Comparison of different methods of GPS paradata usage in CAPI surveys for interviewers' monitoring

Daniil Lebedev
Aigul Klimova
Department of Sociology, HSE University

Mobile Apps and Sensors in Surveys Workshop, Utrecht University
22-23 April 2021

Research questions:

- Which factors associate with lower GPS data quality?
- Which of the possible GPS-paradata methods for fieldwork monitoring in computer-assisted personal interviews on tablets are more efficient in terms of detecting "suspicious" (at risk of cheating) interviews?

Background

Separate locations:

- **"Strand length"** - comparison of interview location and that of the sampled houshould [Mohajer & Edwards, 2018; Sikes, 2009] or with interviewers own home [Hasson, 2015]
- **"Geofencing"** - Comparison of locations at the beginning and at the end of an interview [Seeger, 2011; Wang & Biemer, 2010; Mohajer & Edwards, 2018; Choumert-Nkolo et al., 2019]
- **"Curbstoning" test** - checking for presence of too dense groups of interviews' locations [Cecchi & Marquette, 2010; Dajani & Marquette, 2015]

Data

Russian Longitudinal Monitoring Survey (RLMS-HSE) - CAPI:

26th wave – 37 interviewers, 7 regions, 491 interviews

27th wave – 53 interviewers, 9 regions, 631 interviews

GPS locations – information regarding latitude, longitude of a tablet in the beginning and at the end of the interview and accuracy of the measurements (SurveySolutions application) – active measurement

Tablets – Samsung Galaxy Tab A 16.0 SM-T355

Methods

Fieldwork monitoring:

- Geofencing
- Curbstoning
- Interwave geofencing

Distance difference (thresholds):

- 50 meters – mean accuracy of GPS measurements 25 meters
- Accuracy-based – sum of accuracy of both measurements

Data quality:

- Fieldwork monitoring GPS-based methods comparison:
  - Completion times
  - Criterion validity
  - Test-retest reliability
- GPS data quality:
  - Missing data
  - Measurement accuracy

GPS-paradata quality

Missing data

Measurement accuracy

Missing data: 26th wave

Dependent variable: Missing data of location measurements either at the beginning or at the end of the interview - 105 cases (22,3%)

	Sig.	Exp(B)	Sig.	Exp(B)
(Constant)	.002	7344.991	.056	244.740
Age	**.001**	**.910**	.105	.952
Tablet availability	.216	.739	.886	1.044
Confidence with tablet	**.000**	**.549**	**.000**	**.642**
Expectation's index	.238	.903	.488	.940
Confidence with tablet * Age			**.004**	**.986**
Solikamsk (*small region*)	**.000**	**.003**	**.000**	**.006**
Kazan (*medium region*)	.998	.000	.998	.000
Kurgan (s)	**.000**	**.007**	**.000**	**.008**
Volsk (s)	**.002**	**.126**	**.004**	**.142**
Moscow region (m)	.251	.584	.957	.970
Berdsk (s)	**.000**	**.044**	**.000**	**.065**

Moscow – control group

GPS-paradata accuracy: 26th wave

Dependent variable: GPS paradata measurement accuracy (in meters)

Mean accuracy – 23.6 meters (SD = 11.3)

26 wave	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	24.491	1.407		17.407	.000
Solikamsk (s)	-.530	2.235	-.015	-.237	.813
Kazan (m)	**-6.417**	2.583	-.149	-2.484	**.013**
Kurgan (s)	-1.047	1.975	-.036	-.530	.596
Volsk (s)	.295	2.550	.007	.116	.908
Berdsk (s)	-1.022	1.802	-.040	-.567	.571
Moscow region (m)	.622	2.286	.017	.272	.786

Moscow – control group

GPS-paradata accuracy: 27th wave

Dependent variable: GPS paradata measurement accuracy (in meters)

Mean accuracy – 23.4 meters (SD = 9.4)

27 wave	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	22.293	1.037		21.488	.000
Solikamsk (s)	2.080	1.673	.072	1.243	.215
Kazan (m)	**4.811**	1.506	.194	3.194	**.002**
Volsk (s)	**-4.614**	1.640	-.164	-2.813	**.005**
Berdsk (s)	-.797	1.561	-.030	-.511	.610
Saratov (m)	**6.291**	1.602	.232	3.927	**.000**
Nizhniy Novgorod (m)	-1.589	1.781	-.050	-.892	.373

Moscow – control group

GPS-paradata based fieldwork monitoring strategies comparison

Completion time

Criterion validity

Test-retest reliability

Suspicious interviews

	26th wave		27th wave	
	N	%	N	%
Geofencing (50m)	48	13%	18	5%
Geofencing (accuracy-based)	45	12%	23	6%
Curbstoning (50m)	197	51%	213	56%
Curbstoning (accuracy-based)	196	51%	205	54%
Intrawave geofencing (50m)	-	-	39	23%
Intrawave geofencing (accuracy-based)	-	-	38	22%

Completion time

All methods revealed longer completion times for suspicious interviews compared to non-suspicious at least in one wave

Geofencing: Cohen’s d from medium to large effect size (0.40-0.99) in both waves for both types of threshold measure

Curbstoning and interwave geofencing: Cohen's d was small (0.18–0.34) in both waves for both types of threshold measure

Validity and reliability

Criterion validity:

Few significant differences between suspicious and non-suspicious with no clear pattern

Accuracy-based geofencing may be efficient

Test-retest reliability:

Almost none statistically significant differences

Accuracy-based curbstoning: lower test-retest reliability among suspicious interviews compared to non-suspicious in the 27th wave: 0.41 and 0.60, respectively (**$z = 1.3$, $p<0.1$**)

Geofencing (50 m): lower test-retest reliability between suspicious and non-suspicious interviews in the 26th wave (0.38 and 0.53, respectively), though the difference did not reach statistical significance due to the small proportion of respondents flagged as suspicious (**$z = 1.0$, $p=0.15$**)

Outlook

GPS-paradata quality (missing data and accuracy) may vary in connection with:

- **regions** (lower quality in more developed regions – urban canyons [Lemmens 2011; Gong et al. 2012])
- **interviewers' characteristics** (confidence with CAPI)

GPS-based fieldwork monitoring strategies:

- **Geofencing** (accuracy-based and 50 meters thresholds) was efficient in flagging suspicious interviews that have lower data quality (higher completion time, lower criterion validity and lower test-retest reliability)
- **Accuracy-based curbstoning** flagged interviews with lower test-retest reliability
- **Geofencing** (both types of thresholds) tends to identify less interviews as suspicious, but with lower data quality (more efficient)

Limitations

Respondents' addresses were nor available for use

No valid and reliable **measure of fraudulent interviews** was available (only "suspicious")

Some **technical problems** with 27th wave GPS location measures and data on interviewers with high missing data rate

Limited number of interviews and interviewers

Recommendations

Focus on **interviewer's education** while starting using CAPI which can associate with further increase in GPS-paradata quality and with lower rate of «suspicious» interviews

Use **accuracy** as threshold identification for distance between two locations – GPS-data quality may vary in different regions

GPS-paradata should be used in **conjunction with other methods** of fieldwork monitoring – no exact assumptions about fabrications or falsifications may be done based on GPS-paradata analysis only (nonintentional errors or technical difficulties may be present)

Thank you!

For cooperation, questions and comments please contact:
zenon-daniil@yandex.ru
dlebedev@hse.ru

With appreciation to Dr. Aigul Klimova for considerable impact on this research

The study was funded by Russian Foundation for Basic Research grant "Postgraduates" for the project "Using paradata to measure and improve the quality of data in online surveys" (No. 20-311-90073)