New Frontiers in Detecting Data Fabrication

Michael Robbins[1,2]

[1] Princeton University

[2]University of Michigan

21 June 2018

Overview

Importance of Ensuring Data Quality

- The ability to make valid inferences in the social sciences depends on collecting reliable and accurate data

Fabrication in Survey Research

- "Almost every interviewer will eventually succumb if the incitements to fabrication are made overpowering enough, if fabrication is made to appear the only practicable solution to the problems facing the interviewer."
  - Leo Crespi, 1945

Defining the Problem

- Data fabrication is an intentional deviation from the stated guidelines, instructions or sampling procedures by any member of the survey project, including interviewers, supervisors, data entry personnel, the project leaders or the principal investigator, that results in a contamination of the data.

Defining the Problem

- Fabrication includes, but is not limited to, *intentionally*:
  1. Selecting the wrong respondent
  2. Misreading the question
  3. Misrecording a response
  4. Creating data

Distinguishing Fabrication from Survey Error

- Examples of Fabrication
  1. Intentionally selecting a house not selected by the sampling plan but where people are visibly home
  2. Intentionally entering the same questionnaire twice into the data set
  3. Intentionally reading a 'do not read' response category to the respondent
- Examples of survey error
  1. Accidentally selecting the incorrect house by miscounting the skip pattern
  2. Accidentally entering the same questionnaire twice
  3. Accidentally reading a 'do not read' response category to the respondent

AAPOR's Assurance

- "Where appropriate methods are used, interview falsification is rare, involving only a small percentage of interviewers and a substantially smaller percentage of interviews."
  - AAPOR 2003 report on fabrication
- AAPOR recommendations focus on:
  1. Sampling methodology
  2. Interviewer oversight
  3. Data entry process
  4. Callbacks

Evaulating AAPOR's Assurance

- Estimates of the percentage of surveys containing fabricated data vary widely:
  - Less than 1% to $\approx$ 5% of survey observations in developed countries (Biemer & Stokes 1989; Case 1971; Koch 1995)
  - Fabrication by 6.5% of U.S. Census interviewers (Schreiner et al. 1988)
  - Up to 25% of surveys in non-OECD countries may contain a significant percentage of fabricated observations (Bredl, Kotschau, & Winker 2012; Kuriakose & Robbins 2016)

Effects of Fabrication

- Even small levels of fabrication can have big effects
  - In the German SOEP survey, removing the 2.5% of fabricated observations changes the effect of years of education on log gross wages changes by about 80% (Schrapler and Wagner 2003)
  - If roughly 1% of the sample is fabricated via duplication, the probably of obtaining unbiased statistical estimates is 41.6% and if 10% or more of the sample is fabricated, the likelihood falls to 11.4% (Sarracino and Mikucka 2016)

Key Arguments

1. Fabrication is an underreported problem
2. Fabrication is not only a problem with interviewers
3. Fabrication should be incorporated into the Total Survey Error (TSE) framework
4. The field needs an honest conversation about fabrication
5. Survey leaders must cooperate to prevent fabrication

Arab Barometer

The Arab Barometer: About the Surveys

- More than 45,000 face-to-face interviews
- 37 nationally representative surveys to date
- Four waves across 15 countries
  - Wave 1 (2006-7) in 7 countries
  - Wave 2 (2010-11) in 10 countries
  - Wave 3 (2012-14) in 12 countries
  - Wave 4 (2016-17) in 8 countries
  - Wave 5 (2018-19) planned in up to 15 countries
- Data are publicly available for download and analysis at `arabbarometer.org`

The Arab Barometer: The Fifth Wave

- Largest publicly available survey ever conducted in MENA
- Face-to-face surveys using area probability samples
- Approximately 30,000 respondents
- Covers 90% of MENA's population
- Data publicly availble in mid-2019

Surveys by Wave

Country	2006-9	2010-1	2012-4	2016-7	2018-9
Algeria	✓	✓	✓	✓	✓
Bahrain	✓				
Egypt		✓	✓	✓	✓
Iraq		✓	✓	✓	✓
Jordan	✓	✓	✓	✓	✓
Kuwait			✓		✓
Lebanon	✓	✓	✓	✓	✓
Libya			✓		✓
Morocco	✓		✓	✓	✓
Palestine	✓	✓	✓	✓	✓
Qatar				✓	✓
Saudi Arabia		✓			✓
Sudan		✓	✓		✓
Tunisia		✓	✓	✓	✓
Yemen	✓	✓	✓		✓

The Arab Barometer: Question Areas

- Economic issues & personal well-being
- Evaluation of political institutions & political attitudes
- Engagement in politics
- Traditional & new media
- Governance
- Culture & religion
- International relations
- Current affairs
- Special batteries
- Demographics

Logic of Fabrication

Logic of Fabrication

- Fabrication is a crime of opportunity
- Fabricator generally makes a cost-benefit calculation
- Likelihood of being caught is a key consideration

Fabrication by Mode

- Rates of fabrication vary based on challenges to oversight (Blasius and Thiessen 2015)
  - Web surveys avoid many potential sources of fabrication
  - Call centers allow for significant oversight
  - Fabrication is more likely in face-to-face surveys where oversight challenges are greater

Motivations for Fabrication

- Possible motivations for fabricators
  1. Save time and money
  2. Security of the fieldwork team
  3. Length of instrument
  4. Simplification for the respondent
  5. Sensitive questions
  6. To cover up a mistake
  7. Normative beliefs
  8. 'Good-hearted' reasons
  9. Lack of incentive to improve (for a survey firm)

Who Fabricates?

- Interviewers
- Data entry personnel
  - Helping out an interviewer who missed an interview
- Firm leaders
  - Increase profit margin
  - Lower costs in a competitive marketplace
  - Time pressure
  - Relative impunity

Detection

Observational Methods

- Have supervisors attend interviews
- Record portions of the interview for review by principal investigators
  - Evaluation is time consuming and may require language skills that are unavailable to principal investigators
  - Privacy concerns for sensitive questions

Recontact

- Done by phone, mail, or face-to-face
- Can be used to definitively identify fabrication
- Higher costs, especially if done face-to-face
- Should target potentially suspect observations
  - China Mental Health Study flags suspicious interviews & targets callbacks on this subset

Data Analysis Methods

- Comparing to Benford's Law
- Unusual patterns in the data
- Rare response combinations
- Undifferentiated response combinations
- Short paths through the survey
- High percentages of missing data or incomplete interviews

Use of Paradata

- Interview duration
- Duration between interviews
- Close to deadline
- Time of day
- Surge of interviews
- Missing phone numbers

Limitations of Data Analysis Methods

1. Can flag observations but does not always prove fabrication
   - Using multiple tests, Menold and Kemper (2014) can only identify fabrication with 75% rate of accuracy when real and fabricated observations are known
2. Principal investigators must decide if fabrication is the most likely cause
   - CAPI is giving greater leverage to prove fabrication

Prevention

Preventing Fabrication Before It Happens

- Instrument design techniques
  - Consider interview length
  - Consider techniques for asking sensitive questions
  - Consider when interviewer may be incentivized to simplify for respondent
  - Consider question complexity
  - Consider country context

Preventing Fabrication Before It Happens

- Training
  - Thorough training with independent party present if possible
  - Train extra interviewers and select top performers
  - Convince interviewers that fabricators will be detected and punished
  - Provide interviewers and firm with necessary support

Preventing Fabrication Before It Happens

- Structuring financial incentives
  - Reward interviewers for quality not quantity
  - Structure payments to firms to increase their incentives to prevent fabrication
  - Convince firms that the survey is part of a multiround game

Detecting Fabrication During Fieldwork

- Employment of oversight procedures
  - Ensure strong oversight techniques, especially in early days of fieldwork
  - Have an independent representative on ground during fieldwork
- Anaylsis of partial data
  - Request partial data sets and use data analysis methods to detect problems earlier

New Frontiers

Use of Technology

- Computer-assisted personal interviewing (CAPI)
  - Automatic collection of paradata
    - Multiple timers
    - GPS coordinates
    - Recording
  - CAPI Limitations
    - Collects a vast amount of data that can require significant resources to process
    - Does not make fabrication impossible

Real-Time Detection

- Example: **F**ieldwork **A**lgorithm for **L**APOP **C**ontrol over survey **O**perations and **N**orms
  - Geofencing
  - Interviewer identity verification
  - Respondent gender verification
  - Automated contact tracking
  - Question timing verification
  - Recording (with respondent's permission)

Fabrication via Duplication

- Fabrication is based on real interviews, meaning it is harder to detect
  - Observation is internally consistent
  - Correlations hold across variables
  - Rare response combinations unlikely
  - Profiles can be copied to make up for hard-to-reach demographics

Near Duplicates

- Affinity of observations is the maximum percentage of variables that an observation shares with any other observation in the survey (percent match).
- If observation A shares 99% of values with observation B and shares a lower than 99% match with all other observations, the percentage match for observation A would be 99%.
- *PercentMatch* is a Stata program designed to test for maximum shared affinity between any two observations (see Kuriakose & Robbins 2016)

Flagging Observations using PercentMatch

- Fabricators are likely to repeat profiles when filling out the survey (Waller 2013)
- Interview simplification techniques are likely to yield less differentiated response patterns
- Straight-lining is likely to be detected
- Slight deviations from protocol may be detected

Empirical Analysis

Level of Near Duplicates Across 1,008 Country-Year Surveys

Degree of Likely Fabrication	Percentage of Surveys
No cases	35.8%
1% to <5%	46.8%
5% to <10%	7.2%
$\geq$ 10%	10.1%

Kuriakose & Robbins 2016

Near Duplicates by Country Type

Percentage of Observations Exceeding 0.85 Percent Match

Level of Near Duplicates by OECD Status

Degree of Likely Fabrication	OECD	Non-OECD	Difference
No cases	52.6%	17.3%	*+35.2 pts.*
1% to <5%	42.9%	56.5%	*-13.7 pts.*
5% to <10%	1.9%	10.8%	*-8.2 pts.*
≥ 10%	2.0%	15.3%	*-13.3 pts.*

Kuriakose & Robbins 2016

Use of PercentMatch

- Program is now widely used by cross-national projects to improve data quality
  - Has found a double-posted data set
  - Has been used to identify suspect interviewers
  - Has identified data entry personnel who cut corners
  - Has identified firms that have engaged in massive data fabrication

A Way Forward

Admitting there is a Problem

- Allow for an honest discussion about techniques to prevent fabrication

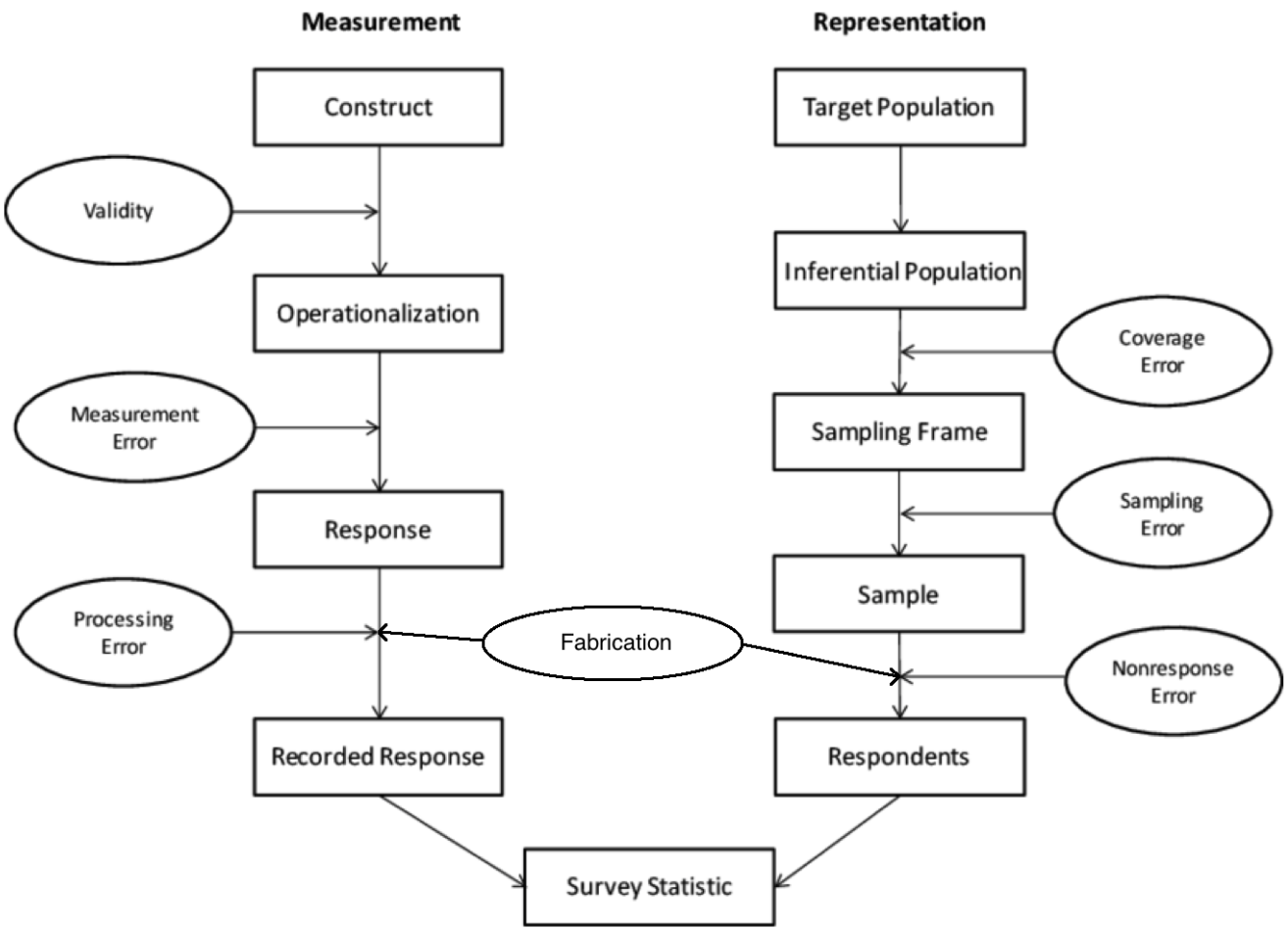
Updating Our Assumptions

- Crespi revised—
  Almost every **person or firm** will eventually succumb if the incitements to fabrication are made overpowering enough, if fabrication is made to appear the only practicable solution to the problems facing the **person or firm**.

Thinking through the Process of Fabrication

- Culprit may be:
  1. Interviewers
  2. Data entry team
  3. Quality control team
  4. Firms leaders
  5. Principal investigators
  6. Project leaders

Revising the TSE Framework



Changing the Approach to Fabrication

- Finding fabrication is akin to detective work
  - *Prove* the data are real rather than *assume* they are real
  - Determine the process most likely to have produced these data
  - Understand detection of fabrication is both an art and a science
    - It may not always be possible to definitively prove if the source of the problem is fabrication or survey error
    - Attempts to address the problem rely in part on this determination

Increasing Cooperation within the Field

1. Make surveys a multiround game, especially in developing contexts
2. Share information between projects to change firm incentives
3. Make data publicly available
4. Change the narrative about fabrication to one that is considered a normal (albeit regrettable) source of error in the survey process

Arab Barometer

www.arabbarometer.org
@arabbarometer
@ArabBarometer