Introduction

We all care how about how well the data we analyze represents the population that we claim it to be representative of.

Those of us who produce data frequently consider fundamental issues such as selection bias and response bias.

Another issue which has the potential to seriously affect how representative our data is, is interview falsification.
The traditional approach

Traditional methods have relied on identifying things like

- Implausible combinations of responses

- Differing levels of variability in responses between interviews

However, even when such methodology is used to flag suspicious cases, it is often difficult to prove that cheating occurred.
Newer methods

CAPI has become the industry standard in many countries, and tablets can be used to record GPS coordinates, timestamps, and even audio files.

This data can allow us not only to flag suspicious interviews, but to substantively prove falsification.
Data Collection
A Brief Overview

The data we collect

How we use it
Quality control data

GORBI's software records data for falsification checks throughout the interview…

- GPS coordinates for each entry
- Timestamps for each entry
- Full audio recording
- Automatically selected respondent
GPS coordinates and timestamps

Every time an interviewer enters anything into the tablet, the GPS coordinates of his or her location and a timestamp are uploaded to GORBI’s servers.
Respondent selection

For most surveys, respondents are selected via a modified Kish Table procedure.

Names, birth dates or ages, and genders of adult household members are entered into a table by the interviewer.

GORBI's software randomly selects one person, records his or her information, and instructs the interviewer who was selected.
13.
**Contact Form for Face-to-Face Mode**

**European Values Study**

<table>
<thead>
<tr>
<th>6. კონტაქტის მდგომარეობის შეყვანა</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = აქტიური მოქალაქე</td>
<td>4 = კონტაქტის მდგომარეობი რამდენიმე სახლში, ხშირად აერ</td>
</tr>
<tr>
<td>2 = ანქარი</td>
<td>5 = კონტაქტის მდგომარეობი, სახლში რამდენიმე ხაჩი</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. კონტაქტის შეყვანა (შეიძლება იქ შემოგვთავაზოთ თუ ქართულ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. მერქვადის დაარსება ორ განთავსებით</td>
<td>4. რამდენიმე რამდენიმე ღირსშენებული, როდესაც იმ თითოეული ადგილ/ გადგილი</td>
</tr>
<tr>
<td>2. საადგილო რამდენიმე სახლი ორი პირ ორი ორი მონახული სახლში, როდე</td>
<td>5. იხილათ იქით ზოგიერთი მონახული სხვა მართვის გამო (თუმცა რეწომ არ არის ურთიერთშ</td>
</tr>
<tr>
<td>3. რამდენიმე რამდენიმე მონახული სხვა მართვის გამო (თუმცა რეწომ არ არის ურთიერთშ</td>
<td></td>
</tr>
</tbody>
</table>

**გათვალისწინებით**

- 14.
Data Processing

Our software automatically flags interviews if they meet any of the set QC criteria.

If flagged, then the data processing team reviews the interview to determine if falsification occurred.

A few examples are provided below:
Short interview or rapid-fire questions

After the pilot, we set cutoffs for the following:
- Minimum interview duration
- Maximum proportion of questions answered in <3 seconds

For these cases we listen to the audio files, and most often find that questions were skipped.
Interview location change

A location change of more than 10 m during the interview

Poti-Zugdidi

Zugdidi

Poti
Odd contact attempt patterns

For surveys with multiple contact attempts, our software flags the following:

- A contact attempt made more than 100 meters away from any other contact attempt for the same household
- A contact attempt made before 9:00 AM or after 9:00 PM
- A contact attempt made too soon after the previous contact attempt
<table>
<thead>
<tr>
<th>No.</th>
<th>Event</th>
<th>Value ID</th>
<th>Date/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Session I start</td>
<td>222</td>
<td>2017-12-20 17:26:36</td>
</tr>
<tr>
<td>2</td>
<td>Session II start</td>
<td>222</td>
<td>2017-12-20 17:29:07</td>
</tr>
<tr>
<td>3</td>
<td>Session III start</td>
<td>222</td>
<td>2017-12-20 21:27:13</td>
</tr>
</tbody>
</table>

**Variables:**
- Inquirer ID
- Inquirer ID
- Operator ID
- Session I start
- Session II start
- Session III start
- Completed
- Respondent Name
- Respondent Address
- Operation Code
- Manager Name
- Store ID
- Store Code
- Double ID
- Double Code
- Double Start
- Double End
- Empty Code
- GPS 1
- GPS 2
- GPS 3

**GPS Data:**
- 2017-12-20 17:26:36
- 2017-12-20 17:29:07
- 2017-12-20 21:27:13
The full picture

The second contact attempt was made more than 100 m from the first, and only 3 minutes after it.

There was a third contact attempt made that evening, even further away.

The interviewer recorded result 1 as "no one home," result 2 as "completed interview," and recorded no result 3.

The timestamps on the answers to all questions indicate that they were entered in sequential order after the second contact attempt.
GORBI's response

We contacted the interviewer. She explained that she found no one home, but on her way out she encountered the single resident of the household. He agreed to an interview, but was in a hurry and during the interview she felt rushed. At the end of the interview she did not save it as completed, and reviewed the answers at home that evening before doing so.

Other data supported her: GPS stamps on the answer uploads matched contact attempt 1, the audio file contained a full interview, and all of her other interviews for the project looked good.
Data Analysis

As there is little literature on the dynamics of interviewer cheating, this analysis was purely exploratory.

Who falsifies interviews?

Looking at accepted interviews
Dynamics of interview falsification

The dynamics of interview falsification in Georgia were very simple, but we had a comparable data set which allowed us to examine the issue in more detail.
EVS Georgia: A short story

Almost all of the interviews rejected from the 2017 EVS data set in Georgia were interviews conducted in the ethnic minority languages, Azeri and Armenian.
Explanation

GORBI has been using this software since 2011, and has a regular group of interviewers from which problem interviewers have long been weeded out.

An average of about one client per year budgets to sample in the minority languages, so we don't offer regular employment to that demographic, and frequently need to hire new staff.
Next question

Our Georgian interviewer pool does not provide sufficient data with which to explore patterns of interview falsification in detail.

However, in early 2019 we conducted a survey for another client with an almost identical sample design.

This survey was conducted in Armenia, with a local partner who had not used our software before.

I decided to look at the results of this survey to see what the nationwide dynamics of interview falsification in a survey like EVS would be under those conditions.
Non-EVS Armenia

The dynamics of interview falsification in this survey were more interesting...

Date

Settlement type

Training and management

Types of falsification
Types of falsification

A new issue

Result

Cheating during household/respondent selection vs. cheating during the interview
A new issue

Most of the new local partner's interviewers were experienced, but they were not used to conducting surveys with fixed target sample sizes and repeat contact attempts.

Early in the survey we noticed PSUs with very low incidences of contact attempt results that required revisits, often combined with extra contact attempts, and had to expand our quality control protocol.
Result

While we rejected only 6% of the 1,321 completed interviews for falsification (74 interviews), we rejected 11% of the 2,857 total contact attempts (314 contacts).

For the purpose of this exploratory analysis I treated them separately, referring to them as interview and procedural falsification, respectively.
Date

I expected falsification rates to decrease sharply as problem interviewers were removed and all interviewers became aware of how intensively they were being monitored.

This was true, but there was one unexpected finding when looking at dates...
Training and management

Our new local partner trained interviewers in two locations, and broke the country down into five regions for fieldwork management.

Training location clearly matters...

Syunik

Falsification rates
Settlement type

Out of curiosity I looked to see how falsification rates varied between capital, urban, and rural areas.

This survey also included a settlement size variable with eight categories, and there was a similar trend there.
The more rural the settlement, the more effort needs to be expended to return to a household for another contact attempt.

The more urban the settlement, the greater the proportion of the interviewer's time that is spent actually conducting the interview.
Accepted interviews

GPS data seems to me to be useful primarily for detecting falsification.

Audio recordings are not used for anything else, although we are caching them for the development of voice recognition software.

Timestamp data, however, might be useful for learning more about interview quality...
Conclusions

How can we use this information to improve our ability to detect interview falsification?

Changes required of our local partner

Allocation of our own monitoring resources
Changes required

After conducting this analysis, we decided to require our local partner to conduct all trainings in the capital.

Both of the contracts we've given them since then have only required single contact attempts, but before conducting another EVS-type survey there we will work with them to develop a detailed training module on selection protocols.
Allocation of monitoring resources

Georgia

Armenia
Georgia

In addition to interviews flagged by the QC program, we usually flag a randomly selected subset of interviews to be listened to.

These findings suggest that in Georgia we should assign a higher selection probability to interviews conducted in minority languages.

We may also consider developing a more sophisticated criteria for flagging quick interviews.
Armenia

We can allocate more of our randomly assigned monitoring effort towards unsuccessful contact attempts, particularly household-level refusals.

We can give contacts attempts in rural areas higher selection probabilities.
Ideas
How might these findings be relevant to EVS?

The example of survey non-response

Ideas for analogous measures
The evolution of the documentation of survey non-response

**EVS 1981:** No information available on response rates.

**EVS 1990:** Nationwide response rates published for approximately half of the countries.

**EVS 1999:** Nationwide response rates published for all countries.

**EVS 2017:** A full data set including non-response cases will be published in the spring of 2020.
Ideas for analogous measures

Could we require summary figures on rejected interviews along with descriptions of verification procedures?

Could we require rejected interviews and rejected contact attempts to be included with future data sets?

Could we request data such as timestamps from countries who are able to provide it?