Validating Interviewer-observed para-data with auxiliary data from Google Street View

Anina Vercruyssen – Geert Loosveldt
Centre for Sociological Research, University of Leuven

1. introduction

The general objective of a survey is to collect information about all sample units. Mostly a survey is considered as the only or at least the most cost efficient way to achieve this goal. Unit non response creates a serious problem because we need information about sample units for which the most appropriate data collection procedure was not successful due to refusals, non contacts, …. Therefore the assessment of non-response bias can be considered as a challenge and a lot of research about non-response bias is focussed on searching for information that is related to the information that we wanted to collect by means of a survey. The collection of paradata about the survey data collection process and the search for relevant external auxiliary variables that are available for all sample units are two procedures that are used to get information that is necessary for a non-response bias assessment.
In face-to-face household surveys, data about the type of house and neighborhood characteristics of the sample unit are considered as relevant information to assess non-response bias. One can ask interviewers to register the house type, the presence of impediments (e.g. intercom), the physical state of the houses, the presence of vandalism/graffiti and litter/rubbish on a contact form during the contact procedure. It is an additional task for the interviewers during the data collection process and the information is available for respondents and non-respondents. Therefore one can consider this information as interviewer-observed para-data. The same kind of house and neighbourhood characteristics can be obtained by using Google Street View as an external source of information. These variables can be coded for all sample units prior to the data collection process. So these variables can be considered as auxiliary data. Some efforts have already been made to try to assess the accuracy of such paradata (e.g. with auxiliary census data, for an overview, see West & Sinibaldi, 2013; also see Pickering et al., 2003; Sinibaldi et al., 2013; Walsh et al., 2013; West, 2013; West & Kreuter, 2013)

Studies that straightforwardly use auxiliary data based on GSV to cross-reference and validate interviewer observations are not new but rare in (social) research (e.g. evaluation of data gathered by listers of sample frames (Eckman & Kreuter, 2013); assessment of neighbourhood indicators in health and epidemiological studies (Clarke et al., 2010; Odgers et al., 2012; Rundle et al., 2011)). So far, there is hardly information available on how the coding procedure works for turning visual GSV-data into useable auxiliary data for surveys.

In this paper we pay extra attention to this process before setting out to explore how useful Maps- and GSV-data can be as auxiliary data to validate interviewer observations and to predict nonresponse in the Belgian data of ESS Round 7 (2014, from her on abbreviated to ESS7BE).

2. Methods

2.1. Data

For this study, we use the interviewer-observed paradata variables from the contact forms of ESS on the type and state of houses and the state of the neighbourhood. The same variables were constructed with auxiliary data from GSV and, additionally, we have variables on distances from Maps.

To explore the usefulness of GSV-data, we took a 20% random sample (N=640) from the ESS7BE-sample. This 20%-subsample is a stratified sample with the same distribution as the main sample according to their final disposition coding (Interview: 55,2%; Non-contact: 5.4%; Refusal: 26.1%; Other non-response: 10.1%; Ineligible: 3.2%). The stratified subsample contains interviewer observations from 126 different interviewers, each being assigned to between 1 and 8 sample units.

2.2. Variables

ESS7 contact files and interviewer observations

As recurrently the case in ESS, interviewers were required to register the house type, the presence of impediments (intercom, locked door/gate or both), the physical state of the houses (5-point scale from 'very good' to 'very bad'), the presence of vandalism/graffiti and the presence of litter/rubbish (4-point scale from 'very large amount' to 'none or almost none') in the contact forms.
In the contact files, we also find the outcomes of each visit. We focus on the final outcome, distinguishing interview obtained, non-contact, refusal, other non-response and ineligible as defined by the AAPOR guidelines (2009).

Google Street View observations

For coding the GSV-data, an anonymized address lists was used with no other variables to enable blind coding. The coding was done by one single coder who first of all registered *how* the exact addresses was found in GSV. Once the house was identified, the same instructions and coding scheme were followed as given to the interviewers to describe the house and neighbourhood. The month and year of the GSV-images were also registered. We define images as outdated when they are from before 2014, the year of the fieldwork for ESS7BE.

Google Maps observations

As GSV is accessed through entering the address in Maps we could also instantaneously use this location in Maps to calculate distances to inform us about the reachability of the sample units. We measured the distance to walk to the nearest train station and bus/tram/metro stop, and the distance to drive to the nearest motor way entrance by car.

2.3. Pitfalls of coding Maps and GSV-data and possible solutions

Using Maps and GSV may be simple in daily life, but it is more challenging to use it to generate auxiliary data. Finding an address in Maps does not guarantee finding the house (easily) in GSV and finding it in GSV does not mean all the observations can be made. We discuss the pitfalls we encountered.

1) Coding auxiliary data from Maps and GSV is more time consuming than expected: On average, twelve houses were coded per hour.

2) GSV can be completely unavailable in some streets:
Although GSV has an extremely high coverage in Belgium, a few streets in more remote villages were not available (yet). GSV was not available for 44 cases.

3) Maps and GSV can be off by a few houses:
When GSV is available for the address entered in Maps, you may end up a few houses further or find yourself facing the wrong side of the street. In 38 cases, we were a few dozen houses off (Table 4) and in 44 cases, we just did not have enough information to determine the exact or approximate location of the house in GSV.

4) GSV can hit physical boundaries:
GSV-cars go as far as the roads allow them and GSV can only give you a glimpse from a distance, as was the case with 18 addresses in our subsample.

5) GSV-images can be too pixelish, leaving smaller features (e.g. house number, intercom, multiple doorbells, …) too blurred to identify.

6) GSV can be censored:
Houses can be deliberately pixelish. People have the right to demand that their houses get blurred on GSV to respect their privacy (Google, 2015).

7) GSV can be outdated:
Although GSV makes regular updates, we only had up-to-date images from the year of the fieldwork (2014) in 123 cases. Often, 2009 was the most recent year (243 cases).

8) Google Maps details can be unavailable:
Maps was used for coding the distances to the nearest public transport options and motor way entrances. We have 439 cases (68.6%) for which we could measure all three distances in Maps.

The experienced pitfalls illustrate that using Maps and GSV-data as auxiliary data is not as easy and straightforward as using it in regular daily life. We will assess how useful the obtained data from Maps and GSV is to improve our knowledge on survey data collection processes.

3. Results

Given the pitfalls mentioned in the previous section we could only spot the exact house number or that of the direct neighbour in 62.3%. 23.6% of the houses were found in less exact ways and for 14.1% we did not find the house or have useable images in GSV.

3.1. Validating interviewer observations with auxiliary data from GSV

Table 1: Concordance between interviewer observations and GSV-observations in ESS7BE

Observations	exact match	lenient match	Total N
Type of house	70.9%	78.7%	588
State of house	44.6%	78.3%	561
Litter	75.2%	76.7%	576
Vandalism	87.1%	88.0%	575
Impediments	43.4%	83.9%	447

70.9% of the house types are an exact match between our GSV-observations and the interviewer observations. The most common mismatches occurred with multi-unit buildings. When we are more lenient and ignore misclassifications of the subtypes of detached houses (farms versus regular detached) and multi-unit buildings (apartments, flats, student house, or retirement homes) that can be hard to visually distinguish even when physically present at the location, we reach 78.7% concordance. Having a match for the house type did not significantly differ with regard to how we found the house in GSV.

Matching the coding for the physical state of the houses was harder, which may be due to the rather subjective nature of these judgements. When we are more lenient and distinguish between '(very) good' versus 'reasonable' to '(very) bad' house states, we end up with 78.3% concordance. The level of agreement for the exact type(s) of impediments was similarly low, but lenient matching by distinguishing if any impediment was observed or not brings us to 84% concordance for the cases without missing codes in GSV. Higher levels of agreements were found for presence of litter and vandalism. The differences between the exact match (categorical: amounts of litter and vandalism) and the lenient match (binary: presence of litter and vandalism or not) are very small. It should be noted, however, that neither the interviewer nor the GSV-coder often observed litter or vandalism.

3.2. Predicting nonresponse with auxiliary data from GSV

We also wanted to compare which data works best to predict non-response: GSV-data or paradata from interviewer observations? We restrict the presentation of the results in table 2 to the impact of these data on (non-)contact rates. Model 1 excludes the observations on impediments as we otherwise lose quite some cases with the GSV-data due to missing values, Model 2 includes the impediments resulting in 101 less cases in the regression models. In Model 1 we see that the GSV-data predict more than the interviewer observations for contact: when littering and vandalism was observed in GSV, the sample units turned out to be more likely to be non-contacts while the interviewer observations do not show such significant effects. The effect of litter or vandalism on contactability is also in line with results from other studies (see e.g. Beullens, 2013, Billiet et al., 2009; Blom et al., 2011).

Table 2: Predicting (non-)contact in the 20%-subsample of ESS7BE with GSV-observations versus interviewer observations

	GSV						**Interviewer Observations**					
	Model 1			**Model 2**			**Model 1**			**Model 2**		
CONTACT	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Constant	3.063	.235	***	3.931	.468	***	2.865	.223	***	4.476	.597	***
Litter (yes. ref.cat = no)	-.831	.426	*	-.709	.495		.918	.925		.794	1.064	
Vandalism (yes. ref.cat. = no)	-1.13	.518	*	-.703	.647		-1.14	1.015		-.691	1.169	
State of house (reasonable to (very) bad. ref.cat.= (very) good)	.298	.577		1.478	1.053		-.448	.440		-0.485	0.503	
Impediments (yes. ref.cat = no)				-1.535	.525	**				-2.331	.633	***
Nagelkerke	0.049			0.128			0.014			0.156		

In Model 2, both the GSV-data and interviewer observations of impediments significantly predict non-contact in ESS7BE. That impediments hamper response and contact is also in line with the literature on contactability (see e.g. Blom et al., 2011; Groves & Couper, 1998).

The impact of the GSV-data and interviewer observations on response is limited. There is only a borderline significant effect of vandalism in model 1with the GSV- data. This effect disappears after controlling for impediments (Model 2). For both types of data there is a significant effect of impediments on (non-)response (Not in table 2).

Neither the GSV-data nor the interviewer observations significantly predict refusal. In both models there is no significant effect of the dependent variables.

3.3.Predicting nonresponse with auxiliary data from Maps

As for using Maps auxiliary data about the distance to the nearest train station, bus stop and motor way entrance to assess potential issues with regard to the reachability of sample units, we did find significant effects of more distance to the nearest train station leading to a higher odds of contact. This may be explained by the distance to train stations usually being longer in non-urban areas, which typically generate better contact rates than urban areas. Else, we saw no significant effects of distances to public transportation and motorway accessibility on the response processes in our sample.

4. Discussion

Using Google Street View seems like a straightforward, cost-saving approach for gathering auxiliary data for surveys and cross-checking interviewer-observed paradata. However, in practice, we encountered quite some pitfalls.

Although it is hard to say what the actual cause of discrepancies between the GSV-observations and the interviewer observations is, we need to note that the GSV-coder only had to focus on the observations whereas the interviewers also need to focus on establishing contact and, even more important, obtaining an interview. As such, the GSV-coder may have more time and motivation to create these supplementary data. Moreover, the GSV-coder registers the observations immediately when spotting the house, whereas we do not know whether interviewers always fill in the contact forms immediately at the location. Although we cannot provide evidence for the latter statements, we do believe that coding of the type and state of the houses, presence of impediments and the state of the neighbourhood by one, single coder leads to more stable and reliable data than the current interviewer observations made by a large group of different individuals.

However pitfalls of using GSV-data currently affect the ease of use and the quality of this auxiliary data to a high degree. The coding takes more time than expected, the picture quality issues of outdatedness and pixelishness do not allow yet to actually validate the interviewer-observed paradata, GSV-data does not predict non-response that much differently than interviewer observations and the distances calculated with Maps also do not really bring new insights. Probably it is only relevant to collect specific distance information based on Maps when this information is closely related to the topic of the survey (e.g. mobility survey).

Nonetheless, it is a promising, freely available source of auxiliary data that allowed us to get some insights in potential improvements with regard to filling in the contact forms correctly that should be addressed in the upcoming interviewer training sessions for ESS Belgium. Furthermore, the GSV-data seems to enable the detection of potential contactability issues even before the actual fieldwork starts. Although it does not seem to help us to tackle the bigger non-response problem of refusal (26.1%) in comparison to contactability issues (5.4%), GSV can be useful to e.g. predict contact issues in advance of the fieldwork and foresee extra contact attempts when litter, vandalism or impediments are spotted in GSV. Also, with more updates and technological improvements, GSV auxiliary data may soon become easier to process and more reliable to use.

So when we take into account different aspects of GSV as a sources of auxiliary variables (data quality and predictive power) there are mixed feelings and we recommend careful use of this kind of data.

References

Barbier, S., Wuyts, C., Italiano, P., & Loosveldt, G. (2016). *European Social Survey Round 7 Belgium – Process evaluation for the data collection*. Leuven: Centre for Sociological Research, KULeuven.

Beullens, K. (2013). *The use of paradata to assess survey representativity. Cracks in the nonresponse paradiMaps.* Leuven: KULeuven.

Billiet, J., Vehovar, V., Beullens, K., & Matsuo, H. (2009). Nonresponse Bias in Cross-national Surveys: Designs for Detection and Adjustment in the ESS. *ASK. Research&Methods*, (18), 3-43.

Blom, A. G., de Leeuw, E. D. and Hox, J. J. (2011). Interviewer effects on nonresponse in the European Social Survey. *J. Off. Statist.*, 27, 359–377.

Clarke, P., Ailshire, J., Melendez, R., Bader, M., & Morenoff, J. (2010). Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health & place*, *16*(6), 1224-1229.

Eckman, S., & Kreuter, F. (2013). Undercoverage Rates and Undercoverage Bias in Traditional House Listing. *Sociological Methods & Research*, 0049124113500477.

ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 1.0. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Google (2015). Image Acceptance & Privacy Policies. Retrieved from https://www.google.com/maps/streetview/policies/

Groves, R., & Couper, M. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *Journal of Child Psychology and Psychiatry*, *53*(10), 1009-1017.

Pickering, K., Thomas, R., & Lynn, P. (2003). *Testing the shadow sample approach for the english house condition survey*. National Centre for Social Research.

Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. *American journal of preventive medicine*, *40*(1), 94-100.

Walsh, R., Dahlhamer, J., & Bates, N. (2013). Assessing Interviewer Observations in the NHIS. Retrieved from www.Amstat.org.

West, B. T. (2013). An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 211-225.

West, B. T., & Kreuter, F. (2013). Factors affecting the accuracy of interviewer observations evidence from the National Survey of Family Growth. *Public Opinion Quarterly*, 77(2), 522-548.

West, B. T., & Sinibaldi, J. (2013). The Quality of Paradata: A Literature Review. In Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic uses of process information*, pp.339 –360. Hoboken, New Jersey : John Wiley & Sons, Inc.