30th International Workshop on Household Survey Nonresponse

**Nonresponse in a smart living conditions survey**

Goran Ilic & Peter Lugtig (Utrecht University) and Barry Schouten (Statistics Netherlands & Utrecht University)

*Summary: Smart surveys use the storage, computing and internet functions of smart devices in collecting part or all of the survey data. One appealing option to reduce respondent burden and to improve data quality in such surveys is to replace part of the questions by sensor measurements. Such measurements can be made actively or passively, i.e. with or without explicit interaction with respondents. The use of sensors is promising from the survey questionnaire perspective when survey topics are cognitively burdensome, and/or when they require specialist knowledge and/or when questions are weak proxies of the concepts of interest. This promise is only real when sensor data can be accessed, are accurate and comparable, are not burdensome to respondents and not privacy intrusive.*

*We conducted an experiment in a living conditions survey where we replaced three topics by respondent photos. We randomized the sample across three conditions: a question condition, a sensor/camera condition and a choice condition. We discuss the nonresponse mechanisms to the three conditions as well as the quality and information of the data obtained.*

1. **Introduction**

This paper explores nonresponse to a so-called smart survey in the context of living conditions. We investigate to what extent offering sensor options affect willingness to participate and improve data quality.

A smart survey starts from the viewpoint of the device on which the survey runs. This device will often be a desktop, laptop, tablet or smartphone. To a survey the following features may be added:

1. Device intelligence: It can use the intelligence (computing, storage) of the device
2. Internal sensors: It can employ the sensors that are available in the device
3. External sensors: It can communicate through the device with other sensors that are close by
4. Public online data: It can go online and extract publically available data
5. Personal online data: It can go online and request access to existing external personal data
6. Linkage consent: It can ask consent to link external personal data already in possession of the survey institute

Smart surveys use device intelligence, the first feature, and may in addition employ one or more of the other features. Doing so, they will give new missing data patterns, in particular new forms of unit-nonresponse. In this paper, we consider the internal sensor feature: we offer respondents the opportunity to take photos of their housing conditions as an alternative to answering survey questions. These sensor measurements are integrated into a larger survey.

Sensor measurements may be added value from two perspectives: They may reduce respondent burden and/or they may reduce measurement error. Sensor measurements, consequently, seem especially interesting for three types of survey topics: 1) topics that require high (cognitive) effort, 2) topics that require specialist knowledge, and 3) topics for which questions offer weak proxy measures of the concepts of interest. Living conditions surveys concern all three topics. It may be time consuming to describe one’s setting in detail, it may requires expert knowledge to do so, and the underlying concepts of living conditions are complex constructs involving several dimensions.

Since smartphone devices contain integrated sensors (Link et al., 2014), researchers now have the opportunity to ask respondents to activate them and perform different tasks in addition to the survey (Wenz, et al., 2019). Thus, smartphone devices can be utilized to collect a variety of data, such as participants’ geolocations, physical movement or internet usage (Keusch et al., 2019; Link et al., 2014). Nowadays, visual data can also be easily and quickly collected since smartphone devices contain a camera. This feature presents an opportunity of asking participants to take a photo during a survey. Potentially, some survey questions can be substituted by the task of taking a photo (Bosch et al., 2019; Link et al., 2014). This is specifically useful for a survey aiming to investigate the objective aspects of a participant’s situation which they could find hard to report such as living or housing conditions.

Despite the potential benefits, there are four broad challenges in implementing visual data in survey methodology. The first challenge is the willingness of participants to perform a task. In the case of housing surveys, the question is how willing would respondents be to take a picture of their private space such as their home. Research so far shows that participation willingness differs over the task and participant characteristics (Jäckle, et al., 2019; Revilla et al., 2019; Wenz, et al., 2019). The second challenge is associated with the ethical aspect. This must be carefully considered when using visual data (Kelly et al., 2013). For instance, depending on their content, pictures could disclose personal information outside the main scope of the survey (Wenz, et al., 2019). In the housing survey context, a picture might contain other household members or information that might disclose respondents’ privacy such as bank card number. Thus, measures of privacy protection have to be designed and implemented. The third issue is related to technical challenges in developing an app or web survey that allows the collection of pictures. The development and testing of apps or web surveys can be time-consuming and expensive (Link et al., 2014). Moreover, other obstacles might occur such as an inability to develop apps applicable to all operation systems (Kreuter et al., 2018). Consequently, users of certain operating systems cannot be included in the research. The fourth challe first is related to the analysis of visual data. Since visual data is different than traditional survey data it requires a different analysis approach. The information of interest has to be detected, extracted and encoded from an image before the statistical analysis can be conducted. This can be done using developments in the computer vision field or manually by a human coder (Bosch et al., 2019). For a survey where the sample size is large, human coding of pictures can be time-consuming, therefore the computer vision approach might be a solution to increase a speed of visual data analysis. This paper is a first exploration into these challenges. The main research questions that we answer is 1) to what extent offering sensor measurements improves participation, 2) to what extent offering sensor measurements improves data quality.

1. **Design of the study**

Smart surveys affect both representation and measurement. In this paper, we focus on nonresponse. The introduction of sensor measurements, however, implies a clear trade-off between nonresponse and measurement. Improved data quality may, in fact, be the main motive to replace/supplement survey questions by measurements. Up to nonresponse, the errors for smart surveys coincide with those for traditional surveys. There is a target population that is represented by a sampling frame, which may be incomplete, contain ineligible units and/or contain double records. A sample is drawn at the hand of a sampling design from this frame leading to sampling error. Sample units are contacted and invited, but may not respond, Respondents then proceed to perform the survey tasks. It is here that new causes of missing data may start to occur. In Figure 1 these are displayed for the case of sensor data. Respondents need to have access to sensors, they need to be willing to perform sensor measurements and/or provide access to the sensor data, they need to execute the sensor measurements and the sensor data should be processed and transmitted. In all these steps, population units may drop out and cause representation to be selective and unbalanced across relevant population characteristics.

In our case study, we employ cameras available in mobile devices. Transmitted response, thus, consists of visual data from which meaningful features need to be extracted. Such feature extraction is typically done by machine learning models that need to be trained with the help of labelled pictures. Machine learning model predictions even in the best of cases have an accuracy below 100%. The problems in accuracy of computer vision can occur for a number of reasons (Gollapudi, 2019): 1) Problems with the camera sensor in low or bad light, 2) The same object looks different from different angles, and 3) An object in motion can look different than stationary. Here, we ignore these processing errors, but they should not be ignored.

Our study is linked to a living conditions surveys, in which sample units are asked to report on various aspects of their housing. To answer the research questions, we asked panel members from the LISS panel to participate in a short survey about their dwelling’s characteristics. The LISS panel is managed by CentERdata, a research organization based in Tilburg, the Netherlands. The panel has been recruited from samples drawn from the Dutch population register (Das & Knoef, 2019). Since its start, a number of refreshments samples have been included. The panel consists of around 8000 members in a total of 5000 households. The smart survey study was conducted in October, 2019. The collection of data started on 7th October 2019 and finished on 27th October 2019. Only households with at least one smartphone or tablet were invited. From 2700 LISS panel members invited to participate, 1759 completed the web survey. Individuals were randomly assigned to one of the following conditions:

1. Condition 1 (photos) - tasked to take pictures instead of answering three questions.
2. Condition 2 (text questions) – participants answered survey questions.
3. Condition 3 (choice) – participants had the opportunity to choose either the picture or question condition at the beginning of the survey. If participants agreed to take a picture, the routing was the same as for the condition 1.

Before the data was made available, CentERdata researchers went through the pictures and masked the content that could lead to a violation of privacy of the participants.

*Figure 1: Representation errors in smart surveys.*

Respondents

Undercoverage of sensors

Sensor owners

Willing to use sensors

Sensor participants

Sensor missing data

Complete sensor measurements

Transmission sensor data

Transmitted response

We asked participants in condition 1 and the condition 3 photo group to take three pictures. Before each task, we provided respondents with a task description containing a question if they are willing to perform the task. In case they said yes, they would be able to take a photo. If they rejected the task, we asked an open-ended follow-up question about the reasons for rejection together with the text question about the topic. In condition 2 and the condition 3 text group, we asked participants to answer text questions that measured the same concept of interest as the pictures. The three pictures respondents were asked to take were the favourite place in their house, the dwelling’s outdoor area, and the type of heating device. The latter was used to derive energy efficiency. We included the favourite place photo and question as we deemed it to be sensitive to respondents. The other two photos, outdoor area and heating device, were included as they require expert knowledge.

1. **Results**

In total, 2700 LISS panel members were invited to participate in the survey. 65.1% (n=1759) of the invited panel members decided to fill in the survey, which results in survey nonresponse of 34.9% (n=941). In condition 3, 56.7% (n=320) of the participants chose to take a picture instead of answering survey questions. The complete distribution of the participants per condition is presented in Table 1. Between the photo and choice options, response rates were very similar, but in the question option response rates were higher. These must be attributed to break-off during the survey. In the question/tekst option break-off rates were negligible.

*Table 1: Participation rates. Condition 1 = photo, condition 2 = questions, condition 3 = choice between photos and questions*

|  | n | % |
| --- | --- | --- |
| Invited | 2700 | 100 |
| Response |  |  |
| Condition 1 (Photo) | 540 | 60.0 |
| Condition 2 (Question) | 655 | 72.8 |
| Condition 3 (Choice) | 564 | 62.7 |
| Total | 1759 | 65.1 |
| Condition 3 |  |  |
| Preferred photo | 320 | 56.7 |
| Preferred question | 244 | 43.3 |

3.1 Willingness to participate

Our first research question is about willingness to participate. Table 2 shows the actual rates at which respondents consented to perform the tasks. This conforms tot he final step, transmission of responses, in figure 1. In the photo condition, respondents had the option to skip the question. We make three observations. The first is that overall item response rates are relatively low under the forced photo condition, ranging from 26.5% for the outdoor area to almost 40% for the favourite place. The second is that for heating device also the question condition showed relatively high item nonresponse; only around two third of the respondents provided an answer. The third, and most positive observation, is that under the volunteered photo condition item response rates are much higher.

|  | **Favourite place** | | **Outdoor area** | | **Heating device** | |
| --- | --- | --- | --- | --- | --- | --- |
|  | r  (eligible) | % | r  (eligible) | % | r  (eligible) | % |
| **Condition 1 (Photos)** | 212  (540) | 39.3 | 138  (521) | 26.5 | 155  (444) | 34.9 |
| **Condition 2 (Text question)** | 641  (655) | 97.9 | 626  (632) | 99.1 | 367  (549) | 66.9 |
| **Condition 3 (Choice)** |  |  |  |  |  |  |
| **Photo group** | 248  (320) | 77.5 | 154  (306) | 50.3 | 166  (276) | 60.2 |
| **Text group** | 233  (244) | 95.5 | 228  (231) | 98.7 | 117  (203) | 57.6 |
| **Total condition 3\*** | 248  (564) | 44 | 154  (537) | 28.7 | 166  (479) | 34.7 |

Table 2: Willingness to take a picture or answer text questions per condition

So what were the causes for these relatively low response rates? Table 3 attempts to break down nonresponse into various causes. Since LISS panel sample units all had a smartphone, sensor undercoverage was not an issue. When rejecting to take a picture, respondents were asked for the reason. This was done only for the favourite place and heating device photos. From table 3, we can conclude that for favourite place the main reason was privacy concerns, while for heating device the main reason was that they could not reach the heating device (i.e. were unable to take the picture). In both cases around 10% indicated they had technical issues with their smartphone camera, which can in the majority of cases be resolved in future surveys by improving the questionnaire IT, which had only been tested on relatively new smartphone models.

|  | **Favourite place** | | **Heating device** | |
| --- | --- | --- | --- | --- |
|  | n | (%) | n | (%) |
| **Reasoning** |  |  |  |  |
| **Unwilling** |  |  |  |  |
| Privacy concerns | 202 | 56.4 | 58 | 15.1 |
| Don’t want or feel like it/unnecessary | 26 | 7.5 | 14 | 3.7 |
| **Unable or incapable** |  |  |  |  |
| Technical problems with survey | 40 | 11.2 | 39 | 10.2 |
| Respondent’s related technical issues | 13 | 3.6 | 8 | 2.1 |
| Do not have favourite place\*\* | 20 | 5.6 |  |  |
| Unreachable\* |  |  | 196 | 51 |
| Not sure how home is heated\* |  |  | 1 | 0.3 |
| Not at home | 13 | 3.6 | 11 | 2.9 |
| Other reasons | 17 | 4.8 | 15 | 3.9 |
| Did not provide a reason | 26 | 7.3 | 42 | 10.9 |
| Total | 358 | 100 | 384 | 100 |

Table 3: Coded reasons for rejection of taking a picture of favourite place and heating device

From the analyses, we conclude that actual response rates are relatively low, but depend strongly on the content of the photos. The topic that we considered to be sensitive, favourite place in the house, indeed led to privacy concerns. The topic that we considered to require expert knowledge led to practical issues. We have not yet fully analyzed the answers under the question condition, but we suspect that these may be subject to measurement errors.

* 1. Data quality of respondent photos

The second research question is the impact on data quality. We consider the heating device photo as example.

All photos were manually coded and we also had a separate meeting with a heating device expert in order to discuss the extraction of device model and type. Some heating device brands have no text display on their exterior, so that text recognition is impossible. For such heating devices, model and type prediction must be entirely based on image recognition. Given that heating devices are a very specific set of objects, there are currently no pre-trained machine learning models available. However, in the Netherlands, and most likely also in other countries, there is a large database of models and types including technical information. This database is maintained by the government in order to keep track of energy efficiency of available devices. This is used for environmental tax deduction that the government provides to households if they use efficient devices. Although, we did not perform a full text and image recognition analysis, we expect that machine learning models may reach high accuracy levels if photos are taken properly and under the right light conditions.

Table 4: Analysis of the content in the pictures of heating device

|  | In the line with a task | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Yes | | No | | Total | |
|  | n | % | n | % | n | % |
| Picture of heating device |  |  |  |  |  |  |
| Number of observed pictures | 281 | 87.5 | 40 | 12.5 | 321 | 100 |
| Privacy control | 12 | 4.3 | 0 | 0 | 12 | 4.3 |
| Information extracted from the pictures\* |  |  |  |  |  |  |
| Only brand of a device | 101 | 31.5 | 28 | 8.7 | 129 | 40.2 |
| Model of a device | 124 | 38.6 | 0 | 0 | 124 | 38.6 |
| More detailed model specification | 18 | 5.6 | 0 | 0 | 18 | 5.6 |
| No brand or model information | 38 | 11.8 | 12 | 3.7 | 50 | 15.5 |
| Energy efficiency label (GASKEUR sticker) | 7 | 2.2 | 0 | 0 | 7 | 2.2 |
| Pictures contain |  |  |  |  |  |  |
| Small part of a device without model information | 0 | 0 | 33 | 10.3 | 33 | 10.3 |
| Half or more of device including/excluding model specification | 212 | 66 | 0 | 0 | 212 | 66.0 |
| Only part of a device with device model specification | 59 | 18.4 | 0 | 0 | 59 | 18.4 |
| Not device but includes information about device model (stickers) | 8 | 2.5 | 0 | 0 | 8 | 2.5 |
| Not informative/ not device of interest | 0 | 0 | 7 | 2.2 | 7 | 2.2 |
| Others (GASKEUR, privacy censored picture) | 2 | 0.6 | 0 | 0 | 2 | 0.6 |
| Total | 87.5 | 40 | 12.5 | 321 | 100 | 100 |

Table 5: Analysis of the information obtained in text responses

|  | n | % |
| --- | --- | --- |
| Specification of a heating device |  |  |
| Brand |  |  |
| Only brand including/excluding year | 202 | 41,7 |
| Brand and kind of a device (e.g. high efficiency, combination device) | 41 | 8,5 |
| Brand and model type |  |  |
| Brand, model including/excluding year | 92 | 19,0 |
| Brand, detailed specification of model including/excluding year | 50 | 10,3 |
| Only year | 49 | 10,1 |
| Kind of a device (e.g. high efficiency) | 21 | 4,3 |
| Do not know | 14 | 2,9 |
| Uninformative answers (e.g. type thermostat) | 15 | 3,1 |
| Total | 484 | 100 |

Table 4 gives a summary of the quality of respondent photos. Around 88% of the photos was in line with the instructions, i.e. a photo of the frontside of the heating device including model and type, if present, at relatively close distance and in either daylight or sufficient artificial light. Respondents had to include the whole device frontside in case model and type were not displayed. The 12% photos that did not qualify mostly displayed only a part of the device without information on model/type. Out of the 281 photos, 12 had some form of privacy control where parts of the photo were blurred/masked by LISS panel staff. Even for some of the photos with only part of the device depicted in the photo, the expert suspects they can be classified correctly, because the shape and colour are still sufficiently distinctive. This mean an accuracy of more than 90% should be possible.

Table 5 gives an evaluation of the answers that respondents gave under the question/tekst condition. Out of the 484 pictures only 142 answers specified both the brand and model, i.e. around 30%. For these answers, it would be possible to derive the energy efficiency of heating devices.

We conclude that data quality of photos is superior tot hat of the question answers, even though part of the photos did not conform to instructions.

1. **Discussion**

This paper explores missing data patterns in a smart survey on living conditions that employed sensor measurements via mobile devices cameras. Smart surveys introduce a trade-off between nonresponse and measurement that turned out to be very prominent in the results of the living conditions case study.

Willingness to partcipate and to consent to taking photos was relatively low. In the forced photo condition around 20% of the sample provided photos, whereas in the question condition around 60% of the sample provided answers to the corresponding questions. Offering a choice between a photo and a question lifts the actual item response rates. A small minority preferred taking pictures when offered a choice and the majority of thema also did take the pictures. We conclude tentatively that offering a choice is to be preferred. We need yet to evaluate differences in background characteristics of the response under the various conditions.

Data quality was clearly higher for the photo conditions where the majority of heating devices could be classified. For the open text answers only around 30% had sufficient information. We still need to evaluate the other two pictures of the favourite place in the house and the outdoor area.

Summarizing, a smart living conditions survey demands for a choice between higher response rates and higher data quality. It must be stated, however, that the data collection strategy and respondent instructions may be improved given the results of our study. We may lift response rates by improving technical performance of the questionnaire and by avoiding sensitive topics for photos. We may also lift data quality for questions by giving different instructions/explanations.

More generally, sensor measurements are promising when the measurements themselves are not burdensome to respondents. In the case of taking photos, especially of hard to reach places in the house, may still be time-consuming and, thus, burdensome. Other sensor measurements may have more favourable respondent burden features.

We have the following questions for discussion:

1. Do you have suggestions for improving the data collection strategy?
2. Do you have an explanation for why offering choice between questions and measurements is more preferable?
3. Would an adaptive survey design where we view sensor measurements as a separate ‘mode’be an option? If so, how would you envision such a design?
4. Would it be useful to promis/offer individual feedback to respondents about the outcomes of the machine learning models, e.g. “the energy efficiency of your heating device is …”?
5. Suggestions for follow-up studies?