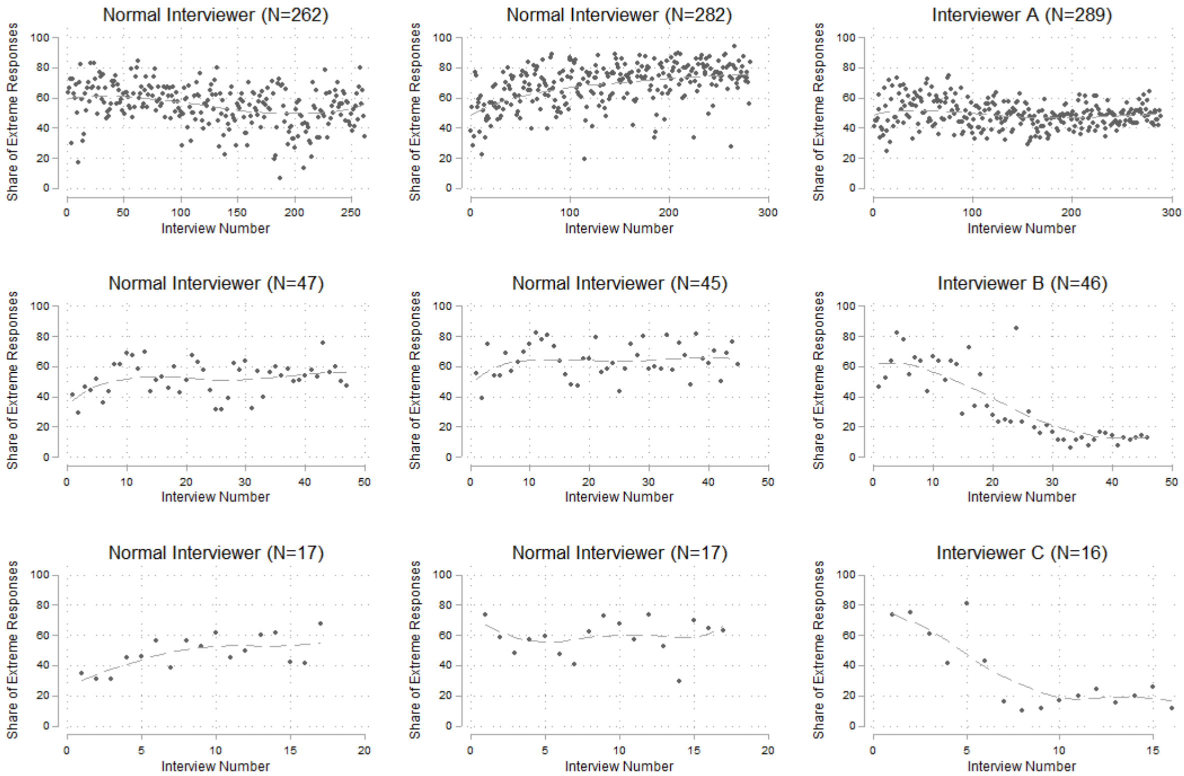
iv.) Rounding: rounded values require less effort from a falsifier's perspective compared to providing non-rounded, exact values.

v.) Non-Differentiation: lower variation on same-scale item batteries enables the falsifier to answer fast and in a consistent manner.

For brevity, we show these analyses only for the interviewers who revealed deviant behavior based on the previously-used indicator-based identification strategies (share of suspicious indicators, meta-indicator analysis, and cluster analysis). In the following figures, we present the development of the five above-listed indicators for the confirmed falsifier (interviewer A) and the other two suspicious interviewers identified by the previous analyses (interviewers B and C) in Sections 4.1, 4.2, and 4.3. For visualization purposes, we compare each of these interviewers with arbitrary interviewers who conducted similar numbers of interviews. Accordingly, we provide nine diagrams for each indicator. In each row, the first and second diagrams show unsuspicious interviewers and the third diagram shows the confirmed falsifier or one of the other two suspicious interviewers. The values of the indicators are plotted on the Y-axis and the interview number (sorted from early-to-late) is on the X-axis. For example, the interview number 20 denotes that this is the twentieth interview conducted by the respective interviewer.

For the extreme response indicator (see Figure 4), Interviewer A's average share does not change over time. However, we do observe a decreasing dispersion over the field period. This is not the case for the arbitrarily chosen interviewers. Furthermore, the other two suspicious interviewers (B and C) differ substantially from the comparators. At the beginning of the field period, their values are similar to those of the other interviewers, but then reveal a strong negative slope. A closer look at interviewer B starting around the 20th interview shows that the extreme responses fall continuously with each additional interview and the variance of this indicator decreases significantly. In turn, we do not find a similar trend for his/her counterparts. For interviewer C, the share of extreme responses decreases from the 6th interview and remains on a rather low level. The indicator values of the arbitrary interviewers fluctuate in the same interval as the previous comparators.

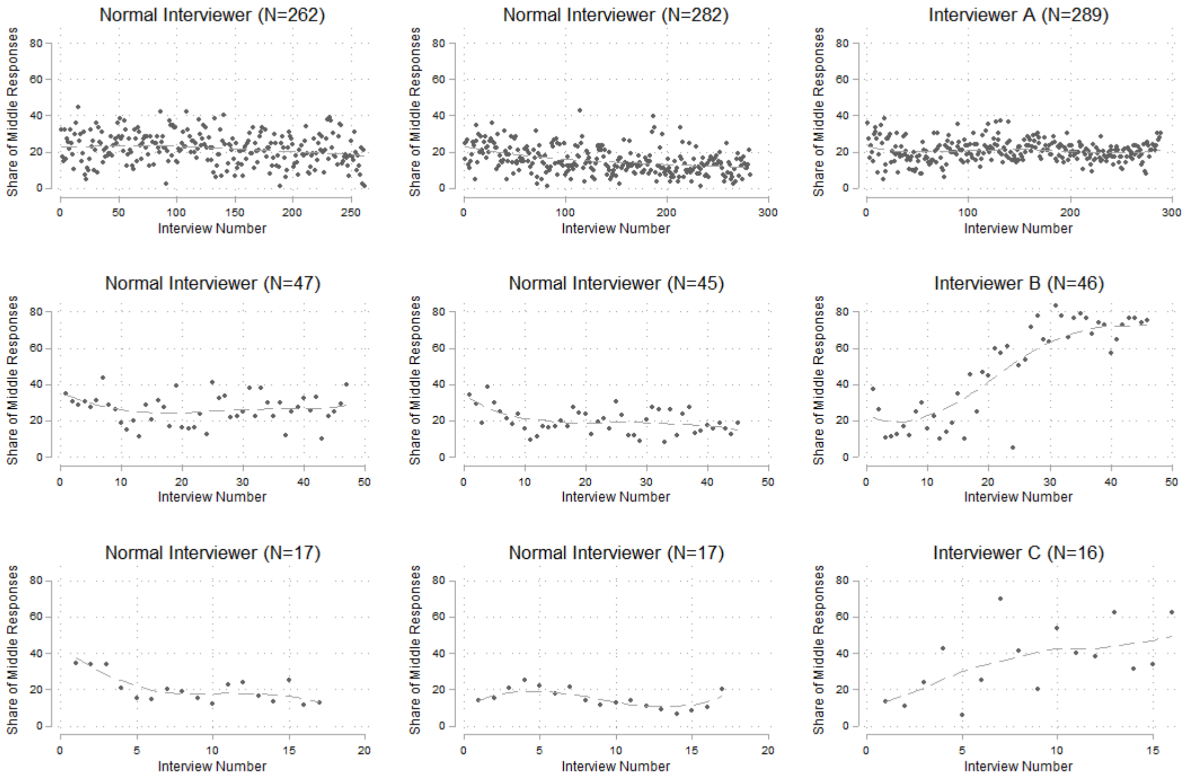
Figure 4. Indicator values for extreme responding style over the course of the field period



Notes: Indicator values are calculated as the share of extreme responses to scale questions in all answered scale questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

A similar pattern emerges for the share of middle-category answers (see Figure 5). The confirmed falsifier has much smaller variance in their share of middle-category answers (var = 35.6) compared to the two arbitrary interviewers (N=282, var = 55.8; N=262, var = 73.8). For suspicious interviewer B, the share of such answers increases significantly over time and then remains at a high level. In turn, there is no such trend for the arbitrary interviewers – the values fluctuate around 20 percent. None of them reaches values similar to the suspicious interviewer. For interviewer C, we see a positive slope and rather extreme values as well.

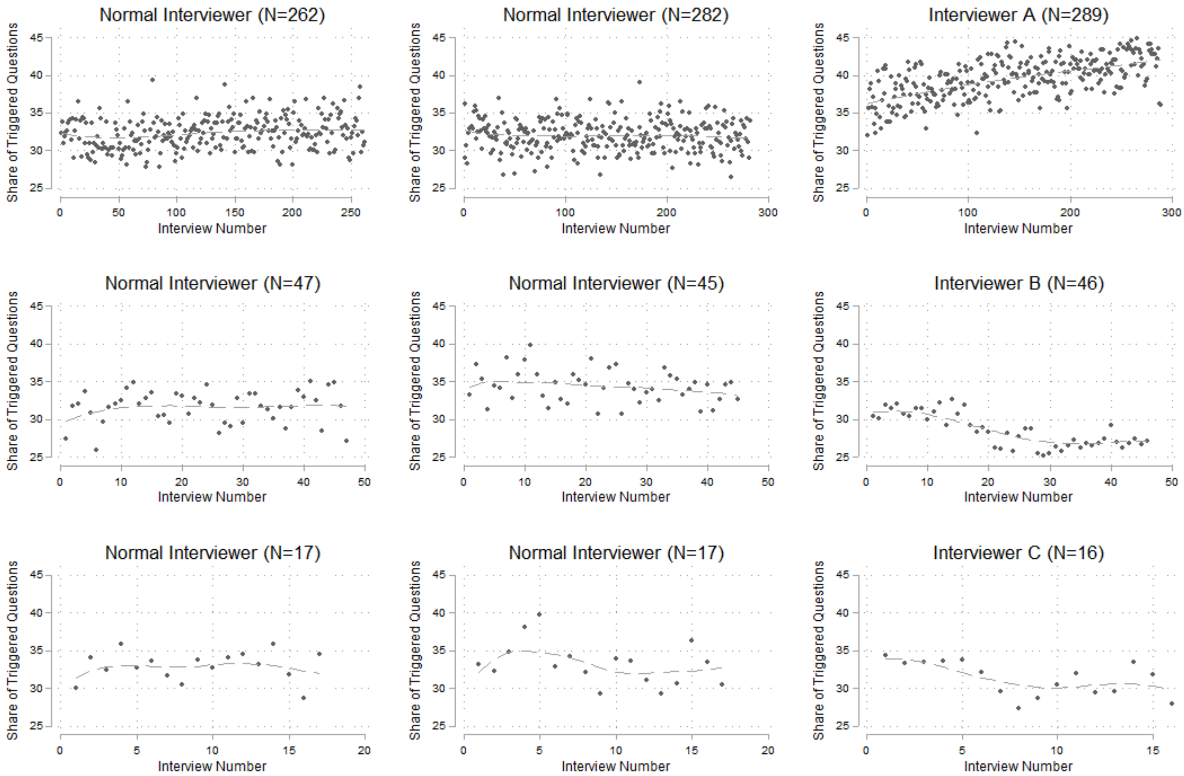
Figure 5. Indicator values for middle responding style over the course of the field period



Notes: Indicator values are calculated as the share of middle responses to scale questions in all answered scale questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

Figure 6 reveals a rather puzzling pattern for the confirmed falsifier. The share of triggered follow-up questions increases over time, which implies that the falsifier increased his/her falsification effort. To the contrary, the share of triggered follow-up questions declines over the field period for the two suspicious interviewers. The indicator values also show a relatively low dispersion for these interviewers. Again, from roughly the 20th interview onward we see a break in the pattern. For the arbitrary interviewers, no such trend is observed. Due to the small number of interviews, the explanatory power for interviewer C is limited. Nonetheless, we can observe a slightly negative slope. Note that for the comparator we do not observe any trends over their field period.

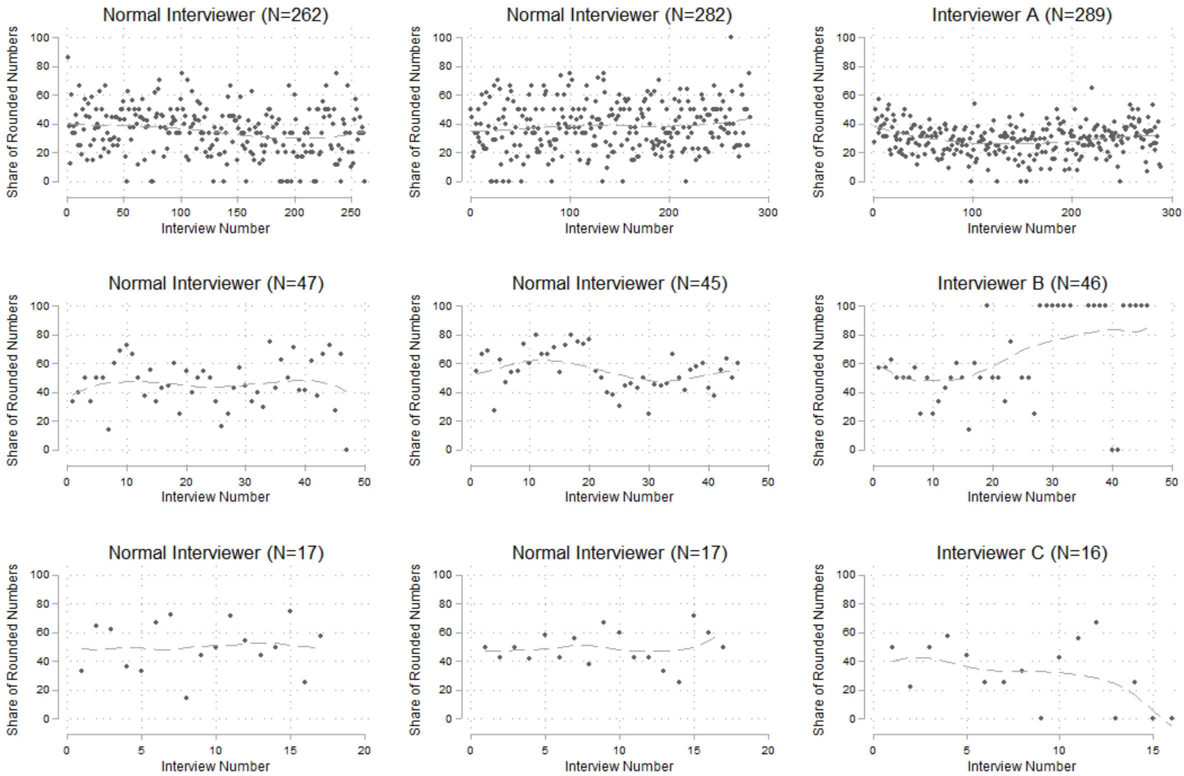
Figure 6. Indicator values for triggered follow-up questions over the course of the field period



Notes: Indicator values are calculated as the share of triggered follow-up questions in all potentially triggered follow-up questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

The indicator for the rounded values, shown in Figure 7, also displays rather suspicious patterns. The confirmed falsifier has a reduced share of rounded numbers in the beginning of the field period, but remains at the same level with reduced variation afterwards. Suspicious interviewer B resembles the non-suspicious interviewers at the start of the field period, but shows extreme values from the 28th interview onward. This hinges on the fact that the suspicious interviewer answered questions on numerical values only once or twice per interview, whereas he did not give valid answers to the other questions on numerical values or did not trigger them. As before, the evidence for interviewer C is restricted due to the small number of interviews.

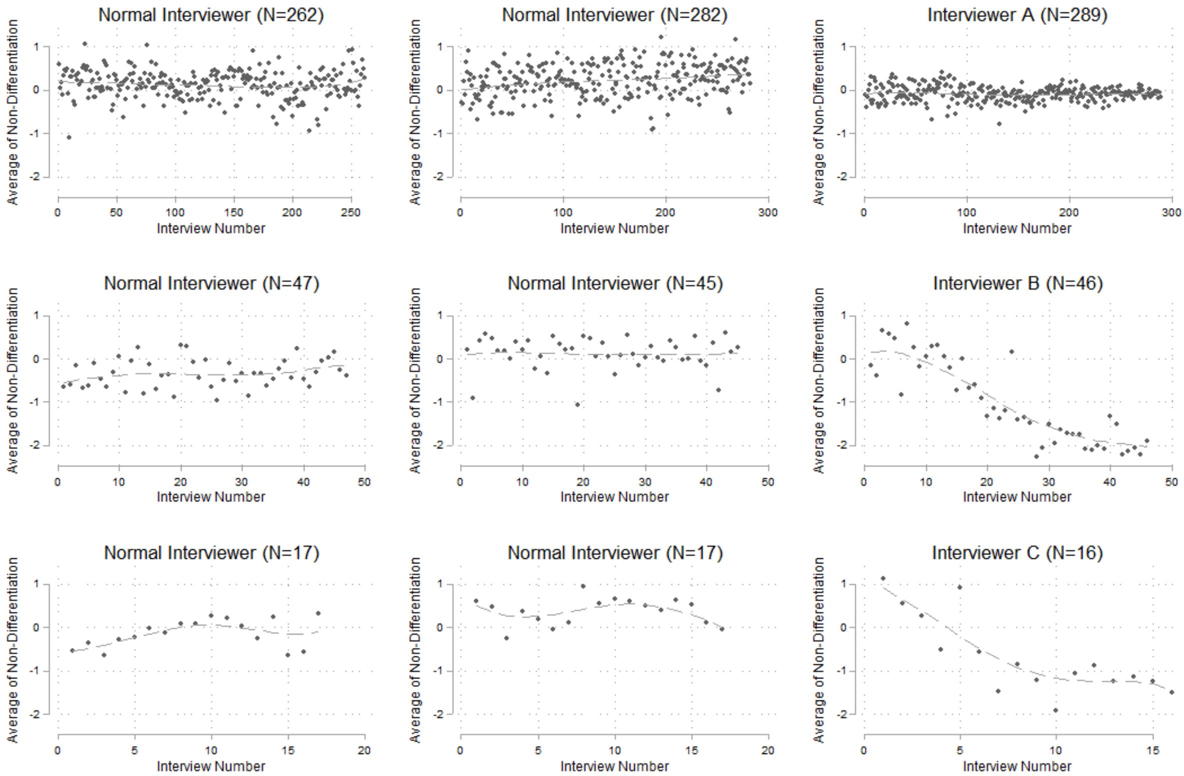
Figure 7. Indicator values for rounding over the course of the field period



Notes: Indicator values are calculated as the share of rounded numbers in the total frequency of responses to open-number questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

Figure 8 shows the time patterns for non-differentiation. The confirmed falsifier's variation differs significantly from the other interviewers; that is, it is much smaller than the comparator interviewers. Moreover, there are clear trends for the two suspicious interviewers. The average standard deviation of responses to the same-scale item batteries decreases for both of them whereas the non-suspicious interviewers do not show such a trend. Towards the end of the field period, the suspected interviewers take on extremely low indicator values.

Figure 8. Indicator values for non-differentiation over the course of the field period



Notes: Indicator values are calculated as the average standardized standard deviation of multiple item batteries.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

Collectively, the results reveal very striking patterns for the confirmed falsifier and both suspicious interviewers. This is also true for other indicators (e.g., duration of interview, item non-response; results not shown). We note that at the beginning of the field period, the suspicious interviewers show similar patterns to the other interviewers, however, the patterns deviate significantly over the course of the field period. This suggests that they began to falsify interviews rather late in their respective field periods.

4.5 Content abnormalities

Table 3 presents selected results from content-based comparisons between the confirmed falsifier, suspicious interviewers, and the unsuspicious interviewers.

According to the data collected from the confirmed falsifier, the majority have participated in vocational or university studies, while no one attended such studies according to suspicious interviewers B and C. In contrast, the respondents interviewed by unsuspicious interviewers reported that roughly one-fourth participated or obtained a degree in vocational or university studies. Almost all respondents of interviewer A reported to have participated in a BAMF Integration course. According to interviewer B, none of his respondents participated in such a course. For the unsuspicious interviewers, the average participation rate is roughly 35%. There is also a strong deviation in the estimated unemployment rate: the confirmed falsifier underestimated the unemployment rate (19.7%) severely compared to the unemployment rate of the unsuspicious interviewers (89.2%). Regarding family composition, 76% of interviewer A's respondents have children, whereas the value is only 4.4% for one of the suspicious interviewers (Interviewer B) – a strong deviation from the 62.9% reported by the unsuspicious interviewers. In sum, we observe significant deviations in content-specific information collected from the confirmed and suspected falsifiers and the unsuspicious interviewers. In sum, we observe significant deviations in content-specific information collected from the confirmed and suspected falsifiers and the unsuspicious interviewers.

Table 3. Content-specific information collected during or after the interview

Variable	Total sample (excluding interviewers A, B, and C)		Interviewer A		Interviewer B		Interviewer C	
	Share	N	Share	N	Share	N	Share	N
Respondent characteristics								
Vocational education or university studies before arrival	24.6	4420	73.0	289	0	43	0	15
Participation in integration measures:								
BAMF Integration course	35.2	4404	99.7	289	0	45	60.0	15
ESF-BAMF language course BA	2.9	4393	91.4	289	0	46	0	15
BA introductory language programme	8.0	4384	51.9	289	0	46	0	14
BA Perspectives for Refugees Other	2.2	4384	57.8	289	0	46	0	15
BA Perspectives for young Refugees	0.5	4386	6.6	289	0	46	0	15
Other German language course	37.3	4419	24.6	289	2,2	46	6.7	15
Not-working	89.2	4465	19.7	289	100.0	46	93.75	16
Children (Yes/No)	62.9	4419	76.1	289	4.4	45	37.5	16
Residence: shared accommodation	35.7	3273	38.7	217	67.7	34	46.2	13
Interviewer questionnaire								
Very good German language skills	10.2	4465	78.9	289	89.1	46	87.5	16
Usage of translate texts (for each question)	57.9	4465	0	289	89.1	46	37.5	16
Usage of Audio-help files (for each question)	6.3	4465	0	289	84.8	46	31.3	16

Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016, own calculations.

In the lower part of Table 3 we explore responses to the post-interview evaluation form that interviewers were instructed to complete on their own. Approximately 10 percent of the total evaluations completed by the unsuspicious interviewers evaluated the language skills of respondents as being (very) good during the interview. Rather strikingly, this share exceeds 78 percent for the confirmed falsifier and both suspicious interviewers. However, this information conflicts with the information on the use of audio files or translated texts given by one of the suspicious interviewers: for about 89% of interviewer B's respondents, frequent usage of translated texts and audio files to aid the interview was reported. For the interviews conducted by the unsuspicious interviewers, we find stronger and more plausible consistency between evaluations of respondents' language skills and usage of additional translation tools.

Overall, we find stark differences between the suspicious interviewers and the remaining crop of interviewers on content-specific responses. The person- and household-level interview data both show unusual response patterns with respect to these sets of interviewers. The answer patterns avoid many follow-up questions about integration measures, employment, and children, which likely saved the suspected falsifiers much time and effort.

4.6 Panel response rates

The suspicious interviewer behavior can be further investigated using survey outcome data from the second wave. Table 4 shows disposition codes for participation in the second wave for the two suspicious interviewers, the confirmed falsifier, and all other interviewers combined. Roughly 56-65% of wave 1 respondents who were assigned to the two suspicious interviewers did not take part in the second wave. This value is in stark contrast to the non-participation rate of the remaining interviewer sample (about 34 percent).

Table 4. Survey Outcome results for the second wave

Variable	Total sample (excluding interviewers A, B, and C)		Interviewer A		Interviewer C		Interviewer B	
	Share	N	Share	N	Share	N	Share	N
Carried out	42.5	2067	17.3	50	37.5	6	32.6	15
Partly carried out	23.4	1139	36.0	104	6.3	1	2.2	1
Refusal	27.3	1330	26.0	75	56.3	9	54.4	25
Other nonresponse	6.8	332	20.8	30	0.0	0	10.9	5
Total	100	4868	100	298	100	16	100	46

Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016 and 2017, own calculations.

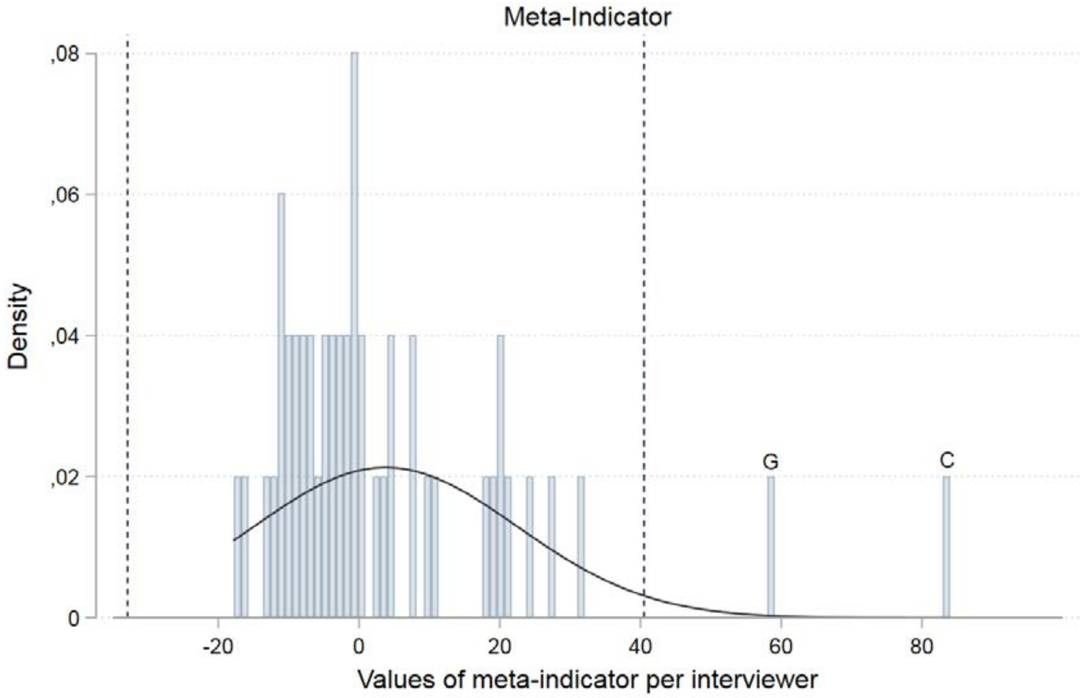
Since we suspected that both suspicious interviewers began to falsify interviews later in the field period, we additionally considered the development of the re-interview rate over time (results not shown). Strikingly enough, we found that none of the interviewees who were interviewed in the first wave at the break points or later (see Figures 4-8) participated in the second wave of the survey.

4.7 Interviewer falsification in the second wave

For the second wave of the IAB-BAMF-SOEP Survey of Refugees in Germany we proceeded to investigate potential falsifications in a similar manner as for the first wave. However, the process was complicated by systematic differences in the questionnaire content for the panel and new respondents, compelling us to conduct separate analyses for the two groups.

As for the first wave, we began with the meta-indicator analysis. Figure 9 presents the distribution of the meta-indicator values[7] for the new and panel respondents combined. The results exhibit right-sided critical values for two interviewers (84.00 for interviewer C and 58.71 for interviewer G). Both meta-indicator values exceed the 97.5th percentile of the distribution and are sufficiently deemed as outliers. In particular, it should be noted that interviewer C was already classified as suspicious in the first wave by the previous analyses. Interviewer G did not conduct any interviews in the first wave.

Figure 9. Interviewer-level meta-indicator with aggregated values of all indicators.



Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, own calculations.

Table 5 summarizes the main results from the Single-Linkage clustering method. Here we find that – exactly as with the meta-indicator – interviewers C and G are marked as the most outlying interviewers. The same result is given by the Ward's Linkage clustering algorithm: as Figure 10 shows, one small cluster (Cluster 2) containing only two interviewers: again, interviewers C and G.

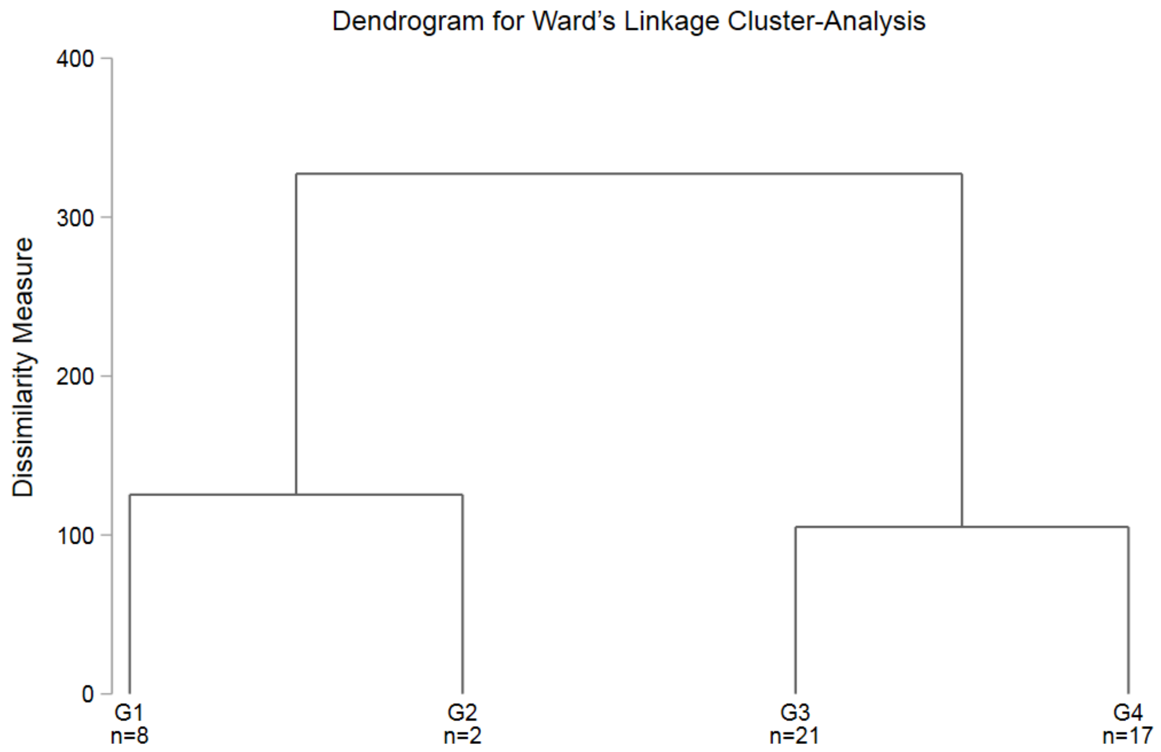
Table 5. List of clusters identified by the Single-Linkage cluster analysis

Clusters	Number of interviewers per cluster	Interviewer ID	Number of conducted person interviews
Cluster 1	46	-	-
Cluster 2	1	C	22
Cluster 3	1	G	83

Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, own calculations.

[7] Similar as for the meta-indicator for wave 1, indicators presented in Table 1 were used. Indicators were calculated separately for new and panel respondents and jointly analyzed. Overall, 34 indicators were used.

Figure 10. Dendrogram for optimal cluster solution (Ward's Linkage)



Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, own calculations.

Finally, we checked the development of the indicators over the field period. Due to different questionnaires for first-time and panel respondents, this analysis is restricted to panel respondents only since the suspicious interviewers C and G mainly conducted panel interviews. The rounding indicator was excluded from this analysis since the number of open-numeric questions was very small. Overall, we find that the suspicious interviewers' development of indicator values over the field period does not substantially differ from arbitrary interviewers with similar numbers of panel interviews (see Figures A1-A4 in Appendix).

Regarding the content of the suspicious interviewers' respondents, we observed substantial deviations from the rest of the sample (see Table A2 in Appendix). For instance, 5.8% of interviewer G's panel respondents and 15.8% of interviewer C's panel respondents stated that they had a foreign educational degree, whereas the share is 57.9% in the rest of the panel sample. There are similar deviations for questions on the respondents' participation in integration measures and responses to the interviewer questionnaire.

5 Conclusion

Interviewer falsification can have a substantial impact on survey data quality and can lead to population estimates that are biased (see Schnell 1991; Schräpler and Wagner 2005). Especially in the context of policy research, unidentified falsification can be problematic since the falsified data might produce results that lead to wrong conclusions and a misallocation of public funds.

In order to prevent or rather identify such situations, the aim of this case study was to evaluate new and existing methods enabling efficient identification of interviewer falsification. For this purpose, we utilized the IAB-BAMF-SOEP Survey of Refugees in Germany in which a high-profile case of fraudulent interviewer behavior was reported. As part of the identification strategies employed, we focused on various indicators of falsification proposed in the literature as well as new ones. These indicators were explored via (1) a meta-indicator analysis, (2) cluster analyses, and (3) their variation over interviewers' respective field periods. Through the combined use of the different analytical tools and a rich set of indicators, a broad control and identification of different types of fraudulent interviewing behavior was achieved.

Through the course of our analysis, the methods were not only successful in retrospectively identifying the confirmed falsifier in the first wave but also discovering further suspicious interviewers in the first and second wave of the survey. These three suspicious interviewers conducted 167 person and 105 household interviews in total. Suspicious interviewer B only conducted interviews in the first wave, suspicious interviewer G only in the second wave and the last suspicious interviewer C was responsible for interviews in both waves. All of these interviewers were jointly identified by all applied methods. The meta-indicator, the cluster analysis as well as the development of indicator values over time showed promising results. Interviewers that were marked as suspicious by the meta-indicator were also marked, or rather grouped, by the cluster analyses. As has been shown by the development of indicators over time, suspicious patterns can be detected early in the field period. This is a useful finding as most published studies on interviewer falsification implement their methods on the full survey data set. Yet, the implementation of fraud detection methods earlier in the field period would be more valuable as intervention steps could be taken to prevent dishonest interviewers from causing more harm.

Follow-up control checks conducted by the survey institute for the first wave of the IAB-BAMF-SOEP Survey of Refugees in Germany supported our empirical results of dishonest interviewer behavior for both suspicious interviewers (B and C) from the first wave but also confirmed that some honest interviews took place.[8] In the course of our second wave analysis, further interviewers showed suspicious patterns for specific parts of their respective field periods or single falsification indicators. However, given that only two interviewers (C and G) showed suspicious patterns in almost all of our analyses, we considered them as particularly problematic. Based on our suggestion, the principal investigators agreed to exclude the data collected by those interviewers from IAB-BAMF-SOEP Survey of Refugees in Germany in the official data release (v34). We may conclude, therefore, that the combination of methods presented in this study (1) may be useful to other studies with interviewer participation, and (2) could be used as part of an efficient and cost-effective quality control strategy to ensure high-quality interviewer-led data collections in surveys. It is important to stress that the majority of interviewers follow the prescribed rules, whereas dishonest interviewer behavior is rather a rare event. Still, methods for identifying suspicious interviewers should be used to ensure the quality of survey data.

[8] For the respective announcement, see https://www.diw.de/sixcms/detail.php?id=diw_01.c.616027.de

References

AAPOR. (2003). Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection, and Repair of Its Effects. *Survey Research Newsletter*, *35*(1), 1–9.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden* (14th ed.). Berlin, Heidelberg: Springer Berlin Heidelberg.

Blasius, J., & Friedrichs, J. (2012). Faked Interviews. In S. Salzborn, E. Davidov, & J. Reinecke (Eds.), *Methods, Theories, and Empirical Applications in the Social Sciences* (pp. 49–56). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bredl, S., Storfinger, N., & Menold, N. (2013). A Literature Review of Methods to Detect Fabricated Survey Data. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers' Deviations in Surveys – Impact, Reasons, Detection and Prevention* (pp. 3–24). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bredl, S., Winker, P., & Kötschau, K. (2012). A statistical approach to detect cheating interviewers. *Survey Methodology, 38(1), 1-10*.

Brücker, H., Croisier, J., Kosyakova, Y., Kröger, H., Pietrantuono, G., Rother, N., & Schupp, J. (2019). Zweite Welle der IAB-BAMF-SOEP-Befragung: Geflüchtete machen Fortschritte bei Sprache und Beschäftigung. *IAB-Kurzbericht*, *3*.

Brücker, H., Rother, N., & Schupp, J. (2017). IAB-BAMF-SOEP-Befragung von Geflüchteten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. Berlin: DIW Berlin, German Institute for Economic Research

Brücker, H., Rother, N., Schupp, J., Gostomski, C. B. Von, Böhm, A., Fendel, T., … Vallizadeh, E. (2016). IAB-BAMF-SOEP-Befragung von Geflüchteten: Flucht, Ankunft in Deutschland und erste Schritte der Integration. *IAB-Kurzbericht*, *24*.

Brüderl, J., Huyer-May, B., & Schmiedeberg, C. (2013). Interviewer behavior and the quality of social network data. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers' Deviations in Surveys – Impact, Reasons, Detection and Prevention* (pp. 147–160). Wiesbaden: VS Verlag für Sozialwissenschaften.

De Haas, S., & Winker, P. (2016). Detecting Fraudulent Interviewers by Improved Clustering Methods – The Case of Falsifications of Answers to Parts of a Questionnaire. *Journal of Official Statistics*, *32*(3), 643–660.

DIW. (2018). IAB-BAMF-SOEP-Befragung Geflüchteter in Deutschland: Revision des Datensatzes. Retrieved from http://www.diw.de/sixcms/detail.php?id=diw_01.c.574021.de

Harrison, D., & Krauss, S. (2002). Interviewer cheating: Implications for Research on Entrepreneurship in Africa. *Journal of Developmental Entrepreneurship*, *7*(3), 319–330.

Hood, C. C., & Bushery, J. M. (1997). Getting More Bang from the Reinterview Buck: Identifying “At Risk” Interviewers. In *Proceedings of the American Statistical Association (Survey Research Methods Section)* (pp. 820–824).

IAB. (2017). *Revidierter Datensatz der IAB-BAMF-SOEP- Befragung von Geflüchteten*. Nuremberg.

Kosyakova, Y., Skopek, J., & Eckman, S. (2015). Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach. *International Journal of Public Opinion Research*, *27*(3), 417–431.

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effects of asking filter questions in interleafed versus grouped format. *Sociological Methods and Research*, *40*(1), 88–104.

Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). *Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4)*. *SOEP Survey Papers 477: Series C*. Berlin.

Krosnick, J., & Alwin, D. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*, 201–219.

Kwasniewski, N., Maxwill, P., Seibt, P., & Siemens, A. (2018). Manipulation in der Marktforschung. Wie Umfragen gefälscht und Kunden betrogen werden. Retrieved March 9, 2018, from http://www.spiegel.de/wirtschaft/unternehmen/manipulation-in-der-marktforschung-wie-umfragen-gefaelscht-werden-a-1190711.html

Menold, N., Winker, P., Storfinger, N., & Kemper, C. J. (2013). A method for ex-post identification of falsifications in survey data. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers’ Deviations in Surveys – Impact, Reasons, Detection and Prevention* (pp. 25–47). Wiesbaden: VS Verlag für Sozialwissenschaften.

Messick, S. (1966). The Psychlogy of Acquiescence: An Interpretation of Research Evidence. *ETS Research Bulletin Series*, *1966*(1), 1–44.

Murphy, J., Baxter, R., Eyerman, J., Cunningham, D., & Kennet, J. (2004). A System for Detecting Interviewer Falsification. *American Association for Public Opinion Research 59th Annual Conference*, 4968–4975.

Murphy, J., Biemer, P., Stringer, C., Thissen, R., Day, O., & Hsieh, Y. P. (2016). Interviewer falsification: Current and best practices for prevention, detection, and mitigation. *Statistical Journal of the IAOS*, *32*(3), 313–326.

Porras, J., & English, N. (2004). Data-Driven Approaches to Identifying Interviewer Data Falsification: The Case of Health Surveys. In *Proceedings of the Survey Research Methods Section, American Statistical Association* (pp. 4223–4228). Retrieved from http://www.amstat.org/sections/srms/Proceedings/y2004/files/Jsm2004-000879.pdf

Reuband, K.-H. (1990). Interviews, die keine sind. “Erfolge” und “Mißerfolge” beim Fälschen von Interviews. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *42*(4), 706–733.

Schäfer, C., Schräpler, J.-P., Müller, K.-R., & Wagner, G. G. (2005). Automatic Identification of faked and fraudulent interviews in survey by two different methods. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift Für Wirtschafts- Und Sozialwissenschaften*, *125*(1), 183–193.

Schnell, R. (1991). Der Einfluß gefälschter Interviews auf Survey- Ergebnisse. *Zeitschrift Für Soziologie*, *20*(1), 25–35.

Schräpler, J.-P., & Wagner, G. G. (2005). Characteristics and impact of faked interviews in surveys - An analysis of genuine fakes in the raw data of SOEP. *Allgemeines Statistisches Archiv*, *89*(1), 7–20.

Seibt, P. (2018a). Sozio-Ökonomisches Panel. Wie eine Interviewerin Befragungen von Flüchtlingen fälschte. Retrieved March 9, 2018, from http://www.spiegel.de/politik/deutschland/fluechtlinge-wie-eine-interviewerin-befragungen-faelschte-a-1191086.html

Seibt, P. (2018b). Gefälschte Umfragen. Die Macht der Marktforscher. Retrieved March 9, 2018, from http://www.spiegel.de/wirtschaft/unternehmen/gefaelschte-umfragen-so-maechtig-sind-marktforscher-a-1190705.html

Stokes, L. S., & Jones, P. (1989). Evaluation of the Interviewer Quality Control Procedure for the Post-Enumeration Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 696–698).

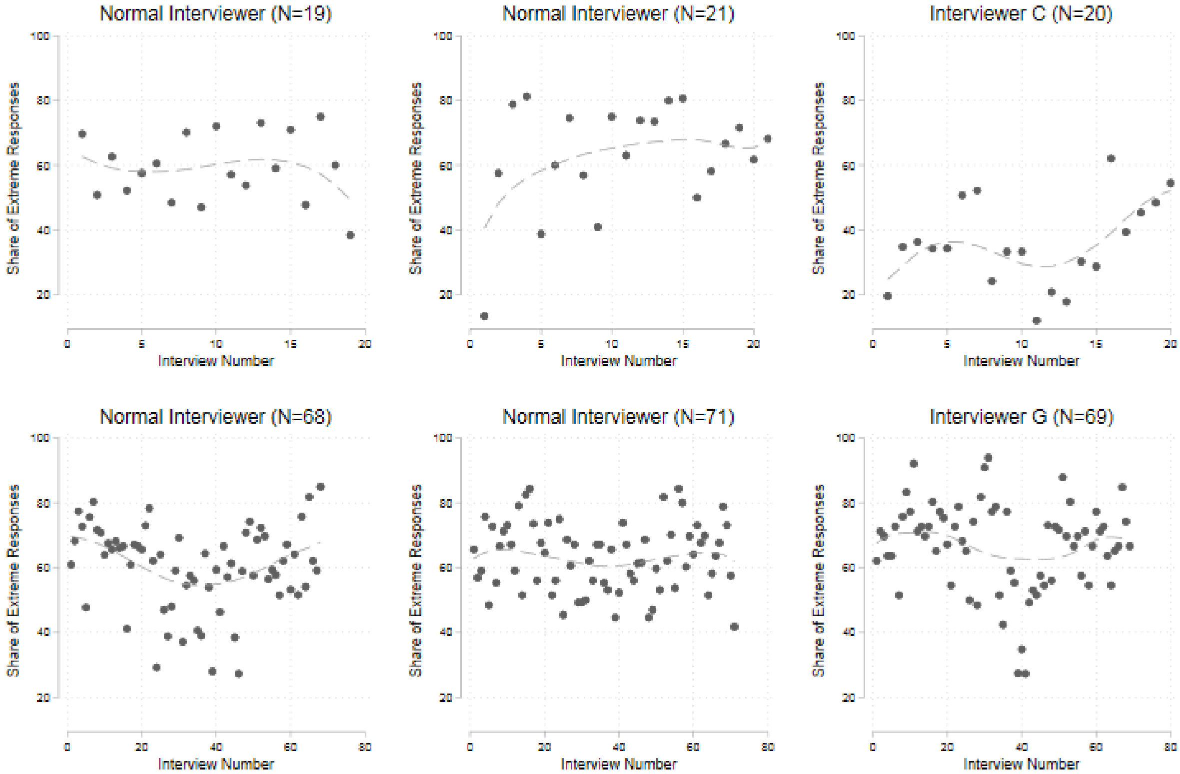
Storfinger, N., & Winker, P. (2013). Assessing the performance of clustering methods in falsification using bootstrap. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers’ Deviations in Surveys – Impact, Reasons, Detection and Prevention* (pp. 49–65). Wiesbaden: VS Verlag für Sozialwissenschaften.

Swanson, D., Cho, M. J., & Eltinge, J. (2003). Detecting Possibly Fraudulent or Error-Prone Survey Data Using Benford ’ s Law. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 4172–4177.

Tourangeau, R., Rasinski, K., Jobe, B. J., Smith, W. T., & Pratt, F. W. (1997). Sources of Error in a Survey on Sexual Behavior. *Journal of Official Statistics*, *13*(4), 341–366.

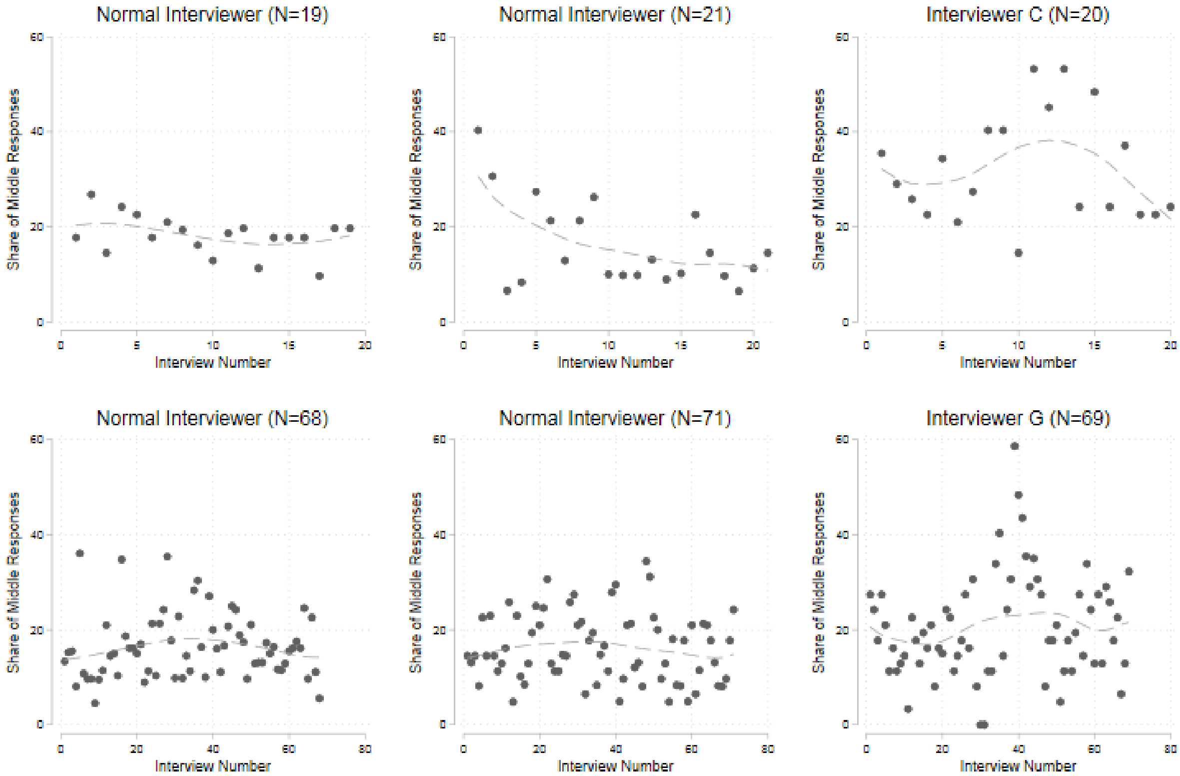
Appendix

Figure A1. Indicator values for extreme responding style



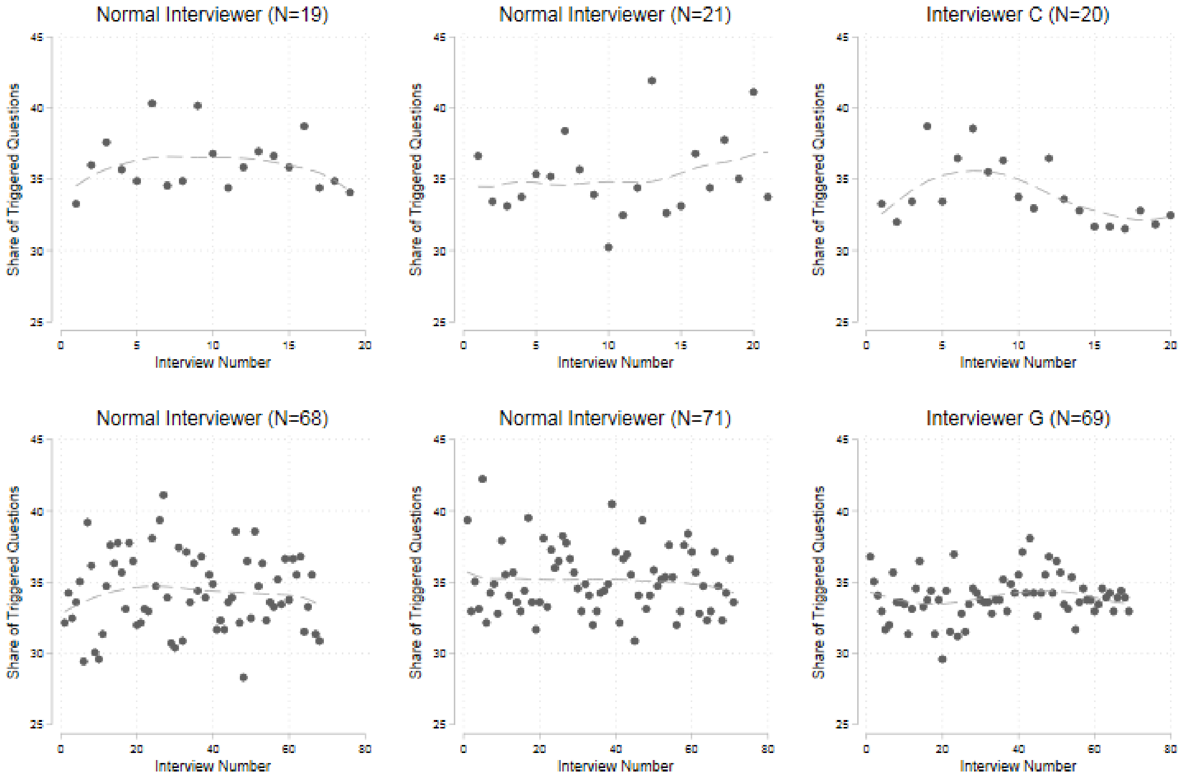
Notes: Indicator values are calculated as the share of extreme responses to scale questions in all answered scale questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, panel respondents, own calculations.

Figure A2. Indicator values for middle responding style



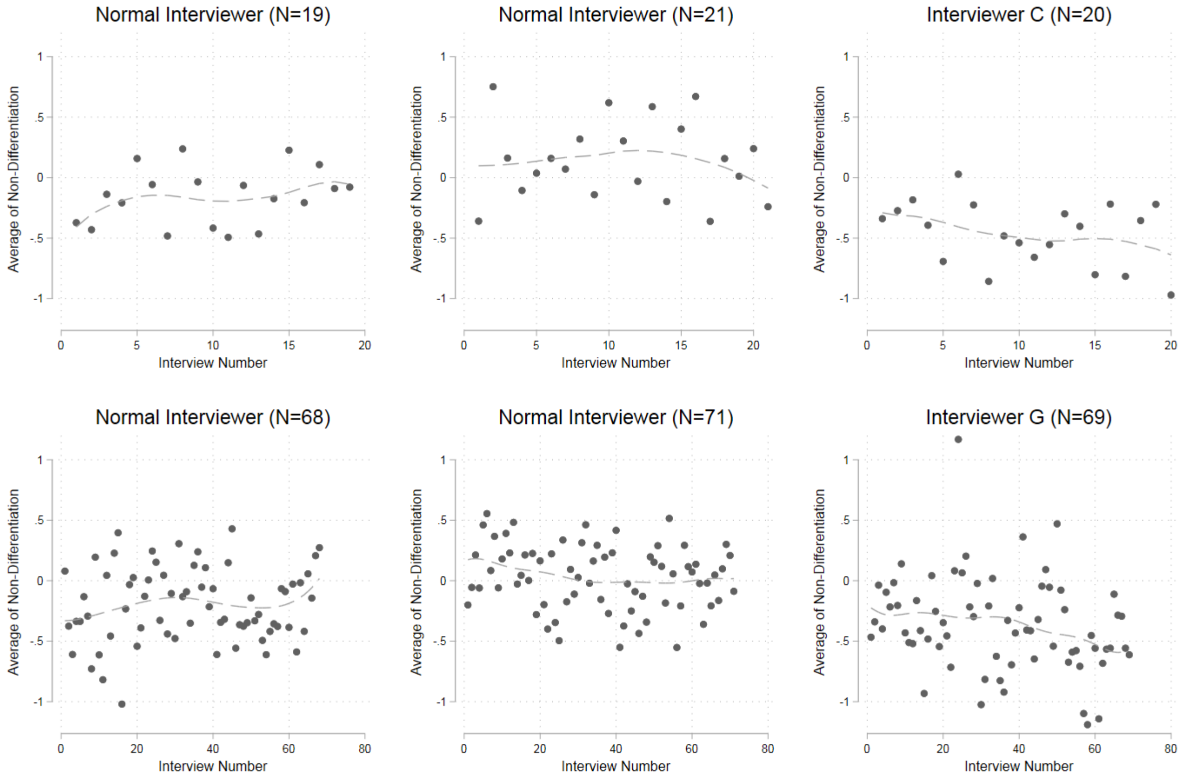
Notes: Indicator values are calculated as the share of middle responses to scale questions in all answered scale questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, panel respondents, own calculations.

Figure A3. Indicator values for triggered follow-up questions



Notes: Indicator values are calculated as the share of triggered follow-up questions in all potentially triggered follow-up questions.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, panel respondents, own calculations.

Figure A4. Indicator values for non-differentiation



Notes: Indicator values are calculated as the average standardized standard deviation of multiple item batteries.
Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, panel respondents, own calculations.

Table A1. Identified Interviewers

Identification	Wave 1		Wave 2	
	Person interviews	Household interviews	Person interviews	Household interviews
A	289	218	-	-
B	46	34	-	-
C	16	13	22	13
D	1	1	-	-
E	2	2	-	-
F	1	1	-	-
G	-	-	83	45

Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2016 and 2017, own calculations.

Table A2. Content-specific information collected during or after the interview

Variable	Total sample (excluding interviewers C and G)		Interviewer C		Interviewer G	
	Share	N	Share	N	Share	N
Respondent characteristics						
Foreign Educational Degree (Yes/No)	57.9	2632	15.8	19	5.8	69
Participation in integration measures:						
BAMF Integration course	49.1	5561	81.8	22	54.2	83
ESF-BAMF language course	6.5	5469	0	22	6.0	83
Other German language course	34.2	5543	0	22	7.3	82
Not-working	84.1	5616	77.3	22	90.4	83
Interviewer questionnaire						
Very good German language skills	18.3	5592	90.9	22	38.6	83
Usage of translate texts (for each question)	51.1	5592	0	22	14.5	83
Usage of Audio-help files (for each question)	4.5	5592	0	22	4.8	83

Source: IAB-BAMF-SOEP Survey of Refugees in Germany, 2017, own calculations.

Imprint

FDZ-Methodenreport 02|2019 EN

Date of publication

04 April 2019

Publisher

Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nuremberg

Download

http://doku.iab.de/fdz/reporte/2019/MR_02-19_EN.pdf

Documentation version

IAB-BAMF-SOEP-SUF1617_EN_v1_dok1, DOI: 10.5164/IAB.FDZM.1902.en.v1

Dataset version

IAB-BAMF-SOEP-SUF 1617 v1, DOI: 10.5684/soep.iab-bamf-soep-mig.2017

All publications in the series "FDZ-Methodenreport" can be downloaded from

https://fdz.iab.de/en/FDZ_Publications/FDZ_Publication_Series/FDZ-Methodenreporte.aspx

Website

http://fdz.iab.de

Corresponding author

Yuliya Kosyakova
Phone: +49 911 179-3643
Email: yuliya.kosyakova@iab.de