

Surveys with sensors. The future of data collection?

Ole Mussmann (CBS) and Barry Schouten (CBS and Utrecht University)

General population official surveys are subject to nonresponse and to measurement error. This has been the case since the early days of surveys and, in time, methodologists have explored and implemented several measures to effectively battle these survey errors. Official surveys have relatively high response rates, despite the response burden, and questionnaire design often involves various rounds of testing and revision. Nonetheless, response rates have been decreasing universally and measurement of certain survey topics remains hard. The decline in response rates has driven survey designers to use multiple modes and to adapt the data collection strategy to sample units. Within the online survey mode, the diversity of devices has grown; especially, smartphones have become omnipresent and are used intensively and frequently. Survey topics that require in-depth knowledge (like travel locations or health), that are cognitively burdensome (like expenditure or time use diaries) and/or that are simply hard to translate to questions (like physical condition or happiness) do not lend themselves for questionnaires. The emergence of sensors in mobile devices and wearables brought promising alternatives to questionnaires. All in all, there seem to be new chances and new possibilities. But will the general population agree, as the devices also bring new privacy issues?

Topics for sensor surveys

In 2016, Statistics Netherlands and Utrecht University established a joint research program on ICT innovation in primary data collection under the name WIN (short for Waarneem-Innovatie Netwerk).

Simultaneously, Statistics Netherlands organized a number of internal brainstorms to identify surveys at risk and/or surveys containing topics that satisfy one or more of the problematic measurement properties. From these the following topics were identified: travel/mobility, ICT use, budget expenditure, time use, health, living/housing conditions and working conditions. These were later pooled to four WIN subprojects: 1) travel, 2) time and media use, 3) budget expenditure and 4) health and life style.

For each of the four subprojects, research is on-going but in different stages. Statistics Netherlands has developed an open-source cross-platform travel app (the CBS verplaatsingen app) that can be used for different surveys. It uses GPS-GSM-Wi-Fi and motion sensors to detect stops and travel, which can be used in different surveys. It is developed using Xamarin in order to be able to link to the Blaise app. In November a large-scale field test is planned exploring a variety of incentive strategies and stop detection rules. The app performs active sensor data, i.e. asks respondents for checks and additional information on purposes of the travels. The same app serves as a starting point for an online event-based time use diary that inserts geo-location data. This app will most likely be further developed within a joint research project with the Dutch Social and Cultural Planning Office and in collaboration with a Eurostat taskforce on the HETUS (Harmonized European Time Use Survey). For budget expenditure another app is explored that can scan receipts, can assist respondents using geo-tagging shopping and other spending area locations that they have visited and asks for consent to link to big data sources. This project is submitted as a proposal within a Eurostat call. Health and life style are the most diverse of the four topics and most likely demand for supplementary wearables such as activity trackers or smartwatches. Here, a first case study has been initiated to explore the fitness of a combination of smartphones and wearables for measuring type of activity and calorie usage. This research is done with several other institutes and is oriented at determining accuracy, cost and user-friendliness trade-offs of different options.

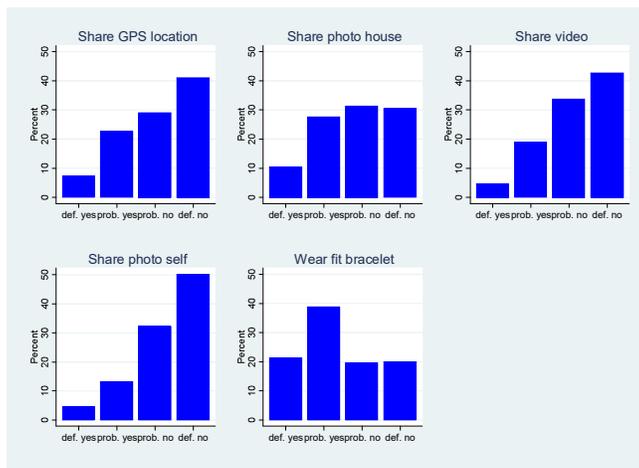
These explorations and research projects are not at all unique, not even at National Statistical Offices. Sessions are frequently included in the program of large survey methodology conferences and in 2019 a dedicated workshop, MASS, will take place at Mannheim university. However, to date, experience about the implementation in official surveys is still very thin. Also, apps are rarely developed open source, having in mind that multiple users may work on the code and benefit.

Consent to provide sensordata

A obvious potential obstacle is privacy. What about person's willingness to provide mobile device sensordata? In order to explore, the consent to sensordata requests, two studies have been conducted within WIN and a third is currently being prepared.

The first study was fielded in the LISS-panel of CentERdata in 2017 and asked for consent to five sensor measurements. Consent rates (taken from Struminskaya et al 2018a) are shown in figure 1.

Figure 1: Consent rates to five sensordata measurements: sharing location, making a photo of one's house, making a video of surroundings of one's house, making a selfie photo and wearing a fit bracelet.



The second study was fielded at Statistics Netherlands with the help of the Blaise team and combined consent questions with browser-initiated sensor measurements in a Blaise 5 questionnaire. Whenever a respondent gave consent he/she was asked to also provide the sensor data. The theme of the survey was general (Struminskaya et al 2018b). The sample consisted of former respondents to a range of surveys. The consent rate to location data was 67%, whereas the consent rates to taking photos and short movies was much lower and hovered around 15%.

A possible reason for the relatively low consent rates to the camera measurements is the artificial nature of the general survey. For this reason, a substantive survey on housing conditions will be fielded in which sensor measurements are more logical and remove much of the respondent effort.

Surveys with sensors?

Summarizing, a new type of survey seems to emerge, one that combines questions with measurements. It is yet unclear how far respondents are willing to go and what are effective data collection strategies, but it seems inevitable that they will be implemented for at least part of the traditional survey topics.

What more can be expected? Experience sampling where respondents are asked to provide their emotions and/or satisfaction depending on decision rules employing sensor data. New incentives to respondents in the form of personal feedback. Multi-person apps that communicate whenever persons are close to each other. Sensors that react to external stimuli such as NFC and Bluetooth. Sensor measurements that help linkage to big data, such as supermarket scanner or loyalty card data. These are just a few examples; varying in utility and practical feasibility, but they all hold in common that they are technically feasible.