The potential of smart surveys. Four case studies in the context of the ESS

Keywords: Trusted smart statistics, Sensor data, App-assisted surveys

1. INTRODUCTION

We discuss the potential of so-called smart surveys. We see three push factors that drive us away from surveys as the dominant mode for data collection: reduction of cognitive burden, better proxies of the concepts of interest, and reduction of respondent inability to provide specialist information. We also see three pull factors of sensor and digital data: the wider availability of such data, the ability for respondents to control the content and access to such data, and coherence of data being collected.

Smart devices offer attractive options to collect more traditional types of data (e.g. survey questions, along with new forms of data. A smart device offers the following features for collecting, linking or processing data

1. Device intelligence: It can use the intelligence (computing, storage) of the device, e.g. it can apply pre-trained machine learning models for image recognition;
2. Internal sensors: It can employ the sensors that are available in the device, e.g. the location sensors, camera or motion sensors;
3. External sensors: It can communicate through the device with sensor systems close by, e.g. a smart watch or an indoor climate system;
4. Public online data: It can go online and extract publicly available data, e.g. open streetmaps data;
5. Personal online data: It can go online and request access to existing external personal data, e.g. bank transaction data or shop loyalty card data;
6. Linkage consent: It can ask consent to link external personal data already in possession of the survey institute, e.g. shop scanner data or public transport data.

We term surveys that use any of these features smart surveys, because they use the extra functions that smart devices offer. The last two features amount to data donation; the data already exist but the respondent acts as intermediary to actually combine the data with the survey data. The hybrid data collections that emerge from smart surveys mix different types of measurement tools and mix new and existing data, e.g. Link et al (2014) and Japec et al (2015).

The mere fact that data can be combined is not a sufficient reason to also do so, of course. These data collections imply a need for new logistics and new IT infrastructure. As a consequence, there must be a strong business case to support such an endeavor. ESSnet Smart Surveys explores the potential of smart surveys in the context of the Europe an Statistical System (ESS).As part of that larger project, we explore the business case at the hand of four topical case studies in the ESS context: household expenditure time use, physical activity and living conditions. We believe that for these topics, the business case is positive.

Crucial questions are what the impact is on data errors, what methodology solutions can be applied and how this translates to architecture and logistics for smart surveys. In order to answer such questions, it is imperative that the total survey error framework is revisited with the smart features in mind. Smart surveys bring new sources of error, both in representation and in measurement. As a consequence, a new methodology is needed to reduce such errors. By merging different types of data, smart surveys may combine the best of multiple worlds, e.g. Callegaro and Yang (2018). One goal of data fusion methods is to combine estimates from both sources in such a way that total errors are reduced or eliminated (Braaksma & Zeelenberg, 2015; Tzavidis et al 2018; Zhang 2012). A same approach is needed for smart surveys.

2. TOTAL DATA ERROR FOR SMART SURVEYS

The Total Survey Error (TSE) Framework is popular for evaluating candidate designs of a survey and for structuring the impact of design features on accuracy of statistics (Weisberg 2009). Several publications have outlined how the TSE can be extended to alternative data sources (Groves, 2016; Japec et al 2015; West et al 2017). TSE has two main components: representation and measurement. Smart surveys using sensors affect both components.

For representation, the errors for smart surveys coincide with those for traditional surveys up to response. 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 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. Respondents need to have access to sensors or external data, they need to be willing to perform sensor measurements and/or provide access to sensor data and other forms of data, they need to execute the sensor measurements or link data, and the resulting 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. Smart surveys thus introduce new errors into a survey,.

For measurement, changes are much more drastic as they involve the very concepts themselves. It is in measurement where sensors and new forms of data add value to surveys, and where the ultimate goal is to reduce TDE. Typically, the measurement process starts by operationalizing how an abstract construct can be operationalized using empirical measurements. For example, a survey may try to measure a number of air quality parameters over a specified time period in specific areas of dwellings. These are proxies for the construct indoor climate, and, more generally, the housing conditions. Asking questions about air quality may be difficult or even impossible for respondents. Sensor measurements can here be used to improve the validity of measurement by employing sensors that directly measure aspects of air quality. However, the respondent needs to be able and motivated to operate the sensors such that they measure these parameters, possibly following instructions given within the survey invitation. New types of measurement error may here be introduced in what we call operating error: respondents may incorrectly initialize measurements or position devices in the wrong spots. When the sensor is put into operation, then it may produce measurement errors itself. Depending on sensor quality and age, sensors may produce both random measurement error and systematic measurement error. In the example, the concentrations of various pollutants or moist is for example subject to imprecision. When measured by two copies of the same sensor at the same time, the two sensors will give slightly different readings. Furthermore, sensors need to be calibrated to known absolute levels, but may deviate from those in time. Periodic recalibration is needed and without such recalibration show a time-dependent systematic error. Even so, sensors may simply not be able to detect certain minimum levels or may not be able to differentiate between certain events. In the example, air quality sensors may not be sensitive enough to differentiate between particulate matter particles of different sizes. Finally, sensor measurements need to be processed, e.g. different sensors may be combined, and errors may occur in doing so.

Ideally, the different error components are estimated and compared to similar errors in survey questionnaires. In general, such an exercise requires advanced experimental designs and may be costly and time consuming. We propose to perform a preliminary evaluation into the potential utility of smart surveys using sensors using a reduced set of criteria. When these criteria are met, then there may be a positive business case and in-depth evaluations may be initiated. We consider three viewpoints: the survey questionnaire, the sensors being used and the respondent who is participating in the study. Positive scores on the survey questionnaire criteria imply that survey questions may be at risk of relatively large survey errors, i.e. there is an incentive to replace or supplement survey questions. Positive scores on the sensor criteria imply that sensor errors may be acceptable, i.e. sensors may offer an alternative way of measuring the construct of interest. Positive scores on the respondent criteria mean that respondents may be willing and able to perform the task, i.e. sensors are also an option for respondents. See Schouten and Mussmann (2019) for a detailed discussion and motivation of these criteria. In the paper, we will align them with the four case studies.

3. TWO EXAMPLES OF CASE STUDIES

We elaborate two of the four case studies where survey and sensor data may be combined. In the presentation, we will elaborate all case studies.

Household expenditure surveys typically stretch out over a number of weeks in which respondents need to list all or a subset of expenditures. Surveys are typically burdensome due to the length and detail of required information. Surveys also require information that respondents do not readily have available, such as quantities and prices of products they bought and/or calorie intake associated with the consumption. As an archetype example of such surveys, we take the Household Budget Survey (HBS), which is a mandatory survey in the European Statistical System (ESS) and conducted in relatively similar designs across many countries. A standard design is a mix of a recruitment survey plus paper and/or online diaries. The HBS deals with all kinds of purchases and expenditures, both small and large, and both frequent and infrequent. Some of these purchases are done on site, such as shops, restaurants and cinemas, others online such as booking websites or retail website, and yet others are periodic payments such as insurances, mortgages and energy.

What kind of sensor data are promising? Time-location sensor data may be employed to assist respondents in memorizing or recalling locations where products or services have been purchased. Some purchases are done online and part of those may be done through a mobile device. The use of certain online shopping apps may be detected and mentioned to assist the respondent. In all of these cases, direct access to the type, amount and cost of products and services themselves will, generally, not be possible due to privacy restrictions on the apps. Another option is to use the camera to scan shopping receipts, to scan barcodes on products and/or to take pictures of food that is consumed. Receipt scanning is mostly useful for purchases that involve many products/services simultaneously and that are burdensome to insert into a diary/questionnaire. Barcode scanning does not provide product prices, but can be linked to GTIN product codes that themselves can be linked to product descriptions, including ingredients. Figure 1 shows screenshots of an HBS app in which expenditures can be included through manual entry and through scans of receipts.

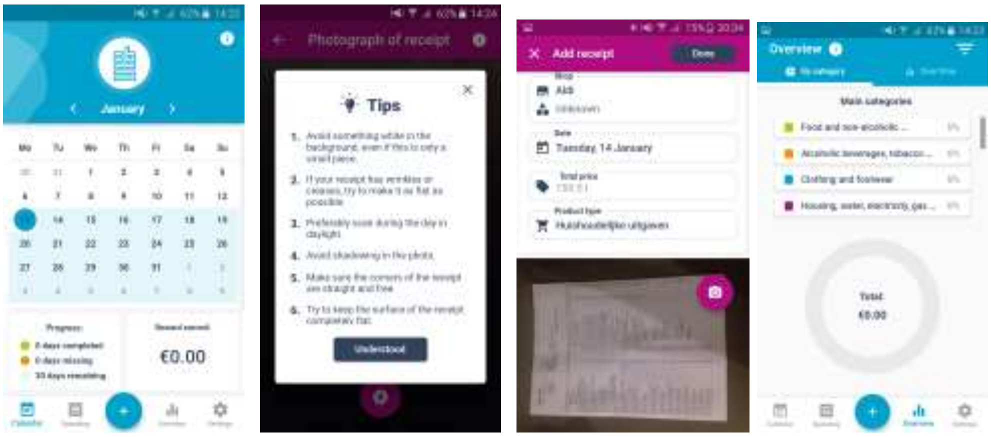

Figure 1: Screenshots of the Household Budget Survey app.

Health is perhaps one of the broadest survey themes, ranging from living conditions, working conditions, hospitalization, medicine use to health determinants such as life style and physical activity. To limit the scope, we focus on physical activity, which is measured in many surveys through a series of questions. Although the interest is in average physical activity over a longer time period, typically, surveys tend to measure activity in cross-sectional samples at one point in time. An example set in the European statistical System is the European Health Interview Survey (EHIS). Respondents need to provide estimates of type and duration of activities, and, sometimes, also intensity. Physical activity questionnaire modules are cognitively burdensome, as they refer to averages durations over a long time period. They are also non-central, as respondents do not know exact distinctions between different types of activity and their durations and intensities. Whereas the real interest is in estimating the mean and variance of individual calorie use over a certain period, a survey delivers a relatively weak proxy measure. As the amount of calories used cannot be asked, they are usually estimated by multiplying average durations by standard metabolic rate values. These standard values suppress individual variation. Physical activity and, more generally, health, thus, scores on all three survey questionnaire criteria.

What sensors can be used? A multitude of wearables are aiming at health, fitness, life style indicators. Activity trackers, smart scales and smart watches may be employed for at least a specified duration in order to derive physical conditions, and, possibly, also mental conditions. See for example Huberty et al (2015) and Kerr et al (2017). Wearables employ motion sensors and some allow for measurements of heart rate. Dedicated wearables exist that measure heart rate and respiration. Often the wearables are combined with mobile devices applications. To strengthen predictions of activity, wearables may be combined with location measurements. Figure 2 gives a small collage of wearables and sensors for the measurement of activity.

Figure 2: Examples of wearables and other sensors designed for measuring physical activity. From left to right: respiratory gas analysis, inertial measurement unit (IMU), smart sensor shirt, fitness bracelet and upper leg motion sensor.