Mixing Online Panel Data Collection with Innovative Methods

Annette Scherpenzeel
Munich Center for the Economics of Aging (MEA)

1 Introduction[1]

Mixed mode studies in survey research have been around since a long time. Whenever a new mode of data collection emerges, it has been combined with existing traditional modes of data collection, and compared in terms of resulting response rates, response bias, data quality and measurement error (Couper 2011, for general overviews of traditional mixed mode studies see de Leeuw 2005 and De Leeuw et al. 2008). The first studies focused on comparisons of the traditional modes paper-and-pencil and face-to-face interviews. These modes were, in turn evaluated again after telephone interviewing evolved. The next stage was a series of studies comparing the effect of computer assisted forms of telephone interviewing (CATI) and face-to-face interviewing (CAPI). Since the emergence of web interviewing (CAWI), most mixed mode studies have compared this mode with the foregoing modes (for overviews of such studies, see Lozar et al. 2008; de Leeuw and Hox 2011).

The most recent comparative mode studies focus on the comparison of web interviewing on computers versus mobile devices, such as smartphones and tablets (see, for example, Antoun 2015; de Bruyne 2015; Lugtig and Toepoel 2016). Furthermore, technological development continues to offer ways of

---

1 We thank CentERdata for enabling the studies reported in this paper, in particular Josette Janssen, Salima Douhou, Maarten Streefkerk and Joris Mulder who organized the complex field work for all studies, CentERdata's director Marcel Das, and Maurice Martens, Iggy van der Wielen, Bart van Nieuwburg and Lennard Kuyten who developed the innovative apps and software.

S. Eifler und F. Faulbaum (Hrsg.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*, Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, DOI 10.1007/978-3-658-15834-7_2

data collection that go beyond asking survey questions, for example by continuously tracking people's behavior. It is this development, the mixing of interviewing with completely new ways of measuring behavior and feelings, which is the subject of the present paper. We will illustrate the possibilities of new mode mixtures with studies that have been conducted in the LISS panel, a probability based online panel which was especially developed for scientific research and for experimentation. The purpose of the paper is to not only demonstrate the interesting and truly innovative ways of data collection which have become possible, but to also evaluate the problems encountered and the resulting response rates and data quality. We will discuss the different ways in which the new techniques were combined with the normal panel interviews using the framework of mixed-mode designs given by De Leeuw and Hox (2011). They distinguish four main groups of mixed-mode design, which can be applied to traditional mixed-mode studies as well as to the most recent multi-mode web surveys:

1. Contacting by Different Modes: to reduce nonresponse, e.g. Advance letter, reminders, screening
2. Another Mode for Specific Questions in the Questionnaire: e.g. Self-administered part for sensitive questions
3. Different Modes for Different Respondents: same questions in different modes, to reduce nonresponse or noncoverage. E.g.: for non-internet group; for nonrespondents; respondents in different countries
4. Alternating Modes in a Longitudinal Design. E.g. first wave face-to-face, second wave Internet

We present the applications of new techniques and innovative modes within the framework of this categorization in order to show the purposes which such techniques can fulfill in survey data collection.

2 Mixed modes in probability based online panels

In the past decade several online panels for scientific purposes have been set up, based on true probability samples. In the United States, Knowledge Networks and RAND have for example built such panels. In Europe, scientific online panels have been established in different countries, such as the Longitudinal Internet Studies for the Social Sciences (LISS) Panel in the Netherlands, the German Internet Panel (GIP), the Longitudinal Study by Internet for the Social Sciences (ELIPSS) Panel in France, and the GESIS Panel in Germany (for an overview of all four European panels see Blom et al. 2015).

As Blom et al. (2015) describe, these panels all use traditional, offline modes to contact and recruit the sample members for the online panels. They thus fit into the first category of mixed mode designs distinguished by de Leeuw and Hox (2011): Contacting by Different Modes. The mixed mode recruitment is an essential part of the design of such panels, since it is the only way to cover the complete probability sample of the general population which they are based on. Hence, the panel members are recruited using face-to-face contact, telephone contact or a combination of these two modes before they start participating in the online interviews. Next, while some of these panels provide internet access and computers or tablet to those sample units who do not have their own internet access, some other panels include these sample members by offering them a different mode of interviewing, for example mail questionnaires. The GESIS panel is an example of this mixed design, using online and mail questionnaires for different subsamples in the panel, and hence an illustration of the third category of mixed-mode designs described above: Different Modes for Different Respondents.

The LISS panel is often regarded as the first of the scientific probability based online panels in Europe. However, it followed a design which had been in use for a longer time, first at the Telepanel in Amsterdam during the early nineties (Saris 1998) and from 1997 onwards at the CentERpanel in Tilburg (Hoogendoorn and Daalmans 2009). Building on these experiences and the existing infrastructure of the CentERpanel, the LISS panel was able to progress quickly and take the forefront in integrating new modes of datacollection into the online panel. Within one year after its start, in 2008, the first experiments with new data collection techniques were designed. These experiments used self-tests to collect biomarkers in the panel (Avendano et al. 2011). Measures of cholesterol level, daily cortisol pattern and waist circumference were obtained, which were all three considered to be strong predictors of health but relatively noninvasive to collect (Avendano et al. 2011). These experiments fit into the second type of mixed-mode study distinguished by de Leeuw and Hox (2011): Another Mode for Specific Questions, since specific health measures were collected which could not be obtained in the regular online questionnaires. However, as Avendano, Scherpenzeel and Mackenbach (2011) report, the biomarker measures were a considerable burden for the respondents and the participation rates were low (15% of those invited to participate actually performed the tests in the cholesterol and cortisol experiments, 26%

in the waist circumference experiment). In part, the large response loss was due to the written consent[2] that had to be obtained from the panel members before they could participate in these experiments.

For these reasons, the self-test biomarker measures were considered to be unsuitable to be used in mixed mode designs in the online panel and were discontinued. In the next years, experiments were developed using more user-friendly and technologically innovative methods. We will discuss the following four studies using this type of innovative modes in the LISS panel:

1. The weighing study: A continuous measurement of weight and fat percentage over time using Internet weighing scales
2. The smartphone Time Use study: The collection of time use data, experience sampling data, and data about the use of social media, by means of a smartphone app
3. The Mobile Mobility study: The collection of mobility data by means of a smartphone application using GPS data
4. The Actigraphy study: The measurement of the physical activity of respondents using accelerometers.

The aim of the paper is to give an overview of the methodological aspects of the four studies: the design, the feasibility of implementing the methods in an online panel, the willingness of respondent to participate and the difficulties encountered. We will give a few summarized examples of substantive results as well, to illustrate the value of using new techniques in an online panel. However, since the substantive results are not the subject of this paper, we will refer to other publications for further descriptions of such results.

3 Four innovative studies in the LISS panel: General design

The LISS panel was founded in 2007 and consists of almost 8000 individuals who complete online questionnaires every month. Households that could not otherwise participate are provided with a computer and Internet connection. A detailed description of the LISS panel design, recruitment and response can be found in Scherpenzeel and Das (2011). The panel is operated by CentER-

---

2 The experiments were subject to approval by the Medical Ethical Commission (METC) of Erasmus MC Rotterdam, the Netherlands. This METC ruled that a written consent was required.

data, Tilburg University, and is the central part of a larger project, titled the Advanced Multidisciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS). The MESS project and its core part the LISS panel were funded by the Netherlands Organisation for Scientific Research.

For all four innovative modes studies, subsamples of the LISS panel were used. In general, an online invitation to participate in a special study was first send to all LISS panel members, explaining that only a limited number of devices was available and we would probably have to make a selection from the panel members that were willing to participate. This "scarcity" principle was assumed to increase the willingness to participate, following Groves, Cialdini and Couper (1992). The invitation was programmed into the online panel member page, as a regular questionnaire. Links to instruction videos, specially developed for the LISS panel studies, were given in the invitation to show what the participation involved and to take away respondent's fear of not being able to handle the new technology[3]. In addition, we developed refusal conversion screens which were presented following a (closed) question asking for the reason of the refusal. The refusal conversion argument shown on the screen was tailored to the refusal reason the respondent had selected. For example, when a respondent selected as refusal reason for the weighing study: "because I don't want to measure my weight", the following text was shown:

(Translated from original Dutch version)

> *"You indicate that you don't want to participate because you don't want to measure your weight or you don't want to weigh yourself frequently.*
>
> *It would be a pity if only people would participate who are already frequently weighing themselves. That would give a biased impression of the weight and health of the Dutch population. It is exactly your contribution that would really help the researchers of the Tilburg University. In addition, you will receive 3.50 euros for each month that you participate. For that, you only have to stand on the weighing scale for a moment, even with your eyes closed if you wish."*

3 A pilot study preceding the weighing study had shown that the most common reason for refusal was the fear of not being able to install the scale and connect it to the Internet.

Several concepts are applied in this argument which, according to Groves, Cialdini and Couper (1992) should increase the response rate: The text emphasizes the scientific importance, mentions an authority (Tilburg University), explicitly uses the word "helping" and indicates what people get in return (reciprocity). The refusal reasons which respondents could select were based on the most frequently given answers to an open-ended question in pilot studies. If too many panel members were willing to participate, a random subsample was drawn from all potential participants. Those who were not selected received a thank you email with explanations about the limited number of devices available and the high number of willing participants.

In all four studies, intensive efforts were made to stimulate participation, by active helpdesk calls before the fieldwork started, during and after the fieldwork. In addition, instruction videos were available for panel members and feedback of results was provided on the panel member's personal LISS web page. The collection of the special data by the devices was combined with specific questionnaires, for example a monthly questionnaire on activity and health in the weighing scale study and a day reconstruction questionnaire in the accelerometer experiment.

We will now give a short description of the objective and specific design characteristics of each of the four studies, before we turn to the general overview of response rates, difficulties and data quality outcomes.

4 Weighing study

Background and Objective

Social science research on health and health behaviors often uses BMI based on self-reports of weight and height as an indicator of health risks. However, a large literature documents biases in self-reported weight. In their review article Connor Gorber et al. (2007) report that studies comparing self-reported and objectively measured weight usually find under-reporting for weight and BMI, with a great deal of individual variability. In addition, a BMI based on self-reports cannot distinguish fat from fat-free mass such as muscle and bone. Fat is a stronger predictor of morbidity and health (in particular cardiovascular diseases and diabetes) than is total body mass. Another major weakness of weight measurements in social science surveys is its low frequency, typically once a year in panel studies, or one single time in cross-sectional studies. This may mask substantial weight fluctuations over time and relationships with health behavior or ageing. Technological advances can help to obtain high frequency objective measures of weight and body fat. The objective of the LISS panel weighing study was to use such measures to validate self-reports of weight as well as to follow weight and body fat variability over time in an accurate way to improve our understanding of the relationships between weight, health, and health behaviors.

Sample and study design

The LISS weighing study was a longitudinal study which lasted 3 years, following the same panel respondents over time. A first sample of 950 households in the LISS panel started to participate in it in 2010, a refreshment sample of 300 new households was added after 2 years. We provided these households with a device that measures body weight and fat percentage (based on bioelectrical impedance), and that wirelessly sends this information to the LISS database. Since the weight measurement was embedded in the LISS panel, the data are linked to a wealth of socioeconomic variables, including health related behaviors like smoking, drinking, sporting and the use of health care.

Device

The scale connected to an internet gateway via a radio signal. Both devices did not need to be in the same room. If a panel member stepped onto the scale, the scale measured weight and bioelectrical impedance, subsequently sending this information to the panel's central server via the gateway. Respondents were requested to step on the scale bare-footed and always at the same time of the day, wearing similar clothes. The scale used electrodes on the scale to send a low voltage current through the body. With this system the total impedance of the body is calculated and provides an estimate of the body's water volume. This total body water volume is then used to estimate fat-free body mass and, by difference with body weight, body fat. A more detailed description of this estimation and the use of the scale is given by Kooreman and Scherpenzeel (2014).

5 Smartphone Time Use Study

Background and objective

Time Use Research (TUR) is usually carried out using questionnaires and diaries. Respondents complete, for example at the end of the day, all their activities of one day, spread over fixed time-slots. With current technology, such as smartphones and "apps", TUR can be set up in a completely different way. Respondents having a smartphone with them can enter their activities several times during a day. In addition, with smartphones much additional data can be collected, such as the location of the respondent at the time of the activity or photos and videos of the activity performed. The smartphone Time Use study in the LISS panel was carried out in cooperation with The Netherlands Institute for Social Research (SCP), which carries out the regular TUR in the Netherlands within the framework of the Harmonized European Time Use Survey (HETUS, Eurostat 2009).

Sample and study design

The major disadvantage of smartphone data collection is the confounding of mode effects and sample selection bias, since only respondents having a smartphone with Internet access can participate. In the LISS panel, we therefore wanted to explore the possibility to lend smartphones to people, comparable to lending computers to households that did not have a computer to participate in the LISS panel.

Between November 2011 and May 2012, a series of pilot experiments were conducted to test the idea of lending smartphones to respondents. The goal was to find out if inexperienced respondents are willing and able to complete a time use survey on a smartphone, to optimize the time use app for inexperienced users, and to test the resulting data quality. We selected 50 participants who owned an Android smartphone and 50 participants without a smartphone who were provided with an Android smartphone by CentERdata.

Following the pilot experiments, we implemented the smartphone app in a larger and representative survey in the LISS panel in 2012/13. This large scale study aimed to compare the paper diary of the Dutch time use survey (conducted in 2011/12) to the smartphone diary (data collected in 2012/13). Furthermore, the study aimed to make use of the special features of smartphone data collection to enrich the normal time use data. The first smartphone function exploited was the push function, which was used to add Experience Sampling questions to the time use app. At six randomly selected moments, spread over the day, the time use app gave a notification and three questions popped up (Figure 1). The three questions asked how the respondent felt at that moment (how happy, how rushed, and how tired), using slider scales. When a respondent did not react, they disappeared from the screen after 10 minutes. The experience sampling data about feelings can thus be combined with the time use data collected in the same app using the same 10 minute time-slots. The second smartphone function we employed to collect additional data was the GPS. The time use app registered GPS data if the respondent had allowed that (consent to collect GPS data was asked at the start of the app) and if they had turned on the GPS of the smartphone. Finally, the app registered the number of telephone calls, text and Internet messages done with the smartphone. Participants were told in advance about this registration and were given the option to stop participating if they did not agree to that.

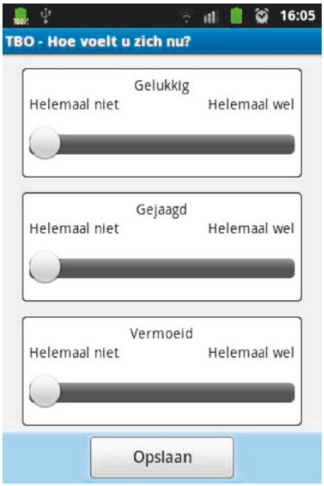

Figure 1 The three pop-up Experience Sampling questions, asking: Time Use Survey - How do you feel at this moment? Happy, Rushed, Tired. The scale labels are: Not at all – Extremely. The button below in the screen says ‘Save’.

Similar to the normal paper-and-pencil TUR survey, the data collection for the main survey was spread over one year. Each month, a different batch of about 170 respondents was invited to participate, in order to obtain a total net sample of approximately 2000 participants after 12 months. In this way, we could account for seasonality in the time use data. Another advantage of this design was that a small number of loan smartphones was needed since they could be lent to a different batch of panel members each month. Each participant filled in the diary on one weekday and one weekend day during the same week. The days were systematically spread over the participants.

Device

The time use app was developed for Android smartphones and for iPhones. The participants owning an Android smartphone or iPhone could download the app from App stores or via a link at the CentERdata website. Participants not owning an Android smartphone or iPhone on which the app could be used were provided with an Android smartphone (Samsung Galaxy GIO) by CentERdata, with the time use app already installed on it. In addition, an instruction video was developed showing how to use the loaned smartphone and the time use app. The app had to be installed and could then be accessed multiple times. Since the app also worked offline, respondents could fill in the diary anywhere at any time, independent of Internet accessibility. The diary data were sent to the research institute automatically ("synchronized") whenever a 3G or Wi-Fi-connection was available, without further action being required by the respondents. The loan smartphones were sent to the participants by post. After one week, in which respondents participated in the time use study on one weekday and on one weekend day, the smartphone were sent back to CentERdata in the same package and without costs. At CentERdata, the returned smartphones were cleaned and re-installed, and send out again to the next participant.

6 The Mobile Mobility study

Background and objective

In most countries, including the Netherlands, the understanding of people's travel behavior is based on cross-sectional travel surveys where only one day is surveyed for each respondent. This is not enough to gain a proper understanding of the dynamics in travel behavior and changes in behavior needed to reverse the long-term trends of growing mobility, congestion, increasing oil consumption and greenhouse gas (GHG) emissions. As part of a comprehensive research program conducted by the Centre for Transport Studies at the University of Twente, a field experiment was setup in the LISS panel using the GPS in smartphones to collect continuous long-term data on individual travel behavior. Smartphones are carried by people throughout the day and are therefore especially suitable for collecting accurate and extensive travel behavior data at low level of respondent burden. However, there was yet little experience with using smartphones as data collection tools in the transport field. The LISS panel experiment collected multiple-week and multiple-year travel behavior data from 600 panel members using smartphones. Similar to the design in the smartphone Time Use study, CentERdata again loaned smartphones to LISS panel members who did not own a smartphone themselves, in order to attain a representative sample of participants in the mobility study. The project was conducted by the Centre for Transport Studies at the University of Twente, in collaboration with CentERdata, the ICT research institute Novay, and the Dutch Ministry of Infrastructure and Environment.

Sample and study design

A pilot study for the mobility smartphone study was conducted in the LISS panel in January 2013. The first wave of the main study was carried out from April to July 2013, the second and third wave in the same months in 2014 and 2015. A random selection of 500 panel members who were willing to participate in the mobility study, including both smartphone owners and panel members without a smartphone, was selected for each wave. Each wave lasted about four weeks of fieldwork. As described for the smartphone Time Use study, the loan smartphones were sent to the participants by post and returned by them to CentERdata at the end of the four weeks.

Device

In the smartphone mobility study, the same loan Android smartphones were used as in the smartphone Time Use study. A mobility app was developed for iPhone and Android platforms by the ICT research institute Novay, in assignment by the University of Twente. The app automatically recorded trips using GSM/UMTS cell, Wi-Fi and GPS sensors. It detected whether or not the smartphone was moving and if so, it collected a location trace of the location measurements. If the smartphone remained stationary for some time, the app finalised the active collection and would send the data (i.e. location trace) to a server. Subsequently, on the server the location trace was cleansed with outlier detection and a trip was constructed from the location trace, using algorithms to automatically detect (multi-modal) tours, trip purpose and transport mode (Geurts et al. 2013).

In addition, an online interface was developed by CentERdata, where participants could check the registrations made by the mobility app and edit them if they were incorrect. In this way, the researchers of the Twente University could study the reliability of the app registrations.

7 Actigraphy study

Background and objective

Physical activity is a prime component of health behavior and accurate measurement is necessary for a better understanding of what drives differences in physical activity, how this is related to health, and how this can be influenced by policy. Obtaining internationally comparable objective measures of physical exercise has proven to be very difficult given its subjective nature. Almost all existing large-scale studies of physical activity have used self-report questionnaires. Although questionnaires are valuable, they also have their limitations, especially for international comparisons. First, they typically provide little information about the pattern of activity across the day and through the week. Self-reported measures are usually limited to certain aspects of daily activity such as structured exercise or walking and daily activities and sedentary behavior are not captured. Second, responses to questionnaires may not be accurate because of the cognitive challenge of estimating frequency and duration of activity and social desirability bias. Third, there may be important cultural differences across countries or across socio-economic groups in what constitutes exercise and vigorous exercise in particular. The development of accelerometers has opened up new possibilities for studying all intensity levels of physical activity from completely sedentary to vigorous activity available over periods of several days. The first aim of the LISS panel Actigraphy study was to understand the degree to which these limitations may distort the measurement and comparisons of physical activity. Secondly, the aim was to study patterns of activity and sedentary behavior through the day and week, using the continuous objective accelerometer measurement.

Sample and study design

The study involved wearing a wrist-worn accelerometer for 8 days, and additional questionnaires eliciting self-reports of physical activity and time use. In September 2012, a feasibility pilot was carried out among 200 LISS panel respondents, using 100 accelerometers. Starting in February 2013, the accelerometer main study was carried out on a sample of about 1000 LISS panel members, using 300 accelerometers. The accelerometers were sent to participants by post, with full instructions on its use and a log book. They were returned by post at the end of the 8-day period, after which CentERdata downloaded the data from the device, cleaned and re-installed it, and sent it to the next participant.

Device

GENEActiv accelerometers were used for the study. This is a wrist-worn accelerometer, which also measures near-body temperature and includes a configurable clock, allowing data to be matched to reported activity or other measures. The accelerometers are charged in a cradle connected to a USB power source. This cradle also allows for communication with the GENEActiv software used to configure the instruments, and download and manage recorded data.

8 Results

Response rates

Table 1 gives an overview of the response of the LISS panel members in the different innovative studies. For the two smartphone studies, we used one combined online invitation questionnaire, in August 2012. Respondents could choose to either participate in both studies (Time Use Study and Mobile Mobility Study), in only one of the two, or refuse both. They could also ask for more information before deciding, in which case they were called by our helpdesk. For the Time Use Study, a different sample of 170 panel members was selected each month from the pool of panel members who had indicated they were willing to participate in the study.

Table 1 Response to the invitations for all four innovative studies. Conditional percentages

	Weighing Study	Smartphone Time Use Study	Smartphone Mobility Study	Actigraphy study
Total invited	7264	7107[1]	7107[1]	2559[2]
Completed invitation questionnaire	72%	75%	75%	77%
Willing to participate[3]	56%	37%	37%	57%
Participated[4]	75%	68%[5]	81%[6]	90%

[1] The invitation for both smartphone studies was included in one questionnaire. Respondents could indicate if they wanted to participate in both studies, choose one of the two, refuse both, or get more information.
[2] A random subsample of the total panel was invited for the Actigraphy study.
[3] Conditionally on completion of the invitation questionnaire.
[4] Conditionally on a (random) selection from those who were willing to participate.
[5] Monthly average from October 2012 to September 2013.
[6] Response in the first wave, averaged over two batches of participants. Response is defined as having at least one trip registration in the mobile app.

In all four studies, about 75% of the panel members completed the online invitation questionnaire, reflecting the usual monthly response in the LISS panel. The willingness to participate, as indicated in these invitation questionnaires, was clearly lower in the two smartphone studies (37% in each) than in the weighing and actigraphy studies (56% and 57%, respectively). The lower willingness could reflect, for example, privacy concerns, a higher perceived respondent burden, or fear of not being able to use the technology. We did not study the reason for the difference in participation willingness. The actual participation rates, after panel members had agreed to participate and received the device, show a different pattern. The rate was lowest in the smartphone Time Use study (68%), which was the most burdensome study of the four, demanding respondents to fill in all their activities for two whole days and the experience sampling questions several times a day. It was followed by the weighing study (75%) which did not require a lot of effort from the respondents, but still that they would step on the weighing scale regularly. However, the relatively low response rate in this study can in part be related to technical problems with the data transfer of the weighing scales in the first period of the study, which we will discuss below. The Actigraphy study, which did not necessitate any response from the panel members other than wearing the wrist device, had the highest final participation rate of the four studies (90%). The Mobile Mobility Study had a high final response at well, 80% of those who were willing to participate actually did so. During the course of the day, this study was non-invasive because the app would record the trips of the respondent automatically. However, the checks and editing of the registered trips which the respondents were supposed to do online were a considerable demand, although they could be done at the end of the day.

If the willingness to participate in the innovative studies is related to specific respondent characteristics, the resulting net samples would be biased. This was studied by Fernee and Sonck (2014) for the smartphone Time Use study. They found no differences in the sample distributions of gender, age, education, household composition, working status and urbanization between the participants in the LISS panel smartphone Time Use study and the regular paper-and-pencil TUR of 2011/12 conducted by the SCP in the Netherlands. In contrast, a large difference in sample composition was found between the LISS panel sample members who participated with their own smartphone and those who were provided with a loan smartphone to participate: The smartphone owners who participated were on significantly younger, higher educated, more likely to have children in the household, working and living in more urbanized regions. This is an important result, since it not only proves the importance of including people who do not own a smartphone in a study that aims to be representative of the general population, but also shows that this can successfully be done by providing people with loaned smartphones to participate.

Technical and operational difficulties

Loaning smartphones and accelerometers to respondents and sending them by post has a certain risk of loss. This risk turned out to be relatively small: Until July 2013, we lost 18 of the 440 (4%) loan smartphones in use for the two smartphone studies and about 13 of the 300 (4%) accelerometers. A much larger problem for the Actigraphy study was the defect rate: over 90 accelerometers had to be returned to GENEActiv during the fieldwork because of defects.

Technical problems also occurred in the weighing study, causing some loss of data during short but recurring periods of time. The weight data were transferred from the gateway at the respondent's home to the CentERdata database through a server owned by the company that developed the weighing scales. At this server, an algorithm which was owned by the company transformed the impedance data to weight and body fat data. The company´s server had a few disruptions during the fieldwork period, which lasted shortly but would disorganise the automatic transfer of the data from many of the gateways in the respondents' homes. Respondents then had to reset the gateway, by dis- and reconnect it. Messages were sent to all participants at each occurrence of this problem, asking them to perform this reset, but these messages were not always read of followed up. This implied that the CentERdata helpdesk had to telephone all remaining participants to get the reset done and to get the data transfer working again. The data loss as a result of these disruptions was small in comparison to the overall amount of data collected with the weighing scales and did not pose a problem for the data analysis, but it did entail a considerable workload for the helpdesk of the survey institute.

The two smartphone studies did not suffer from major technical problems during the main data collection, after the app´s had been extensively tested and improved in the pilot studies. On the contrary, the smartphone Time Use pilot experiments showed that it was feasible to conduct time use research with an app on the smartphone for both experienced and inexperienced users. Respondents were able to use the app on a loaned smartphone to register their time use. With regard to data quality, the Mobile Mobility app in the first wave appeared to have some difficulty in correctly identifying modes of transport, in particular the use of tram, bus and train (Thomas et al. 2014). The algorithm of the app was therefore improved for the next waves of the study. This problem did, however, not cause any technical disruptions during the fieldwork. The only technical issue causing possible data loss was the battery use of the apps. The GPS registrations implemented in both the Time Use and the Mobile Mobility study caused the smartphone batteries to discharge rather fast. This was especially reported by the participants using a loan smartphone, perhaps related to their inexperience with the battery discharge of smartphones with Internet connections, or to the quality of the battery in the loan smartphone. The fast battery discharge was the main complaint of all participants in the two smartphone studies, expressed in an evaluation questionnaire.

The most important difficulty associated with the use of the apps in the LISS panel was the large investment it demanded in terms of man hours for software development and fieldwork operations. The development of an app for specific use in a single study such as the TUR study is a large and rather inefficient investment. A browser-based questionnaire can, when using one of the many user-friendly software packages available nowadays for online questionnaires, be programmed by researchers or other non IT-programmers, and be used on different types of computers, different browsers and often even on mobile devices without much adaptations. This is not the case for apps. Apps have to be developed for each operating system separately (for example for Apple IOS, Android, Nokia OS, Blackberry OS, etc.) and professional software developers are needed for that. Furthermore, after each change in the app, users (in our case the respondents) have to download and install it again. For use in a survey, where response rates are essential, it is therefore important to exhaustively test the app in repeated test- and pilot rounds, to minimize the need for adaptations and corrections after the start of the main fieldwork. The development of the user-friendly and profoundly tested time use app, including all the special features it had like push functions for Experience Sampling and collection of social media use data, required over 400 programmer man-hours for the Android version, and about another 400 for the iPhone version. For the Android version, these man-hours include the complete specification and building of app diary on the basis of the normal paper diary, and all adaptations after two pilot tests. For the iPhone version, the man-hours represent the re-programming of the already designed Android app for the IOS operating system and adaptations after a pilot study. Furthermore, both smartphone studies as well as the Actigraphy study demanded many operational hours, due to the logistics of repeatedly sending out the loan devices, cleaning and re-installing them, registration of incoming and outgoing batches, and extensive helpdesk support for the participants during the entire fieldwork period.

A final operational point that had to be taken into account is the treatment of the GPS data as collected in the two smartphone studies. GPS data can, in combination with other LISS data or even only on the basis of the GPS coordinates of one's home address, easily be used to identify a person. It was therefore necessary to develop extensively secured data transfer protocols and restricted access to the data for research purposes, in contrast with the normal free data access policy of the LISS panel. Researchers can apply for use of the GPS data, which can be assessed remotely on a protected server at CentERdata, without the possibility to download the data to a local environment or directly link them to other data.

Validation of self-report data

In all four studies, the data obtained with the innovative methods of data collection were compared with the self-report data obtained using the regular online questionnaires. The weighing study, the Mobile Mobility study, and the Actigraphy study proved the existence of some clear biases in the self-reports. For example, the average self-reported weight of participants in the LISS panel weighing study is 0.9 kilograms lower than the average actual weight for men and 0.7 kilogram lower for women. Furthermore, participants with a high actual BMI have a stronger tendency to underreport their weight whereas participants with a too low actual BMI, according to WHO standards, tend to over report their weight (Kooreman and Scherpenzeel, 2014). The Mobile Mobility app registered significantly more trips per person per day than is normally found in the traditional travel surveys conducted by Statistics Netherlands, thus proving the existence of an underreporting bias in this survey (Thomas et al. 2014)[4]. In the Actigraphy study, bias in self-reports of activity were related to respondent characteristics. For instance, older age groups self-reported that they are more physically active, in contrast with the objective measure taken by the accelerometers, which revealed that age is negatively associated with physical activity. Likewise, females, students, and high income group mentioned that they are more physically active than their comparison groups, whereas the objective measure showed no such differences (Kapteyn 2015).

The smartphone Time Use study, in contrast, resulted in nearly equal time use distributions as the regular paper-and-pencil TUR of 2011/12 in the Netherlands with regard to the five main categories of activities: Leisure time, travel, employment and study, household and family care and sleep/eat/personal care (Fernee and Sonck 2014). However, some differences in reported time use were observed for a few subcategories of activities, such as food preparation, eating (for both less time was reported in the smartphone study), or computer activities (more time was reported in the smartphone study). As was described in the section on response, the differences are most probably not caused by selection bias, as the sample distributions on some major

4 The app registrations also showed a slight under registration of trips in comparison with the online checks and adaptations to these registrations as done by the LISS panel participants. However, the number of trips per person per day as registered in the app is still clearly higher than the number found in the traditional surveys.

demographic variables were equal in both studies. Rather, they seem to reflect mode differences. We do not know which of the two modes, paper-and-pencil or smartphone app, reflects the true time use distribution, but a change of modes in longitudinal, repeated TUR can hence affect the comparability of the results for some of the more detailed levels of activities.

Data enrichment

Two of the studies reported here aimed to not only increase the accuracy of measurement by using different modes, but also to enrich the existing LISS data with supplementary special measures. We will shortly describe some examples of remarkable findings based on these supplemental data, referring to the publications about these results for more extensive descriptions.

The advanced weighing scales provided a measure of body fat, in addition to weight, which cannot be obtained through questionnaires. Kooreman and Scherpenzeel (2014) showed with the help of this additional measure that in the general population fat-based measures of obesity point at a substantially larger prevalence of obesity than BMI-based measures. Hence, the innovative mode of datacollection provided a true enrichment of the data. In addition, the weight data obtained with the scales documented the existence of a weekly cycle in body mass (Kooreman and Scherpenzeel 2014). This cycle can only be observed by collecting high frequency (daily) precise measures of weight, which is infeasible to do by means of daily questionnaires.

The Experience Sampling measures of mood, as implemented in the smartphone Time Use study are another example of supplemental data going beyond the normal time use data. Sonck and Fernee (2013) observed that the random pop-up Experience Sampling questions showed clear fluctuation patterns in moods during the day and across the days of the week, which cannot be seen based on a general estimation at the end of a fieldwork day. These fluctuations could then be related to the type of activities performed in the same time slots.

9 Conclusions

Technological development continues to offers new ways of data collection and new ways of mixing modes of data collection. In this paper, we have described how smartphone features and life-tracking technology can be used to collect more accurate data and supplemental measures in an online panel.

All four studies can be considered to be examples of the second type of mixed-mode study distinguished by De Leeuw and Hox (2011): Another Mode for Specific Questions, but in a new and pioneering way. A general aim of the four studies was to enrich the existing LISS data with additional measures, offering new, unique analysis possibilities for researchers in the domains time use, health, mobility and activity. Another aim was to validate data obtained with traditional self-report methods with these new and in some cases objective continuous measures. It was shown that the willingness of respondents to participate in such innovative studies as well as final actual response rate was satisfactory. Furthermore, in the smartphone studies, inexperienced users were also able to participate in the smartphone studies, using loan smartphones with the apps installed on it. The principle of providing smartphones to respondents in the studies described here is similar to the strategy of providing computer equipment and internet access to non-internet users in order to enable them to participate in an online panel, as the LISS panel and other similar panels for scientific use have been doing for about a decennium. In the smartphone Time Use study, it resulted in a net sample of participants which was clearly more representative than a sample of smartphone users only.

The weighing study and Mobile Mobility study proved, with the help of the objective measures, the existence of bias in the self-reports of weight and travel behaviour. Furthermore, the weighing study and smartphone Time Use study reported fluctuating patterns in weight and in mood across days and weeks which can only be observed using detailed high frequency measures.

In some cases, further technical improvements in the devices used are still needed. However, the main impediments for this type of study are the amount of time and manpower that is needed to implement the technologies and the complex operations and logistics. As Couper stated in his view on the future of survey methods (Couper 2011), the times demand for survey methods which are easy and undemanding, since respondent's time and willingness have become precious. The weighing scales, mobility app, and accelerometers described here are good examples of methods which automatically register behaviour ("life-tracking" methods) or demand little effort from the respondents. However, these methods move the burden from the respondent to the researcher or the research institute: Whereas the respondents only have to wear a wrist-watch, carry their own smartphone with them or step on a weighing scale like they would normally do, researchers and fieldwork management have to invest much more in the survey preparation and support than when they use only regular computer assisted questionnaires. For each future study, such investments must be weighted against the gains in response, data quality and data enrichment.

References

Antoun, C. (2015). Mobile Web Surveys: a First Look at Measurement, Nonresponse, and Coverage Errors. Doctoral Dissertation, the University of Michigan.

Avendano, M., Scherpenzeel, A., & Mackenbach, J. (2011). Can biomarkers be collected in an Internet survey? A pilot study in the LISS panel. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies*, New York: Taylor & Francis Group.

Blom, A.G., Bosnjak, M., Cornilleau, A., Cousteaux, A., Das, M., Douhou, S., & Krieger, U. (2015). Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Social Science Computer Review 0894439315574825,* first published on March 31, doi: 10.1177/0894439315574825.

Connor Gorber, S., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews*, 8, 307-326.

Couper, M.P. (2011). The future of modes of data collection. *Public Opinion Quarterly*, 75, (5), 889-908.

De Bruyne, M. (2015). *Designing Web Surveys for the Multi-Device Internet*. Doctoral Dissertation, Tilburg University.

De Leeuw, E.D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21, 233-255.

De Leeuw, E.D., Hox, J.J., & Dillman, D.A. (2008). Mixed-mode surveys: When and why. In E.D. de Leeuw, J.J. Hox, & D.A. Dillman (Eds.), *International Handbook of Survey Methodology*. Boca Raton: CRC Press.

De Leeuw, E.D., & Hox, J.J. (2011). Internet surveys as part of a mixed-mode design. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies*. New York: Taylor & Francis Group.

Eurostat (2009). Harmonized European time use surveys, Guidelines 2008. Methodologies and Working Papers. Luxembourg: Office for Official Publications of the European Communities. http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-08-014. Zugegriffen: 26. Juli 2016.

Fernee, H., & Sonck, N. (2014). A comparison between time-use data collected by smartphones and a paper diary [powerpoint slides]. Presentation at the *XV111 ISA World Congress of Sociology*, 13-19 July 2014, Yokohama, Japan

Geurs, K., Veenstra, S., & Thomas, T. (2013). The setup of a mobile mobility panel for the Netherlands. Paper presented at the 13th The World Conference on Transport Research (WCTR), July 15-18, 2013, Rio de Janeiro, Brazil.

Groves R.M., Cialdini, R.B., & Couper, M.P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, 56, 475-495.

Hoogendoorn, A., & Daalmans, J. (2009). Nonresponse in the recruitment of an internet panel based on probability sampling. *Survey Research Methods*, 3, 59-72.

Kapteyn, A. (2015). *Comparing subjective and objective physical measures across countries*. Manuscript.

Kooreman, P., & Scherpenzeel, A. (2014). Advanced Body Mass Measurement, Feedback, and Health Behaviors. *Economics and human biology*, 14, 141-153.

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50, 79-104.

Lugtig, P., & Toepoel, V. (2015). The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error, *Social Science Computer Review,* 34, 78-94.

Saris, W.E. (1998). Ten Years of Interviewing Without Interviewers: The Telepanel. In M. P. Couper et al. (Eds.), *Computer Assisted Survey Information Collection* (p. 409-429), New York: Wiley.

Scherpenzeel, A., & Das, M. (2011). True Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet:*

Advances in applied methods and research strategies, New York: Taylor & Francis Group.

Sonck, N., & Fernee, H. (2013). *Using smartphones in survey research: a multifunctional tool. Implementation of a time use app; a feasibility study.* SCP-publication 2013-22, The Hague: The Netherlands Institute for Social Research.

Thomas, T., Geurs, K., Bijlsma, M., & Douhou, S. (2014). Hoe mobiel zijn we eigenlijk? Eerste inzichten uit het Mobiele Mobiliteitspanel (How mobile are we really? First insights from the Mobile Mobility Panel). *Tijdschrift Vervoerswetenschap*, 50 (3), 138-154.

http://www.springer.com/978-3-658-15833-0

Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung
Eifler, S.; Faulbaum, F. (Hrsg.)
2017, VIII, 272 S. 27 Abb., Softcover
ISBN: 978-3-658-15833-0