Identification of Falsifications in Surveys – a Link to the Cross-Cultural Context

Natalja Menold
GESIS

GESIS Leibniz Institute for the Social Sciences

Member of the Leibniz Association

Overview

- Falsifications of survey data
- Detection Methods
  - Re-contact
  - Para-data
  - Interviewer Characteristics
  - Statistical Analyses
- Conclusions
- Cross-cultural perspective

Interview Falsifications: A German Case

"*Market researchers are supposed to describe how do Germans think. However, according to information of SPIEGEL, …. manipulated data are not an exception...* "

Spiegel-Online: 01.02.2018

Definition

- "Interviewer falsification means the intentional departure from the designed interviewer guidelines or instructions, which could result in the contamination of data." (AAPOR 2003: 1)

- Kinds of falsifications

  ‣ **Fabrications of interviews (our focus)**

  ‣ Falsifying the process data

  ‣ Miscoding the answers to a question in order to avoid follow up questions

  ‣ Interviewing a non-sampled person

Falsified Data:
Frequency of Occurence and Impact

- ## Frequency of occurrence
  - ▶ *"quite low"* (AAPOR, 2003; Crespi 1945, Evans 1961, Guest 1947)
  - ▶ 3-7% in U.S. Bureau of the Census (Biemer & Stockes 1989)
  - ▶ 100% in a non-OECD country; 50 fake interviews; detected by re-interviewing (Bredl, Winker & Kötschau 2008)

- ## Impact
  - ▶ Falsifications may seriously contaminate the results of correlative and multivariate analyses (Schräpler & Wagner 2003)

Detection Methods
(Bredl, Storfinger & Menold, 2013)

Re-Contact

Usage of Para-data

(Ex-post) Statistical analyses

Re-Contact

- Respondents are re-contacted after the initial interview

- Aim: to verify whether the initial interview actually took place

- By postcards, telephone, mail, in person

- Questions on the time, date, topics of the interview, interviewer's behavior

Re-Contact

Study	Country	Surveys	Method	Success
Case (1971)	US	Numerous; Market research	Telephone Random 20% of sample	90% were 27% of the interviews in studies were not properly
Hood & Bushery	US	NHIS / Bureau	Telephone Random	<0.5%
Hood & Bushery	US	NHIS	Telephone Focused	3.6%
Turner et (2002)	US	Survey on sexually transmitted diseases	Telephone Focused: 100% suspicious interviewers; 40% randomly selected	suspicious interviewers: were
Koch (1995)	DE	ALLBUS	Postcards 25% random selection	60% response <0.5%
Koch (1995)	DE	ALLBUS	focused	2.3%

Disadvantages of Re-Contact

- Problems to obtain participation

- Memory problems

- High costs (if large subsamples are re-contacted)

- Random selection of re-contacts: low effectivity

  ➡️ Focusing on "interviewers at risk" seems to increase hit rates

Usage of Para-data

Interviewer-characte-ristics

Statistical analyses

"At risk" interviewers

Para-data

- Implausible success rates

  (Turner et al., 2002, SHARE)

- Lacking contact information

  (Hood & Bushery, 1997)

- Implausible ineligible rates

   (Hood & Bushery, 1997)

- Date and time stamps at computer assistance

 (Bushery et al., 1999; Krejsa et al.,1999; Murphy et al., 2004)

Interviewer Characteristics

- Experience
  - ▶ *"for the newer interviewers it may be useful to re-interview some of their work more frequently"*
    (Schreiner et al., 1988: 496)
  - ▶ falsifications by experienced interviewers are more difficult to detect
    (Hood and Bushery; 1997; Schreiner et al., 1988)

- Young interviewers with a higher level of education produce a higher rate of falsifications (Koch, 1994)

- No effects of gender, age, education
  (Schraepler and Wagner, 2003)

Ex-post Statistical Analyses of Survey Data

- Benford's Law (Benford, 1938)

  ▸ Accurate survey data: first digit follows a logarithmic and scale invariant distribution

- Survey data distributions may deviate from Benford's Law because of rounding (Schräpler & Wagner, 2003; Wang and Pedlow, 2005)
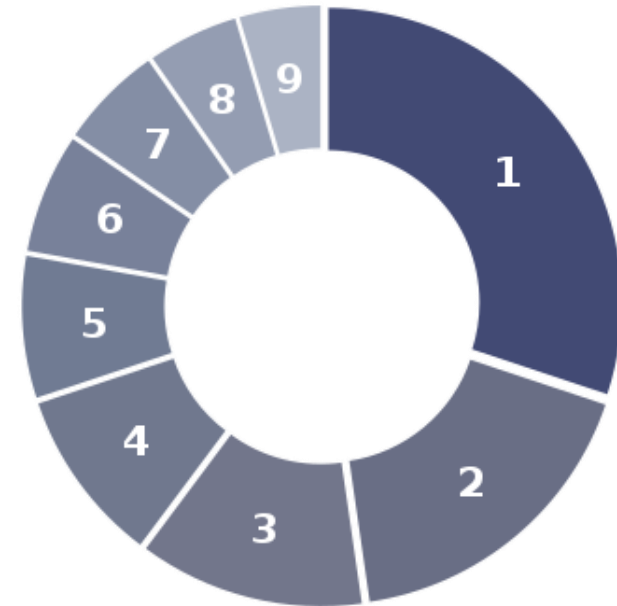
Statistical Analyses (Menold et al., 2013)

- Response behavior (formal) indicators
  - response to filters (FILTER) to avoid further questions
  - usage of semi-open questions (SEMI) less frequently
  - higher survey non-differentiation by falsifiers (SND)
  - Falsifiers claim familiarity with nonexistent items (VOCT)
  - lower recency and higher primacy effects
  - lower item nonresponse (INR); less acquiescent (ARS), extreme (ERS) and middle (MRS) responding
  - differences in rounding behavior
- Experimental study to evaluate indicators: dataset of 700 real and 700 falsified interviews
- Multivariate cluster analysis (global clustering with heuristic optimization)
 - 82% of falsifiers and 92% of non-falsifiers was correctly identified
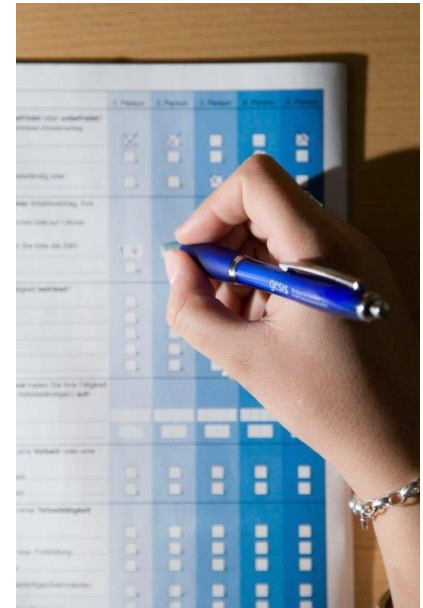
Menold et al., 2013

- Strengths
  - Indicators are derived based on results of the previous studies on real falsifications
  - Method is evaluated when using a large amount of data, which are surely falsified
  - Approach to identify "at risk" interviewers
- Limitations
  - Method evaluation in an experimental setting
  - Can not be used as a stand alone method
  - Application for cross-cultural comparability is limited, as no empirical test of the method in cross-cultural context is available

Statistical Analyses (Blasius & Thiessen, 2013)

- The same response (strongly agree) to a set of related variables: simplifying
- CatPCA: identical factor scores as simplifying
- Was found to be correlated with country (World Value Survey)
- ALLBUS 2008: Three interviewers with 2 to 7 respondents with identical response patterns
  - The interviewers are likely falsifiers / simplifying the task
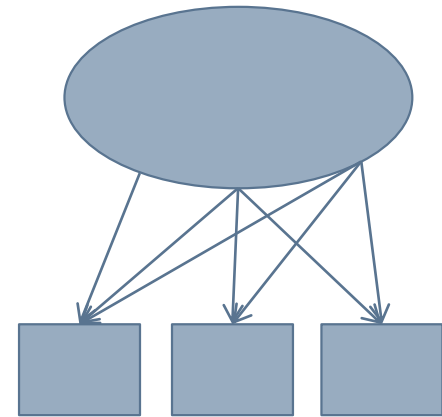
Blasius & Thiessen, 2013

- Strengths:
  - CatPCA as a method to identify simplifying in the data
  - Correlations of simplifying with countries/ interviewers

- Limitations:
  - Method has not been evaluated on the data with known falsifications
  - differentiation between interviewers′ and respondents′ effects is questionable
  - Interviewer effects are confounded with area effects
  - Cannot be used as a stand alone method

Conclusions

- Re-Interviews is the method of choice, but random selection is ineffective
- Problem of selection of "at risk interviewers"
- Possible Procedure: combination of methods
  - Use interviewers' work tracking (e.g. GPS-data)
  - Use para-data
  - Use statistical analysis methods which are evaluated on known falsified data during the field phase
  - Contact respondents and interviewers to verify suspected cases
- Problem of verification
  - Respondents' availability and memory problems
  - Interviewer has to verify falsifying behaviour

Cross-Cultural Perspective

- Research comes mainly from the US and Germany

- (Cross-cultural) research findings by Blasius & Thiessen: link to falsfications by interviewers should be evaluated

- Controlled experiments like those by Menold et al. (2013) in the cross-cultural context are needed

References

- AAPOR: Interviewer Falsification in Survey Research: Current Best Methods for Prevention, 2003, Detection and Repair of its Effects. Available under: http://www.aapor.org/Content/NavigationMenu/ResourcesforResearchers/falsification.pdf (Access: 18.11.2010).

- Blasius, Jörg; Thissen, Vistor: Detecting Poorly Conducted Interviews In: Winckler et al. (Hg):67-88

- Bredl, Sebastian, Storfinger, Nina. & Menold, Natalija. : 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, 2013, pp. 3-24. Frankfurt: Peter Lang Academic Research.

- Benford, Frank: The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, 1938, 78 (1), 551–572.

- Bredl, Sebastian, Winker, Peter and Kötschau, Kerstin: A Statistical Approach to Detect Interviewer Falsification of Survey Data , *Survey Methodology*, 2012.

- Bushery, John M., Reichert, Jennifer W., Albright, Keith A. and Rossiter, John C.: Using Date and Time Stamps to Detect Interviewer Falsification, in: Proceedings of the American Statistical Association (Survey Research Methods Section), 1999, 316–320.

- Case, Peter B.: How to Catch Interviewer Errors, *Journal of Advertising Research*, 1971, 11 (2), 39–43.

- Murphy, Joe, Baxter, Rodney, Eyerman, Joe, Cunningham, David and Kennet, Joel: A System for Detecting Interviewer Falsification, Paper Presented at the American Association for Public Opinion Research 59th Annual Conference, 2004.

- Schräpler, Jörg-Peter and Wagner, Gert G.: Identification, Characteristics and Impact of Faked Interviews in Surveys – An analysis by means of genuine fakes in the raw data of SOEP, IZA Discussion Paper Series, 969, 2003.

- Hood, Catherine C. and Bushery, John M.: Getting more Bang from the Reinterviewer Buck: Identifying 'At Risk' Interviewers, in: Proceedings of the American Statistical Association (Survey Research Methods Section), 1997, 820–824.

- Koch, Achim: Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994, *ZUMA-Nachrichten*, 1995, 36, 89–105.

- Krejsa, Elizabeth A., Davis, Mary C. and Hill, Joan M.: Evaluation of the Quality Assurance Falsification Interview Used in the Census 2000 Dress Rehearsal, in: Proceedings of the American Statistical Association (Survey Research Methods Section), 1999, 635–640.

- Schreiner, Irwin, Pennie, Karen and Newbrough, Jennifer: Interviewer Falsification in Census Bureau Surveys, in: Proceedings of the American Statistical Association (Survey Research Methods Section), 1988, 491–496.

- Storfinger, Nina and Winker, Peter: Robustness of Clustering Methods for Identification of Potential Falsifications in Survey Data, ZEU Discussion Paper Nr. 57, 2011.

Contact :
natalja.menold@gesis.org