

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

2019 Workshop: Interviewers and Their Effects  
from a Total Survey Error Perspective

Sociology, Department of

---

2-26-2019

# Interviewer Falsification in Survey Research: Detection Methods and Impact of Fraudulent Interviews

Silvia Schwanhäuser

*Institute for Employment Research, Germany, Silvia.Schwanhaeuser2@iab.de*

Joseph Sakshaug

*University of Mannheim and Institute for Employment Research, Germany, Joe.Sakshaug@iab.de*

Yuliya Kosyakova

*Kosyakova, Institute for Employment Research, Germany*

Frauke Kreuter

*University of Maryland, University of Mannheim, and Institute for Employment Research, Germany, fkreuter@umd.edu*

Follow this and additional works at: <http://digitalcommons.unl.edu/sociw>



Part of the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

---

Schwanhäuser, Silvia; Sakshaug, Joseph; Kosyakova, Yuliya; and Kreuter, Frauke, "Interviewer Falsification in Survey Research: Detection Methods and Impact of Fraudulent Interviews" (2019). *2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective*. 14.

<http://digitalcommons.unl.edu/sociw/14>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in 2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



INSTITUTE FOR EMPLOYMENT  
RESEARCH  
The Research Institute of the Federal Employment Agency

# INTERVIEWER FALSIFICATION IN SURVEY RESEARCH:

## Detection Methods and Impact of Fraudulent Interviews

Silvia Schwanhäuser  
Joseph W. Sakshaug  
Yuliya Kosyakova  
Frauke Kreuter



# OVERVIEW

---

- What is Interviewer Falsification?
- Case Study - IAB-BAMF-SOEP Survey of Refugees in Germany
- Strategies for Identifying Falsifications
- Results
- Conclusion

# INTERVIEWER FALSIFICATION

---

*“ ‘Interviewer falsification’ means the intentional departure from the designed interviewer guidelines or instructions, unreported by the interviewer, which could result in the contamination of data.” (AAPOR 2003: 1)*

⇒ Includes...

- ... miscoding of answers
- ... misclassification of eligible respondents
- ... deviations from intended mode or selection rules
- ... **fabrication of complete interviews** or parts of it

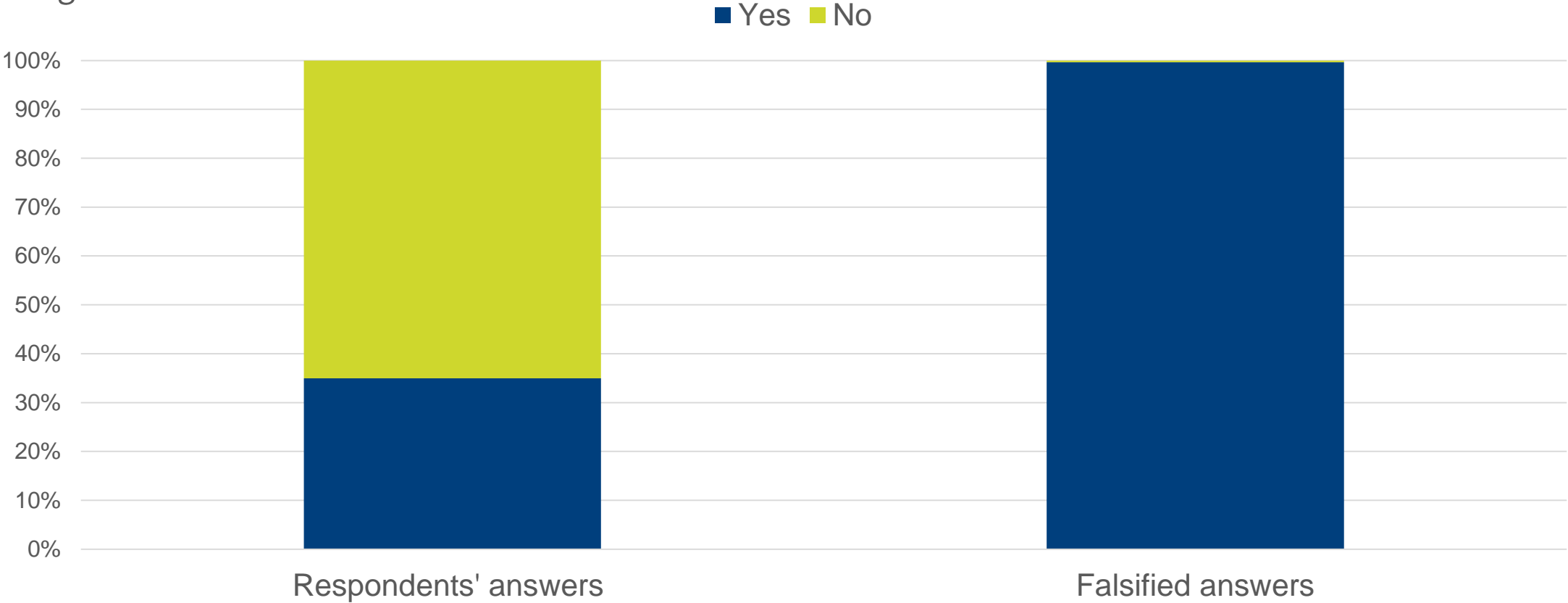
# IAB-BAMF-SOEP SURVEY OF REFUGEES IN GERMANY

---

- Longitudinal household survey
- **Mode:** computer-assisted personal interviewing (CAPI)
- **Field period:** June to December 2016
- **First wave:** 4,816 respondents in 3,554 households
  
- One interviewer identified as falsifier by the survey institute (called “Interviewer A”)
- **Falsified:** 289 person (6.0 percent) and 217 household interviews (6.1 percent)

# IMPACT OF INTERVIEWER FALSIFICATION

Have you attended an integration course organized by the German Federal Ministry for Migration and Refugees?



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

Note: Univariate, unweighted result.

COULD THIS FALSIFIER HAVE BEEN IDENTIFIED  
USING STATISTICAL IDENTIFICATION METHODS?

---

# STRATEGIES FOR IDENTIFYING FALSIFICATIONS

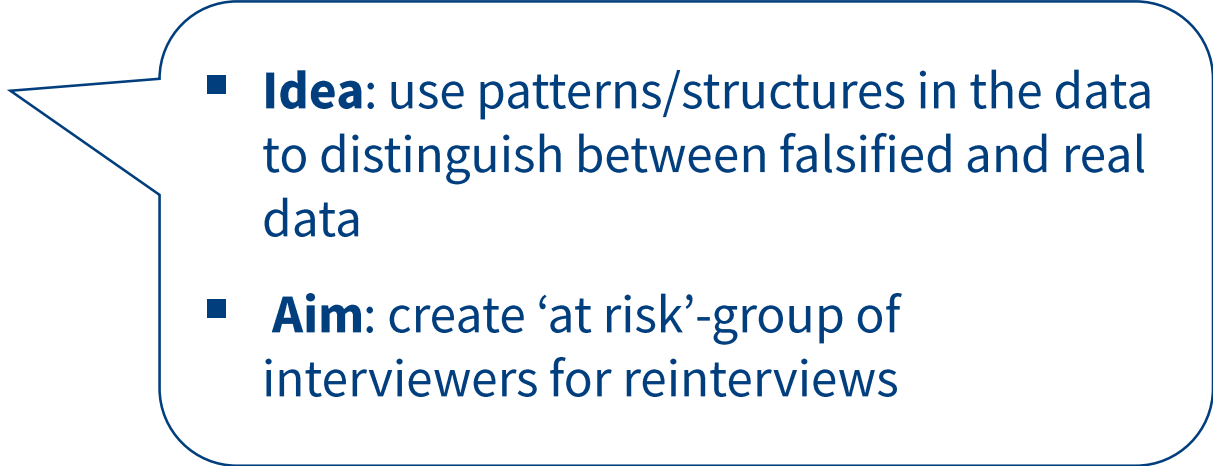
---

## Non-statistical identification strategies

- Monitoring
- Reinterviewing

## Statistical identification strategies

- Duplicate analysis
- PCA / categorical PCA methods
- Analysis of falsification indicators
  - Cluster analysis
  - ...

- 
- **Idea:** use patterns/structures in the data to distinguish between falsified and real data
  - **Aim:** create 'at risk'-group of interviewers for reinterviews



# FALSIFICATION INDICATORS

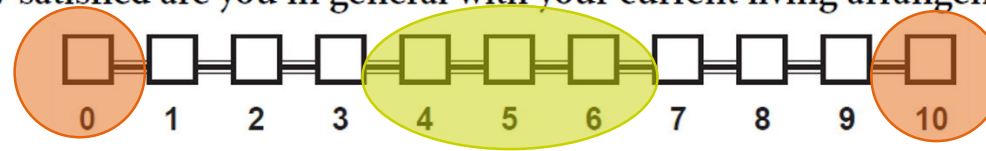
---

- Indicators...
  - ... measure systematic differences between real and falsified data
  - ... can be derived from rational behavior of falsifiers
  - ... can be derived from identified cases of falsification
- Indicators can be calculated from...
  - ... surveyed data (longitudinal / cross-sectional)
  - ... paradata
  - ... interviewer observations

## EXAMPLES FOR INDICATORS

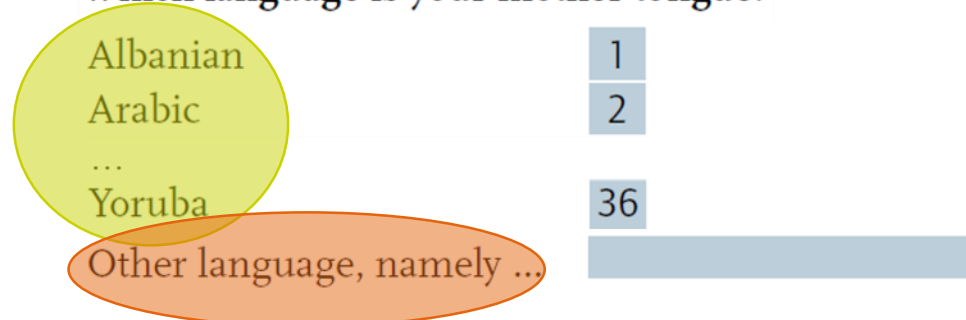
- **Extreme responses:** Lower share of extreme responses on rating scales for falsifiers
- **Middle responses:** Higher share of middle responses on rating scales for falsifiers

How satisfied are you in general with your current living arrangements?



- **Semi-Open responses:** Lower share of responses to “other” in semi-open-ended question for falsifiers

Which language is your mother tongue?



# ANALYZING INDICATORS

---

- Previous approaches:
  - examination of single indicators
  - examination of several indicators but separate analysis
- Our approach: combination of multiple indicators via statistical methods
  - Cluster analysis
  - Meta-indicator

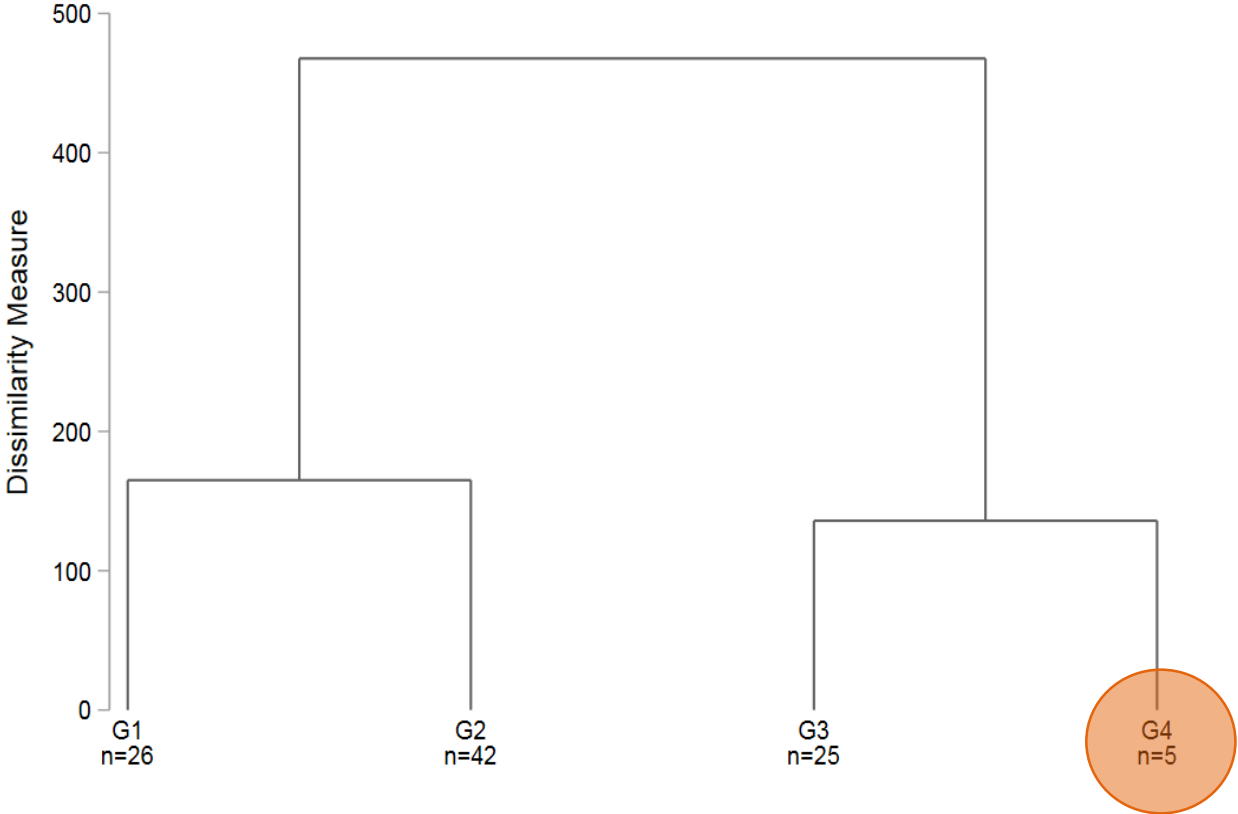
# CLUSTER ANALYSIS

---

- **Idea:** Divide large group (here: group of interviewer) into smaller homogeneous subgroups
- **Aim:** Dividing interviewers on basis of indicators into suspicious and unsuspecting subgroups
- **Clustering Algorithms:**
  - Ward's Linkage: Fuses interviewers that increase cluster variance as little as possible
  - Single Linkage: Fuses similar interviewers first

# RESULTS - CLUSTER ANALYSIS

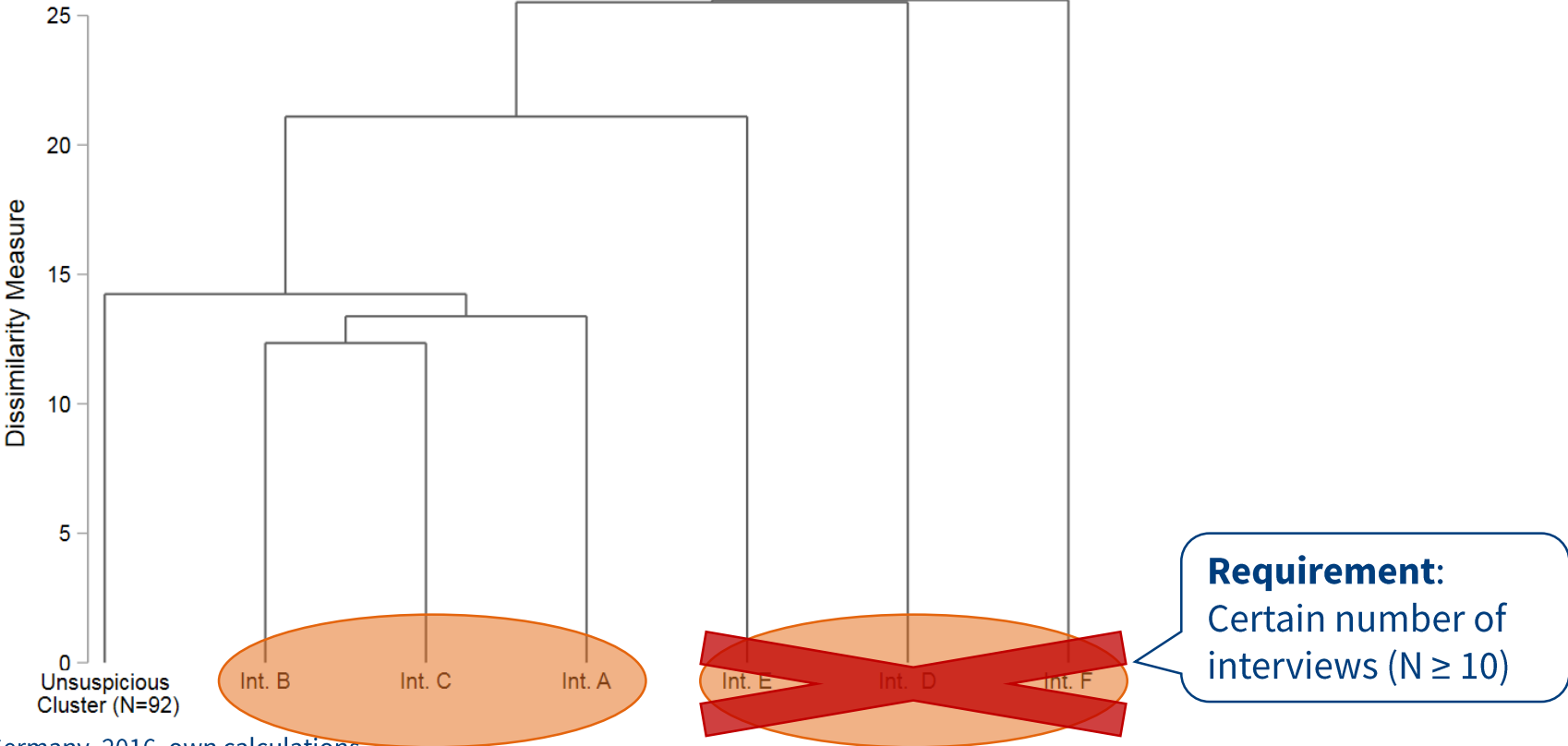
## Dendrogram for Ward's Linkage Cluster-Analysis



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

# RESULTS - CLUSTER ANALYSIS

## Dendrogram for Single-Linkage Cluster-Analysis



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

# META-INDICATOR

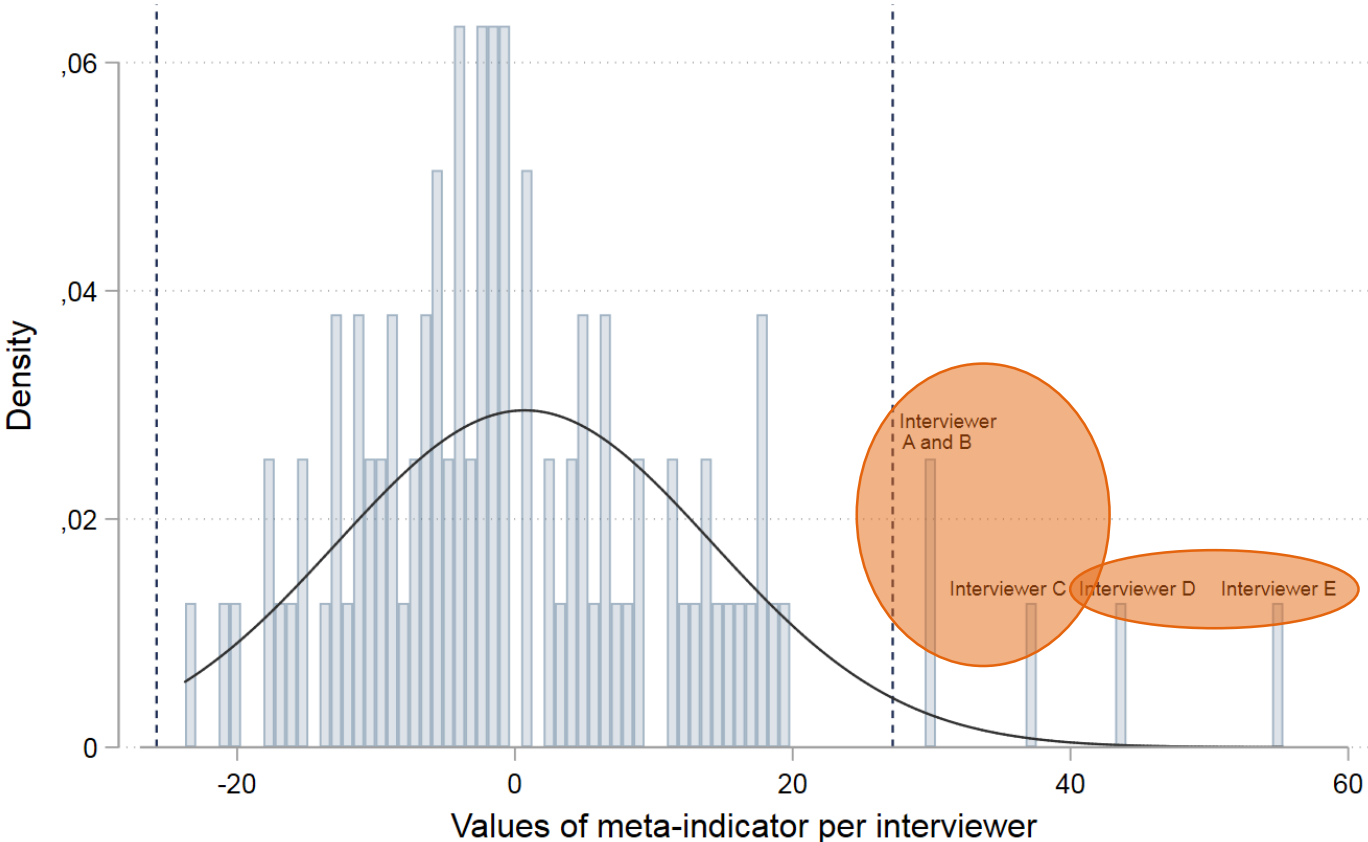
---

**Idea:** Summarizing all indicators in order to interpret them jointly

1. Standardize values
2. Summing up all indicators
3. Checking distribution for outliers

# RESULTS – META-INDICATOR

Distribution of summarized indicator values (meta-indicator)



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



# CONCLUSION

---

## ■ **Summary:**

- Two further suspicious interviewers identified (confirmation in progress)
- First case could have been identified earlier using statistical methods
- Results for meta-indicator and cluster analyses very similar
- Meta-indicator much easier to interpret

## ■ **Outlook:**

- Use similar statistical approaches for the second wave
- Test strategies for identification of further falsification forms
- Application of machine learning algorithms
- Develop strategies that identify cases early in the field period

# CONTACT

---

Silvia Schwanhäuser

 [Silvia.Schwanaeuser2@iab.de](mailto:Silvia.Schwanaeuser2@iab.de)

 +49 911 179 2770

# LITERATURE

---

- AAPOR – American Association for Public Opinion Research (2003): Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects. American Association for Public Opinion Research (AAPOR). Online available [https://www.aapor.org/AAPOR\\_Main/media/MainSiteFiles/falsification.pdf](https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/falsification.pdf).
- Blasius, Jörg; Thiessen, Victor (2012): Assessing the quality of survey data. London: Sage.
- Bredl, S.; Winker, P.; Kötschau, K. (2012): A statistical approach to detect interviewer falsification of survey data. *Survey Methodology - Statistics Canada* 38 (1), 1–10. 85.
- Bredl, Sebastian; Storfinger, Nina; Menold, Natalja (2013): A Literature Review of Methods to Detect Fabricated Survey Data. In: Peter Winker, Natalja Menold und Rolf Porst (Ed.): *Interviewers' Deviations in Surveys*: Peter Lang, 3–24.
- Brücker, H.; Rother, N.; Schupp, J.; Babka von Gostomski, C.; Böhm, A.; Fendel, T.; Friedrich, M.; Giesselmann, M.; Holst, E.; Kosyakova, Y.; Kroh, M.; Liebau, E.; Richter, D.; Romiti, A.; Schacht, D.; Scheible, J.A.; Schmelzer, P.; Siegert, M.; Sirries, S.; Trübswetter, P.; Vallizadeh, E. (2016): IAB-BAMF-SOEP-Befragung von Geflüchteten: Flucht, Ankunft in Deutschland und erste Schritte der Integration. IAB-Kurzbericht, 2016(24).
- 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.
- Guterbock, Thomas M. (2008): Falsification. In: Paul Lavrakas (Ed.): *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc, 267-210.
- Haas, Samuel de; Winker, Peter (2016): Detecting Fraudulent Interviewers by Improved Clustering Methods – The Case of Falsifications of Answers to Parts of a Questionnaire. In: *Journal of Official Statistics* 32 (3), 1. DOI: 10.1515/JOS-2016-0033.
- Hood, Catherine C.; Bushery, John M. (1997): Getting More Bang from the Reinterview Buck. Identifying "At Risk" Interviewers. In: *Proceedings of the Survey Research Method Section, American Statistical Association*, 820–824.

# LITERATURE

---

- IAB – Institute for Employment Research. (2017): Revidierter Datensatz Der IAB-BAMF-SOEP- Befragung von Geflüchteten. Nuremberg.
- Koczela, Steve; Furlong, Cathy; McCarthy, Jaki; Mushtaq, Ali (2015): Curbstoning and beyond. Confronting data fabrication in survey research. In: *SJI* 31 (3), 413–422. DOI: 10.3233/SJI-150917.
- Menold, N.; Winker, P.; Storfinger, N.; Kemper, C.J. (2013): A Method for Ex-Post Identification of Falsification in Survey Data. P. Winker, N. Menold und R. Porst (Ed.): Interviewers' Deviations in Surveys: Peter Lang, 25–47.
- Porras, Javier; English, Ned (2004): Data-Driven Approaches to Identifying Interviewer Data Falsification. The Case of Health Surveys. In: Proceedings of the Survey Research Method Section, American Statistical Association, 4223–4228.
- Schäfer, Christin; Schräpler, Jörg-Peter; Müller, Klaus-Robert; Wagner, Gert G. (2005): Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methodes. Discussion Paper. Deutsches Institut für Wirtschaftsforschung (DIW).
- Schräpler, Jörg-Peter; Wagner, Gert G. (2003): Identification, Characteristics and Impact of Faked Interviews in Surveys. An analysis by means of genuine fakes in the raw data of SOEP. Discussion Paper. Forschungsinstitut zur Zukunft der Arbeit (IZA). Bonn (969).
- Stokes, S. Lynne; Jones, Patty (1989): Evaluation of the Interviewer Quality Control Procedure for the Post-Enumeration Survey. In: Proceedings of the Survey Research Method Section, American Statistical Association, 696–698.
- Storfinger, Nina; Winker, Peter (2013): Assessing the Performance of Clustering Methods in Falsification using Bootstrap. In: Peter Winker, Natalja Menold und Rolf Porst (Ed.): Interviewers' Deviations in Surveys: Peter Lang, 46–65.

# APPENDIX

---

Have you attended an integration course organized by the German Federal Ministry for Migration and Refugees?

Respondents' answers		
	Frequency	Percent
Yes	1560	34.95
No	2904	65.05
Total	4464	100.00

Falsified answers		
	Frequency	Percent
Yes	288	0.35
No	1	99.65
Total	289	100.00

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

# APPENDIX

Indicator	Assumed direction of falsifiers
Acquiescent Responding Style	Lower share of positive connotation independent of question's content for falsifiers
Benford's Law	Poor fit to of Benford's distribution to first digits for falsifiers
E-Mail	Lower share of provided e-mail addresses for falsifiers
Extreme responses	Lower share of extreme responses on rating scales for falsifiers
Filter questions	Lower share of responses which lead to follow-up questions for falsifiers
Interview duration	Shorter duration of completed interviews for falsifiers
Interviewer evaluation	Very positive evaluation of the interview situation for falsifiers
Item nonresponse	Lower item nonresponse rate for falsifiers
Middle category responses	Higher share of middle responses on rating scales for falsifiers
Non-Differentiation	Lower standard deviation across item scales for falsifiers
Primacy effects	Higher share of choosing the first two categories in non-ordered answer option list for falsifiers
Recency effects	Lower share of choosing the last two categories in non-ordered answer option list for falsifiers
Record linkage consent	Higher share of consent to record linkage for falsifiers
Relative interview duration	Shorter duration of completed interviews relative to the triggered questions for falsifiers
Rounding	Lower share of rounded numbers in numerical open-ended questions for falsifiers
Semi-Open responses	Lower share of responses to "other" in semi-open-ended question for falsifiers
Stereotyping	Higher strength of stereotypical response to attitudinal items for falsifiers
Telephone number	Lower number of provided telephone numbers for falsifiers
Variance	Lower standard deviation for one variable between different interviews of an interviewer for falsifiers

# APPENDIX

---

## Results of meta-indicator and number of interviews for suspicious interviewers

Interviewer	Meta-indicator value	Number of person interviews	Number of household interviews
A	<b>29.53</b>	<b>289</b>	<b>218</b>
B	<b>30.08</b>	<b>46</b>	<b>34</b>
C	<b>37.08</b>	<b>16</b>	<b>13</b>
D	<b>43.82</b>	1	1
E	<b>55.33</b>	2	2
F	13.33	1	1

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

# APPENDIX

---

## Response-Rate of suspicious interviewers for wave 2

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

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