ESSnet Smart Surveys

Grant Agreement Number: 899365 - 2019-DE-SmartStat

Link to our CROS website

Workpackage 2

Smart Survey Pilots

Deliverable 2.1: health - measuring physical activity (WP2.3)

Version 2.0, 14-04-2022

Prepared by:

Annemieke Luiten, Vera Toepoel, Barry Schouten, Statistics Netherlands
Marek Cierpiał-Wolan, Katarzyna Kapica, Piotr Szlachta, Statistics Poland
Patrick Lusyne, Hannelore van der Beken, Statistics Belgium

ESSnet co-ordinator:
Shari Stehrenberg and Markus Zwick (DESTATIS, Germany)
smartsurveys@destatis.de

Workpackage Leader:
Barry Schouten (CBS, Netherlands)
Jg.schouten@cbs.nl
telephone : +31 70 3374905

Contents

0. Summary

WP2 performs four diverse pilots to inform WP3 on the specifications of a smart survey platform in the European Statistical System (ESS). Important questions are to what extent design and architecture of smart surveys can be shared across ESS countries and what are country deviations. Deliverable 2.1 describes functionality, usability (Phase 2), methodology and infrastructure (Phase 3) for each pilot. The deliverable 2.1 before you describes work package 2.3 and is about the objective measurement of physical activity by sensors in an accelerometer. In addition, an extensive description is given of the assessment phase (Phase 0) preceding the Phase 2 and Phase 3 activities. Three countries participated in pilots: Poland, Belgium and the Netherlands. Our main conclusion is it is feasible to use objective measurement on a large scale in the general population, although there is much that can and needs to be improved. Response rates in the (web) pilots were generally low and biased towards people who move more than enough according to the WHO guidelines for physical activity. Reaching that part of the population that does not move (enough) will probably require the use of interviewers. Using objective measurement is expensive, both as a result of the devices, but also as a result of the probable need to use incentives. It is in addition logistically complicated. The final and most important question if these costs and complications are worth it in view of added quality and richness of the data can as yet not fully be answered. Substantive analyses are planned for the near future but are out of scope for the present ESSnet.

1. Introduction

Smart surveys are oriented at reduction of respondent burden and improvement of data quality by using the features of smart devices. Six such features have been identified: local storage and processing, employment of internal mobile device sensors, linkage to external sensor systems, linkage to publicly available online data, data donation through the respondent and data donation through the national statistical institute. The inclusion of smart features is promising in particular when survey topics are cognitively demanding, require detailed knowledge or recall, and/or corresponding questions are weak proxies of the concepts of interest. The promise of smart surveys only will only be real when target population response rates are relatively high and balanced and respondents understand the smart survey tasks they are invited to perform. Willingness and usability imply that respondents need to consider the tasks as legitimate and logical and need to trust the statistical institute. These requirements also hold for 'non-smart' surveys, but become more prominent in smart surveys. In these surveys respondent devices are being used and data are being collected or linked that are in part unknown to respondents themselves. Phase 2 functional and usability tests form the stepping stone to field tests and are a guarantee that technically and conceptually smart surveys deliver adequate data. Phase 3 tests inform on how to operationalize methodological and logistical choices.

WP 2.3 (Heath) fits the smart survey eligibility criterion, in as much as we measure physical activity in far more detail and precision than respondents are able to. Physical activity and sedentary behaviour are presently measured by survey questionnaires. An example is the SQUASH module in Health Surveys. The SQUASH is not very accurate however (people overestimate their sports activities, but underestimate how much they move in normal day to day activities). In addition, the SQUASH does not give insight in how people move: do they go to the gym once a week for two hours, or do they move each day for some time; and what is the distribution of muscle strengthening and bone strengthening activities? Sensor measurement, in combination with a machine learning algorithms may form a high quality method that gives precise information on people's physical activity, sedentary behaviour and sleep patterns. In order to be feasibly used in general population fieldwork, the device chosen should be modestly prized; render consistent, reliable, repeatable and valid measurements, and allow easy data extraction. Of course, respondents need to be willing to wear the device for one week and be able to use them without assistance.

Although for the entirety of WP2 the primary question is if it is possible to translate smart survey data to accurate and comparable statistics, the main perspective in WP2.3 is not on the statistics themselves, but on respondent acceptance, respondent burden, respondent support, privacy, logistics, data quality, analysis and costs.

In the project documentation for WP2, a number of phases is distinguished. WP2.3 participates in the following phases:

- Phase 0 – inventory: In this phase an inventory was made of potential devices, a test was done with some of them, leading to a choice of device to be used in the subsequent Phase 2 testing. Part of this phase took place prior to the start of this ESSnet on. Other questions addressed in this phase are: what is the quality of the measurements, on which part of the body should the device be mounted, how long do we need to measure, how do we get the data from the device to the backend server, and how do we make data available for researchers, while preserving respondent privacy. This phase is described in section 2.
- Phase 1 – concepts: This phase was dedicated to the design and preparation of test 1. It addressed the goal of the subsequent field tests, defined for WP2.3 as: to what extent are sample persons willing to wear an accelerometer, can we perhaps use people's own devices, how many people possess an accelerometer or smartwatch or use an app to monitor their behaviour, can we influence willingness, by giving or promising monetary incentives, or by promising feedback on physical activity? What are barriers for participation? Some of these questions are addressed by qualitative research, and some by quantitative research. Section 3 describes this phase.
- Phase 2 – testing: In the testing phase, the tool is tested in a few small and larger scale pilots. Section 4 describes the pilots' setup and findings.
- Phase 2' – process requirements: In the process requirements phase, that runs parallel to the testing phase, the requirements of the tools are specified. This phase is mainly described in Deliverable 2.8 – physical activity. Here, we will focus briefly on the machine learning necessary to derive and summarize statistics from the sensor data.

Although in this WP it is quite obvious what questionnaire elements could be replaced by the objective measurements (i.e., the subjective questions on physical activity and sedentary behaviour, like the SQUASH), the substantive aspect of how this replacement is accomplished is not part of this deliverable. In section 5 is sketched that extensive additional development is needed before the relations between the objective and subjective measurements are understood and how the familiar statistics can be derived from the objective physical activity measurements.

2. Phase 0: assessment

The phase 0 pilot was set up to determine the accelerometer with which to perform the subsequent pilots. The pilot was performed with a convenience sample of colleagues, friends, students, and respondents from a previous health survey who had indicated willingness to participate in follow-up research. The number of participants was 40. The participants came to a laboratory, where they performed a number of prescribed and observed activities, while wearing different sensors. For some of the activities, participants breathed in a respiratory mask to measure their energy expenditure. The sensors that were compared were the activPAL, worn on the thigh, an X-IMU worn on the shin, an X-IMU worn on the wrist, a UKK, worn on the hip, and a Hexoskin, worn on the body. Participants performed the following activities: sitting, standing, walking, running, jumping, climbing stairs, and bicycling. After the lab session, participants wore two of the sensors (the UKK and the activPAL) for one week during a 'free living period' of one week. During that week they filled in a paper diary, registering the start and end time of physical activities like walking, biking, doing sports, non-wear, travelling. The pilot study and the subsequent data analysis are described in detail in Hoek, Windmeijer, Luiten, Bolte and Schouten, 2022.

The following research questions were addressed in Phase 0:

2.1 Which accelerometer should we use?

Machine learning algorithms were developed for each of the sensors, with the aim of predicting the activities performed in the lab and the free living period. The activPAL was shown to have the best performance overall, based on the accuracy of the classification of activities. However, accuracy and quality were not the only criteria to choose the activPAL. Other considerations were

- Costs: although the activPAL is quite costly to purchase (about €300, although considerably cheaper in larger quantities), they are very robust and can be used for many times for many years. Refurbishing is possible against moderate costs.
- Size: the small size of the device makes it easy to send to participants (or to carry a number of them in luggage, should an interviewer be involved), and also for participants to wear unobtrusively.
- User friendliness: the activPAL can be stuck on the leg with a medical plaster and can stay on for the entire measurement week. Older versions need to be taken off for swimming, but the newer versions can stay on even then. The participant may hardly notice it, although some participants complain of itchiness, either from the plaster or from the rubber casing that is generally used to envelop the activPAL. In addition, mounting the activPAL is easy for most participants. Should they do something wrong, however, like pasting it the other way around, the activPAL software detects and corrects that.

- This increases participant adherence. There is less chance that the participant forgets to put the device on again. Also, the device can stay on all day, where devices on the wrist may be forbidden for e.g., care personnel.
- The device is a black box for the respondent. They cannot see their results. This on the one hand makes it unattractive for the participant to keep the device, and on the other hand may limit the risk that the participants change their behaviour.
- There is a large research community in the world, also working with activPAL or comparable thigh mounted devices. They share knowledge and open source algorithms. The community also made it possible for us to lend large amounts of activPALs for our research projects.
- ActivPAL allows downloading of raw data, to use with our own algorithms
- However, various summary data file can be downloaded for those who want to use the activPAL algorithms.
- Some of the output can readily be used for attractive feedback to respondents.

2.2 What is the quality of these sensors?

High quality measurement can only be guaranteed if many conditions are met. The device needs to measure uninterruptedly, the respondent needs to apply it to the intended body part in the intended direction, and will need to keep it on during the designated amount of time. But first and foremost: the device needs to measure what it is intended to measure, i.e., physical activity and sedentary behaviour. In addition, the device needs to be able to differentiate between movements that are real activities and those that may be 'fidgeting', e.g., the movement of the legs while sitting. The large advantage of using a tried and tested device is that this research has already been done by others. All of the sensors tested in phase 0 and the selected device are high grade sensors. The validity of their measurements to determine physical activity has been determined in numerous validation studies in various populations (adults, children) and cultures, e.g., Berendsen, et al., 2014; Crowley et al., 2019; Gorzelitz et al., 2020; Júdice et al., 2015; Wu et al., 2020. Nevertheless, even potentially reliable products are dependent on adequate use and especially relatively long periods of measurement. In general, a wear period of seven days is used in literature, but for reliably classifying some activities, that may not be enough. This topic is discussed in section 2.5.

We have found in the Phase 0 pilot that accelerometers also have their faults, and may for example classify riding a bus as moderately intensive activity (Hoek et al., 2022). In addition, the choice of epoch length and the determination of the cut-off point can all impact the measured intensity. Careful handling of these parameters is necessary in order to measure reliably (Brondeel, 2016).

2.3 On which part of the body should the accelerometer be placed?

The choice for the activPAL also indicated a choice of position on the body. The position on the leg is especially suitable to distinguish bicycling (Tarp, J., Andersen, L., & Østergaard, L., 2015), an important feature in the Netherlands. The UKK, for example, worn on the hip, was not sufficiently able to recognise bicycling (Bikker et al., 2020). The ability to recognize bicycling added substantially to the overall better performance of the activPAL. Other thigh worn accelerometers than the activPAL may perform just as well (Crowley et al., 2019), but see the other considerations for the activPAL described above.

2.4 Is one accelerometer sufficient, or should they be appended with e.g., heart rate measurement?

Hoek et al., (2022) state: "From the point of view of accuracy in the models, all models benefitted from adding heart rate as an additional feature. However, the added specificity of heart rate was not very large: the accuracy of the classification models based on activPAL data increased with 0.034 when heart rate was added. Is was difficult to accurately distinguish moderate intensive and vigorous cycling without heart rate, since these activities consisted of the same frequency of movement in the phase 0 pilot". Additional sensors could be very valuable in adding context to the measurements, however. GPS data could inform on the location of the physical activity for example, a type of information that is sought after by consumers interested in policy and intervention. GPS could also give information on the stretch that is run or cycled in a given time, and hence the intensity of the movement. Adding GPS data would on the other hand make the data far more privacy sensitive. GPS is not included in the activPAL, and would need an additional device. As far as we know, no thigh worn accelerometer includes heart rate at the moment, although other kinds of accelerometers do include those. Any additional device would potentially add to respondent burden, to costs and to the logistic complexities of getting the devices to and fro. For the pilots described here, only one device per participant was used.

2.5 How long do we need to measure?

Hoek et al., (2022) state: "Because the physical activity norm states the required number of active minutes per week, a week is also the measurement unit of time chosen in this study, as well as in most other studies with accelerometers, although shorter periods of four days may also render valid estimates of physical activity (Edwardson et al., 2017). Edwardson et al., (2017) give an overview of studies using the activPAL. In 71% of studies, the activPAL was worn for 7 days. Ideally, the number of days of measurement is determined by the required reliability, which may be different for different types of behaviour (sitting, standing, walking, etc.) and different populations. Most likely, there will be variation over days of the week, over weeks and over seasons in the amount of physical activity. Questionnaires tackle this variability by asking for a subjective estimation of the mean number of minutes over longer periods of time. Longer measurement periods with accelerometers could theoretically inform on within respondent variation and could potentially better inform machine learning models. However, better measurement needs to be balanced with higher respondent burden and diminished adherence. Skin irritation may be the result when wearing the activPAL for longer periods, for example. There is also the practical consideration that there is a finite amount of recording capacity on the device. Edwardson et al., (2017) conclude that, "pending better recommendations, the measurement period should be at least 7 days but ideally up to the 14 day limitation of the activPAL". In the second Polish pilot, a period of 14 days was used, see section 4.3.

2.6 How do we get data from the measurement device to the backend server?

The data from the activPAL are stored on the device during the measurement period. Once returned to the researcher, the device is connected to a computer with activPAL software, by means of a USB cable. The software lets the researcher choose what kind of data is generated. This could be the raw data, but also one of a multitude of summary files. See deliverable 2.8 – Physical activity for more details. Statistics Netherlands chose to download the activPAL software on a separate laptop, and not on the main frame. Hence, this is also the place where the downloaded data are initially stored. A choice can be made whether to deploy the algorithms that summarise the data into the variables needed for the statistic at this moment too. Then, only the ID and summary information need to be imported to the backend server. Although importing the 'mother file' would be wise as well, as that will allow the researcher to generate all the data that someone may later need, without the need to store the bulky raw data; one week of raw data per person amounts to 450MB of data; for 500 respondents this would already be 225GB. The mother files for the same time period are 9MB large.

2.7 How do we make data available for researchers while preserving respondent privacy?

The data on physical activity can as such not be identified as belonging to an individual, nor can they be used to identify an individual. Perhaps in very extreme circumstances, e.g., if someone's data showed that he skated for 12 minutes, 30 seconds and 74 milliseconds on a certain time on February 11th 2022, there would be a high probability that this person would be Nils van der Poel. But otherwise, people's movements are not quite so distinctive. Additional information from the sampled person from either records or survey is needed to identify the person, and that is only done after the data from the activPAL is imported into the secure NSI-backend. The potential hacker of the data would also need to register on the activPAL platform in order to be able to download and use the software needed to read the data.

The data protection officer of CBS deemed the (unlinked) data not sensitive in the sense of the GDPA. Hence, we were allowed to share the data with another NSI, providing that the participants gave informed consent to do so. A secure data transmission was set up between CBS and ISTAT in order to be able to share the micro data. Once the data are linked, however, standard privacy preserving practices come into play.

3. Phase 1: design and preparation of test 1

3.1 Goal of the tests:

In this section, a short overview is given of the pilots performed in the participating countries, and the research questions that we posed in them. The Netherlands has performed three pilots, Belgium one pilot, and Poland two pilots. Section 3.1.1. to 3.1.3 gives a short description of the pilots and the research questions.

3.1.1 The Netherlands

Statistics Netherlands has performed three pilots in Phase 1:

1. a survey among 4997 members of a commercial nonprobability panel, followed by a small fieldwork phase (n=80) with accelerometers among a subsample of respondents.

2. a survey among prior respondents in two regional health monitors who indicated willingness to be approached for additional research (n=6801 and 5711 respectively), with an additional field phase among a subsample of 523 persons to wear the accelerometer.

3. a survey among participants in the LISS panel, a probability sample of the Dutch population. The number of invited panel members was (3874). Subsequently fieldwork with accelerometers was performed among about 500 participants.

The goals of the first pilot in Phase 1 were:

1. Studying willingness of respondents to a questionnaire on physical activity to wear the activPAL
2. Studying possession of private accelerometers and the use of apps for measuring physical activity.
3. Studying the feasibility of using people's own accelerometers, in the form of data copied from their device into a questionnaire. Participants wore an activPAL and their own device during the measurement week. For personal trackers to be useful in physical activity surveillance, a substantial and representative part of the population should possess an activity tracker. Those who possess a tracker must be willing to share their data, and the data must give a valid and reliable impression of people's physical activity. Studies have shown moderate to strong correlations between measurements with consumer-level activity trackers and research-grade devices (e.g., Ferguson et al., 2015). Using people's own devices could thus be a good alternative for professional devices. Using them in large-scale physical activity surveillance has not been studied before, however (Kraakman, 2021).
4. Studying willingness of respondents to a questionnaire on physical activity to upload the data of their own device into a questionnaire
5. Studying the quality of the data provided by the own accelerometers
6. Studying willingness of respondents to donate the data of their own device over a longer period of time by downloading the data from the device's cloud environment.
7. Studying the effect of the promise of feedback on physical activity on willingness to participate in the field study with accelerometers
8. Finding out what participants considered a fair incentive for a week of participation
9. Finding out what respondents perceived as barriers to participation
10. Comparing data of the activPAL and people's own measurement devices with their answers on the SQUASH questionnaire
11. Gauging the usability of wearing the activPAL via an evaluation questionnaire.

While the first pilot was performed in a commercial non-probability panel, the second pilot was performed among the general public: respondents to a health survey who had indicated willingness to participate in follow-up research. The pilot was commissioned and paid for by two public health services in the Netherlands and implemented by the same commercial institute that performed the first pilot for CBS. CBS participated in the analysis of the activPAL data, and was allowed to introduce two experimental conditions, aimed at increasing participation. The goals of the second pilot were to:

12. Study willingness to participate in a field test with the activPAL among the general public
13. Study the effect of two monetary incentives (€12,50 or €25) on participation rate
14. Study the effect of an invitation that was targeted on the participant in terms of their activity level

The third pilot was a large scale replication of the first pilot in a probability panel in the Netherlands (the LISS panel). Apart from the research questions formulated for the first test, in this test we additionally studied:

15. The feasibility of using data from participants' own devices over a longer period of time, by data donation

Details of these pilots will be described in the section 4.

Not all research questions could be answered from the data gathered in pilot 1, as only half of the participants in the accelerometer part also filled in the SQUASH questions (the other half received an experimental alternative on the instigation of the panel holder), making the number of usable comparisons too low. For research questions 5 and 10 we await the results of the third pilot, that is presently still in the field.

3.1.2 Poland.

Poland has performed two tests. The first pilot was a small scale feasibility test with 10 participants from a convenience sample of colleagues. The goal was to:

16. To check how the software and hardware works
17. To check what data are collected and how sensitive and precise they are

The second test was among 20 participants from a convenience sample of people interested in taking part in the survey and the results it provides. The goal was to:

18. To check overall physical activity of people and how it compares to their own estimation of their activity
19. To check how many hours a day/week people stay active.
20. To check if the awareness of wearing such a device had an influence on people physical activity
21. To check if people are willing to take part in surveys using smart devices and whether they need additional motivation to do so.

3.1.3 Belgium.

Statistics Belgium conducted one small pilot study. The focus was on challenges that come with the adoption of this kind of technology as an NSI, less on conceptual or methodological issues and results. Goals of the pilot were:

1. To explore the willingness to wear a motion sensor device
2. To assess participants' experiences with the motion sensor device
3. To obtain participants' opinion regarding passive measuring methods
4. To gauge participants' perception concerning privacy.

4. Phase 2: tests

In this section we start by describing the pilots in greater detail, followed by the findings to the research questions stated in section three.

4.1 NL pilots description.

Three pilots with activPALs were performed during this ESSnet, although not all were specifically performed for the ESSnet. Nevertheless, results that are relevant for the ESSnet will be described here as well. In addition, a pilot was performed by the Dutch National Institute for Public Health and the Environment (RIVM) in collaboration with CBS into the use of accelerometers in population surveillance. The accelerometers used in this project were UKKs (de Hollander, Toepoel, Duijvestijn, Wendel-Vos, Luiten, Schouten and Loyen, 2022). Elements of this pilot are relevant in this context and will be described as well.

The first pilot existed of a survey among panel members of a commercial panel into possession of a smart meter of fitbit, willingness to wear the activPAL for a week, willingness to either copy or upload data from their own accelerometers, smart watches and smartphone apps (n=4997), and subsequently a small fieldwork phase (n=80) with the activPAL, half with participants with and half without own accelerometers. Panel members received an email invitation to fill in the questionnaire, that was introduced as a questionnaire on physical activity. The panel sample was drawn to resemble the Dutch population. No reminders were sent. Participants were promised an incentive of €20 for wearing the activPAL at the moment in the questionnaire where the participation question was posed. At the end of the wear week, participants received an evaluation questionnaire that went into aspects of usability and data quality, and also asked participants with a private measurement device to copy the data of the previous week into the questionnaire. See appendix 6 for the questionnaire.

The activPAL data were compared to data from respondents' own devices and to answers to the SQUASH survey. An experimental condition was included to study respondent willingness: the promise of feedback on activity during the measurement week. Details of this pilot can be found in Toepoel, Luiten and Zandvoort (2021), Kraakman (2021), Kraakman, Luiten and Toepoel (2022).

The second pilot was held among respondents in two regional health monitors who had indicated willingness to be approached for additional research (n=6801 and 5711 respectively), with an additional field phase among a subsample of 523 persons who were invited to wear the activPAL. Sample persons received an email invitation and one reminder to participate in a web survey on health related issues. The questionnaire inventoried physical activity behaviour, and social economical background. Experimental conditions into respondent willingness in these studies were the use of incentives (€12,50 vs €25) and the use of general versus targeted invitations. The targeted information addressed respondents' activity level, that was calculated from the answers to the SQUASH questions in the recruitment questionnaire.

The targeted information for people with low activity levels read:

> *We all know that exercise is good for our health. But how much do we move exactly? Maybe you don't exercise a lot, but that doesn't have to mean that you move (too) little. You will be surprised how much you move by doing everyday things like vacuuming, shopping or cycling to work. Because many people do not realize how much they move by these kinds of daily activities, this information is not properly counted in questionnaires about movement. But these daily movements are very important, and can be measured by an accelerometer.*
>
> *We would like to get a better idea of how much people exercise, especially outside of sports. Therefore, we would like to ask you if you would like to wear an exercise meter for one week.*

The targeted information for people with high activity levels was:

> *We all know that exercise is good for our health. But how much do we move exactly? Do you exercise intensively during the weekend while sitting mostly at a desk during the week? Do you exercise for an hour every day? And how much movement comes from 'home and kitchen' activities? Important questions for exercise research, but difficult to measure with a questionnaire. This information can be measured with an accelerometer.*
>
> *We would like to get more insight into how exercise is distributed throughout the week and how much people exercise outside of sports. Therefore we would like to ask you to wear an accelerometer for one week.*

The neutral information was:

> *We all know that exercise is good for our health. But how much do we move exactly? We can ask you that (and have already), but questionnaires are usually not very good at measuring the important 'home and kitchen' activities that are also a part of our movement. This can be done with an accelerometer meter.*
>
> *For this study we are looking for people who are willing to wear an exercise meter for one week. The purpose of the study is to get more information about how movement is distributed over the week and how much people move, through sports but also outside of sports. Whether you move a lot or little, we need your participation to be able to answer this question.*

All three invitations continued with the following text:

The accelerometer is small and thin. It is stuck on the upper leg with a medical plaster. It can stay there for the entire week.

The meter only measures your movement: whether you are sitting, standing, lying down, or walking. It does not *measure where you are.*

For participating in this study, you will receive a [incentive] worth [€12,50 or €25]...

Are you willing to wear an accelerometer for a week?

An evaluation questionnaire determined if the measurement week was representative of normal weeks in terms of the amount of physical activity.

Regretfully, through a mistake in the operationalization, both experimental conditions were collapsed, such that the high incentive group was also the group with the targeted invitation, and the low incentive group received the general invitation. As a result, there were two experimental groups instead of four: an high intensive and a low intensive recruitment group.

The third pilot consisted of a survey among participants in the LISS panel, a probability sample of the Dutch population. The number of invited panel members was 3370. Panel members were invited via an email to fill in the web questionnaire, that was framed as the standard monthly questionnaire that LISS panel members receive. One reminder was sent. A questionnaire inventoried physical activity by means of the SQUASH questions, possession of accelerometers, background variables, health, sedentary behaviour, willingness to participate wearing an accelerometer, willingness to share own device data by copying data into a questionnaire, and willingness to share data by data donation. The fieldwork among 500 willing participants (400 with own meters, 100 without) allows to compare subjective survey measurement, accelerometer data, own device data that were copied from the device into a questionnaire, and Fitbit data from one month that were donated by the participants. Participants received an incentive of €15 for participating. Experimental manipulations were the absence or presence of a picture of the activPAL with the invitation to wear the activPAL , and textual variations on ease, privacy, and the duration of the period of the donated data.

In none of the pilots people were excluded based on for example having a pacemaker, skin conditions or anxiety disorders. People with questions could contact the contact the panel holder, but no special provisions were taken for them to contact for example medical doctors. Ethical approval for piloting with accelerometers was provided by the CBS ethics board, but no additional medical-ethical approval was asked for, nor deemed necessary.

4.2 BE pilot description

Statistics Belgium performed the test among 10 participants from a convenience sample of colleagues. There were 25 sensors available for the test, but logistical issues due to COVID-19 caused a significant dropout. On the one hand, a number of willing colleagues did not have the ActivPAL monitor in their possession at the start of the test. On the other hand, a number of participants were not able to return the accelerometer on schedule, which meant their data could not be downloaded in time. Therefore, these people were excluded from the test. The test itself consisted of two phases. During the first phase in October, the participants had to wear the ActivPAL monitor for seven consecutive days. They received all instructions via e-mail (see Appendix 1a for a translation of the manual). Afterwards, the ActivPAL monitors were returned to CBS who downloaded the data. In the second phase of the test – nearly two months later – the participants received their personal movement profile as obtained from CBS (see Appendix 2) and were asked to complete a questionnaire about their experience with the ActivPAL monitor and their opinion on the use of such technologies (see Appendix 4 for the survey questionnaire). We made the questionnaire available through MOTUS – the online software platform for data collection of the VUB/Hbits. Participation was possible via a mobile or a web application.

4.3 PL pilot description

In Poland two pilot surveys were conducted during the ESSnet Smart Survey project. Poland has received the accelerometers from CBS.

The first pilot survey was conducted in February 2021. Participants were chosen among people willing to participate in the testing and wear accelerometers for one week. Additionally, participants were asked to keep a general diary of their activities thought a day. The diary had to include main activities like: time of sleep, time of going to and from work as well as shopping and additional activities that participants decided to be important to mark (exercising, walking, etc.). The diary was kept to see if and how the activities are represented by the data collected by the device. Later, in the second pilot keeping a diary has been abandoned as the data well represented the activities of the person wearing the accelerometer.

The second pilot survey was conducted in February 2022. The participants were also chosen among people willing to participate in the testing. This was due to the tight Covid-19 related restrictions in Poland. The participants wore the accelerometers for two weeks. This time participants were instructed on how to wear and mount the accelerometers in person as well as in written form. The written instruction was provided in case there was a need to change the protective patch. The instruction can be found in Appendix 1b to this Deliverable. Additionally, the participants were asked to fill in two surveys (Appendix 3 to this Deliverable). The first survey *General physical activity survey* was to be filled in before the testing started. The questions related to the age and sex of the participants as well as their overall assessment of their physical activity during a day and week. It also included a question on a sport they preform if they sport. The second survey was to be filled after the accelerometers were removed. The questions related to the overall assessment of use of the device. The questions related to the willingness to participate in similar studies in the future (discomfort of use, duration of the testing, compensation). One week after the survey ended, the participants were sent a report generated by the accelerometer software, see appendix 2 for an example.

4.4 Findings

For the description of the findings, the list of goals, as described in section 3.1, will be followed. The findings in sections 4.4.1 to 4.4.5 and 4.4.8 are based on the quantitative Dutch pilots. In the other sections the more qualitative Polish and Belgian findings are incorporated.

4.4.1 To what extent are sample persons willing to wear an activPAL or other accelerometer?

Even though the populations in the three Dutch studies were very different, both in terms of population and fieldwork strategy, there is a remarkable similarity among respondents of the pilots in consent to become participants in the accelerometer fieldwork phase. Table 1 shows results for the three pilots. As the second pilot was performed in two regions, under slightly different circumstances, these results are described separately. In all those pilots, about half of the respondents expressed willingness to participate in an accelerometer experiment. However, in the subsequent step of actually wearing the activPAL, quite a number of respondents dropped out, and here we see differences between the samples and / or fieldwork designs. In the final step, actually wearing the activPAL for the minimal number of days (defined as four), an additional number of participants drops out, so that in the end, depending on the pilot, only about 25% to 44% of the respondents of the questionnaire provided valid data. Not all drop-out in the final stage is due to the participants, however. We also suffered from technical and logistic difficulties, for example with activPALs stopping recording before the week was over, and postal deliverables that did not come in time, causing participants to have to measure fewer days.

The differences between the pilots indicate that there are fieldwork circumstances that influence each step from consent to wearing and to keep wearing. They also indicate that fieldwork needs to be planned and executed with utmost attention.

Table 1. Willingness to wear the activPAL among respondents of recruitment questionnaire, actually wearing the activPAL, and percentage of participants with valid data.

	respondent in questionnaire	consent to wearing activPAL	started wearing activPAL (subsample of respondents)	started wearing activPAL (subsample of respondents)	provided >= 4 days of valid data
	n	%	n	%	%
pilot 1	2276	48.7	48	29.2	26.2
pilot2a	2091	47.4	186	33.9	25.4
pilot2b	1881	55.8	190	40.5	33.1
pilot3	3370	50.5	201[1]	48.5[1]	44.0[1]

1: situation per April 1st, after five (independent) batches of fieldwork, about halfway through fieldwork

Note: the percentage participants of the subsample is calculated based on the percentage consent of the entire sample.

One relevant statistic is not visible in table 1, and that is the number of participants from the sample that received the invitation to fill in the questionnaire and the influence thereof on the final percentage of sample members who deliver valid data. Table 2 shows these findings and also make clear the importance of high response rates in this first step for the final number of participants who deliver valid data. The table also shows that the final number of participants, in view of the sample that was first invited is quite low, around 11%, in three of the pilots. In the LISS panel, this percentage is much higher: 44% of sample. Both the substantially higher initial questionnaire response and the large percentage of consenting respondents that went through with wearing the meter contributed to this finding. The LISS panel response is commonly this high, as a result of the carefully thought out fieldwork and incentive strategy and the long term loyalty of the panel members (see e.g., Scherpenzeel, 2014).

Table 2. Response to recruitment questionnaire, consent to wearing the activPAL, wearing activPAL, and providing valid data as a percentage of the sample.

	sample	participation to questionnaire	consent to wearing activPAL	started wearing activPAL	provided >= 4 days of valid data
	n	%	%	%	%
pilot 1	4997	45.5	22.2	13.3	11.9
pilot2a	6801	30.1	21.4	15.3	11.5
pilot2b	5711	32.9	17.7	12.8	10.5
pilot3	3874	87.0	43.2	48.5[1]	44.0[1]

1: situation per April 1st, after five (independent) batches of fieldwork, about halfway through fieldwork

The findings on participation of sample are reminiscent of the findings in the pilot by the National Institute for Public Health and the Environment (in Dutch: RIVM), that preceded the CBS pilots reported here. In the RIVM pilot, sample persons were contacted by letter with the stated purpose of filling in a questionnaire on physical activity ànd wearing an accelerometers (the UKK). There were six varying incentive conditions, with varying response and recruitment rates, but overall, the response rate to the questionnaire was 13.8%. 12.8% of the sample gave permission to send the meter, and 11.4% of the sample finally wore the UKK for the required number of days. See section 4.2.8. for more details on this study.

The common denominator in all these pilots, as well as the pilot by the RIVM, is that the recruitment of the respondents is either through mail, or within a web survey. It is to be envisaged that interviewers would do a much better job of securing participation. They could have the role of showing the measurement device, showing how to mount it, and showing the feedback material that the participant could receive. In 2023, CBS aims to perform a pilot with accelerometers in a (web-CAT-CAPI) mixed mode population survey, with a large role for interviewers.

4.4.2 Possession of private accelerometers and the use of apps for measuring physical activity.

In each of the Dutch pilots, respondents were asked after the possession and use of commercial accelerometers or apps on their smartphones. Despite the different circumstances and populations in the pilots, the findings are very similar. Around half of the respondents use one of those to monitor their physical activity (49% in pilot 1, 49.8% in pilot 2a, 46.9% in pilot 2b). In pilot 3 the precise rate is difficult to ascertain, as this was done in two steps, with a prior inventory in another survey. In the pilot proper, all persons with an activity tracker (but not those using an app) were invited, plus an additional number of persons without a tracker.

In pilot 1 it was ascertained that of those using an app or owning a tracking device, 16% owned a smartwatch, 8% an activity tracker, and 37% used an app (some respondents used more than one). There are differences between the owners of the different devices in the intensity of their use: smartwatch owners use their device daily to a larger extent than activity tracker owners, and they again to a larger extent than app users. See table 3.

Table 3. Intensity of use of physical activity by monitor device, pilot 1-NL

	use daily	use sometimes	total use	
	%	%	%	n
smartwatch	59	24	83	355
activity tracker	47	30	77	192
app	33	41	75	847

In pilot 3, this pattern was remarkably similar, with 51% of smartwatch owners using their device daily and 26% sometimes (77% in total) , against 35% and 50% respectively of tracker owners (85% in total). In pilot 2, the trackers were more specified, making the comparison difficult.

4.4.3 Willingness to copy data of own device into a questionnaire

Table 4 shows the percentage of people with an activity tracker who are willing to copy data in a questionnaire, and who are willing to upload data. For easy comparison with the number of people willing to wear the activPAL, the latter data are included in table 4 again. The table shows that the number of people willing to copy data is higher than the number that is willing to wear the activPAL, but the number that is willing to upload data is lower than both other values.

Table 4. Percentage of respondents consenting to wear the activPAL, willing to copy their own data and willing to upload data.

	respondent in questionnaire	consent to wearing activPAL	willing to copy own data	willing to upload own data
	n	%	%	%
pilot 1	2276	48.7	58.0[1]	40.6
pilot 3	3370	50.5	56.8[1]	34.2

1. of owners who indicated that they used their device or app sometimes or always to monitor their physical activity

In addition, all participants willing to copy data were asked during the recruitment questionnaire to copy their data from the previous day. Table 5 shows the number and percentage of people in pilot 1 who use a tracker, who copied data and who did not know how to copy the requested data.

Table 5. Number of people using a tracker, who copied data and who did not know how to copy data – pilot 1 NL.

Indicator	Uses tracker [a] that is able to measure this indicator	Copied yesterday's data	Did not know how to copy	Did not want to copy
Number of steps	732 (83.4%)	637 (87.0%)	29 (4.0%)	66 (9.0%)
Distance (traveled)	724 (82.5%)	565 (78.0%)	89 (12.3%)	70 (9.7%)
Active minutes	503 (57.3%)	350 (69.6%)	102 (20.3%)	51 (10.1%)
Calories burned	463 (52.7%)	326 (70.4%)	73 (15.8%)	64 (13.8%)
Heart rate	353 (40.2%)	236 (66.9%)	75 (21.2%)	42 (11.9%)
Sleep duration	244 (27.8%)	209 (85.7%)	16 (6.6%)	19 (7.8%)
Speed (current and/or mean)	278 (31.7%)	128 (46.0%)	119 (42.8%)	31 (11.2%)
Stairs climbed	229 (26.1%)	176 (76.9%)	31 (13.5%)	22 (9.6%)

[a] *A smartwatch, activity (fitness) tracker or phone that measures physical activity. N = 878*

note: table from Kraakman e.a. 2022

4.4.4 Willingness to donate data of the own device

If people would be willing to donate data, that would potentially very worthwhile. Giving people an activity tracker for a week runs the risk that people will behave unnaturally during that week, e.g., by moving more than they normally would. Asking people to copy data runs the risk that they might polish up their data. The data from pilot 1 (NL) gave ground to this suspicion, as the data turned out to be rounded to a large extent (Kraakman, Luiten & Toepoel, 2022). Donated data, on the other hand, gives us the opportunity to look back in time, to a period that people were not aware that they would be monitored, and will potentially give data that are closer to the true activity behaviour.

See table 4 in the previous section for the percentage of people willing to donate their data in the first and third Dutch pilot (the question was not asked in the second pilot).

In pilot 3, an experimental manipulation tried to influence people's decision by varying the information leading to the recruitment question. The variations were: downloading is very easy; downloading takes less time than copying the data from your meter into a questionnaire; the message that 'CenterData will of course ensure that no privacy-sensitive data are shared with other parties (after downloading the overview you can still determine which data you do and do not want to share)', and combinations of two and all three elements. See table 6 for the effect on willingness to donate data of these experimental manipulations.

Table 6. Percentage of respondents willing to donate data

	% willing	n
Downloading is easy	34	100
Downloading saves time	40	107
Privacy is guaranteed	32	95
Easy & time saving	34	93
Easy & privacy	34	106
Time saving & privacy	32	88
Easy & time saving & privacy	33	119
Overall	34	708

Note: $\chi^2_{(12)} = 7.78$, ns.

The manipulations did not have a significant effect on the number of respondents willing to download the data. The slightly higher effect of the manipulation 'downloading saves time' does resonate however in the reasons people give for not wanting to donate data: 'it takes too much time' is with 50% the most given reason. Privacy considerations are second (37% of refusers) and 'it seems too difficult' is the third with 21%.

In pilot 3, participants are asked to actually donate their data. The data collection of this pilot is still ongoing at the moment. A first impression of the results will be included in the overview deliverable of this WP.

4.4.5 Summary: feasibility of using people's own accelerometers

Sections 4.2.1 to 4.2.5 described people's possession of their own activity tracker, their willingness to copy data from those trackers into a questionnaire, and their willingness to donate data. About half of the sample in pilots 1 possessed an activity tracker. Half of them again were either willing to copy their data, and about one third is willing to donate data. However, in pilot 3 we noticed that not all activity trackers lend themselves equally well for data donation. For pilot 3 we had to focus on Fitbits, as those were the only ones that allowed easy access. Fitbits were possessed by 20% of respondents with an activity tracker. That means that in the end, only roughly 50% (possession) * 30% (willing) * 20% (Fitbit) = 3% of the respondents at the most will donate data. That is obviously not enough to base statistics on. Copying data seems to be a more fruitful track. However, the donated data will nevertheless give insight into the important question of respondents' behaviour change and respondents' accuracy in copying data. And, in the end, they may be one additional valuable data stream, next to copied data, questionnaire data, and activPAL data.

Apart from people's willingness to use their own data, there is the obvious question if the data that are collected by the commercial activity trackers are any good in capturing people's actual physical activity. In both pilots 1 and 3, we asked respondents what they thought about the accuracy of their trackers. In pilot 1, 12% of respondents found them to be very accurate, 53% rather accurate, 25% found them accurate for some, but not for all movements, and 9.4% found them inaccurate. In pilot 3, the accuracy was judged a bit higher, with 30% of respondents finding the measurement very accurate, 45 % fairly accurate, 19% sometime accurate, and 5% inaccurate. Reasons given for inaccuracy are not yet fully analysed, but the (non-)registration of bicycling as activity is mentioned often, random hand movements that are counted as physical activity, and also recognition of 'vertical' activity, like climbing stairs.

The topic of the use of people's own devices is also picked up in academic research on physical activity, see e.g., Shcherbina et al., (2017) who compared the commercial Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2 with a benchmark heart rate and energy expenditure measurement. They concluded that heart rate is generally adequately measured, but energy expenditure has a large overall error. The accuracy of the commercial devices varies with sex, body mass, skin tone, and kind of activity.

One aspect that is hitherto not discussed, is the bias that may be introduced when using people's own tracking devices. Kraakman, Luiten & Toepoel (2022), show for pilot 1 that possession of smartphones and other activity trackers is related to age, education, and income, with younger people, people with higher education and higher income being overrepresented. In addition, not surprisingly, people who monitor their physical activity adhere to physical activity guidelines to a (far) larger extent than those who do not. Especially people with a smartwatch or activity tracker comply more often to the guidelines, see table 7 for results from pilot 1. For pilot 3, this analysis is not yet performed.

Table 7. Compliance to physical activity guidelines by use of monitoring device

%	no tracking	smartwatch or other device	phone	ANOVA
complies to complete guideline	45.6_a	63.9_b	57.8_c	F(2,2230)=27.6***
complies to number of minutes	70.5_a	81.1_b	77.0_b	F(2,2230)=11.5***
complies to bone and muscle guidelines	66.6_a	79.5_b	74.7_b	F(2,2230)=15.9***
complies to balance guideline	19.3_a	25.9_b	21.3_{ab}	F(2,2230)=4.5*
n	1136	482	616	

Note: table adapted from Toepoel, Luiten and Zandvliet (2021).

4.4.5 The effect of the promise of feedback

In pilot 1 (NL), an experimental variation on promising feedback was included. Half of the respondents to the questionnaire was promised feedback on their physical behaviour pattern, and the other half was not.

Promising feedback did nothing for the participation rate: when not promised feedback, 49.5% of respondents was willing to participate, versus 48% when promised feedback ($\chi^2_{(1)}$ = .53, ns). This pattern of slightly lower, (though non-significant) consent was consisted for various subgroups: with or without their own meter, adhering or not adhering to the physical activity guidelines, the four age groups distinguished, and for men. For women, however, the difference is significant: 54% consent *without* feedback, versus 48% with feedback ($\chi^2_{(1)} = 4.03$, $p < .05$).

In pilot 2a and 2b (NL), participants were asked if they would like to receive an overview of their exercise behaviour. Although the feedback is apparently not sufficient to seduce people to participate, those who do participate are curious to receive it: 98% both pilots. The targeted text did not have the expected effect of increasing curiosity about the outcomes in any of the four groups (activity x region).

All participants in the Dutch, Polish and Belgium pilots received feedback on their physical activity. The participant received one of the overviews that can be generated by the activPAL software, plus an explanation of what the overview means. See appendix 2 for an example of both. The generated overview was personal, the explanation was the same for all participants.

4.4.6 Incentives

One of the goals in the project was finding out if a monetary incentive is needed, and to which amount. In the Netherlands, this was addressed by asking respondents in pilot 1 what they deemed to be a sufficient amount for the task, and by an experimental variation in pilot 2. In Poland and Belgium the topic of monetary incentives was not discussed, but incentives in the form of feedback were part of the evaluation questionnaires that were sent to the participants in both countries. In the Netherlands this latter topic was also addressed experimentally in the first pilot, see the previous section.

In the tree Dutch field pilots described here, an incentives was promised for wearing the activPAL for a week: €20 in pilot 1 and €15 in pilot 3. In all pilots, the incentive was introduced at the moment of asking consent for the field study. In pilot 2 an incentive experiment was envisaged, comparing €12,50 with €25. As described earlier, through a mistake in the operationalisation, this experimental condition was collapsed with another experimental condition with a targeted invitation text. This made it impossible to distinguish the effect of both measures. However, it was telling that in three of the four groups distinguished in this pilot, no difference was found between the group with the high incentive / targeted invitation and the low incentive / neutral invitation. Table 8 shows the results.

Table 8: Willingness to wear the activPAL in active and non-active respondents by incentive level and targeted invitation.

	Active		Non-active		Active		Non-active	
	%	n	%	n	%	n	%	n
Neutral/low	69.7	617	67.7	450	56.8	482	44.8	451
Targeted/high	72.4	554	68.4	470	59.7	474	53.2	474
	ns		ns		ns		$X^2_{(1)} = 6.49$ $p < .01$ (1-sided)	

Unless the targeted invitation had an opposite effect of the one we envisaged (i.e., decreasing the response rates and hence working against the incentive), it can be cautiously concluded that the higher promised incentive generally did not lead to higher recruitment rates. Further analysis is necessary to determine why they did have the desired effect in the non-active group in pilot 2b. If there is generally no difference between an incentive of €12,50 and one of €25, it makes one wonder if even lower incentives will also result in acceptable recruitment rates.

As described in section 4.1, a pilot with accelerometers was performed by the Dutch National Institute for Public Health and the Environment (RIVM) in collaboration with CBS (de Hollander, Toepoel, Duijvestijn, Wendel-Vos, Luiten, Schouten and Loyen, 2022). In this pilot, an incentive experiment was included, with the following conditions: promise of a €10, €20 or €40 gift voucher, and unconditional gift of a €5 gift voucher plus the promise of a €10 voucher[1]. Table 9 shows results.

Table 9. Percentage of response, recruitment, dropout and participation by incentive

Incentive	n	Response questionnaire	recruitment activPAL	Dropout	Participation >= 4 days
€ 10	1000	12.6	11.3	0.9	10.4
€ 20	2000	16.1	15.2	1.4	13.8
€ 40	1000	19.5	18.5	1.5	17.6
€5+€10	1000	17.1	16.1	2.4	13.7

Note: adapted from Toepoel, et al., (2021).

Not surprisingly, the highest incentive shows the highest response rate and the highest participation rate. The differences, however, are not large. Additional analysis needs to show if the slightly higher participation rates translate to higher quality and if this additional quality is worth the additional costs of a higher incentive.

One last source of information on incentives in this project is the question we asked respondents in the pilot 1 questionnaire under what conditions they would be willing to participate. We will describe the full findings in the next section, but some of the reasons given pertained to incentives. Of all the

[1] Additional conditions were tried also: another gift voucher worth €10, and a condition where a promised €10 voucher was combined with a promise of feedback. In the latter group, however, oversampling took place of some groups with lower response rates; the results are therefore difficult to compare with those in table 8. The vouchers in table 8 are all of the same kind: a popular gift voucher in the Netherlands.

circumstances that would contribute to participation, a higher incentive was mentioned in 13% of cases. However, the amounts mentioned by people were quite steep indeed and higher than would be feasible in practice. All are above €40, while the amount mentioned most is €80 (which was the highest amount suggested in the closed answering categories). See table 10 for the distribution over given amounts.

Table 10. Incentive values mentioned by respondents as condition for participating.

€	n
40	5
50	3
60	13
80	30
100	6
150	3
>=200	6

In Poland, the matter of monetary incentives was addressed in the evaluation questionnaire. During the pilot, no incentives other than an overview of physical activity were given nor promised. Half of the Polish participants indicated that they would appreciate a monetary incentive in future occasions, while 29% indicated that the physical activity report was sufficient, and 21% found that no incentive at all was necessary. Nevertheless, 68% of respondents indicated that they would participate in future research, even without an incentive.

All Belgian participants indicated that it is important to them that they receive feedback on their results and how they relate to an average. All but two of them reported that the feedback received from CBS regarding their movement profile met their expectations:

"I would have liked to see a comparison with the other participants, but on the other hand the feedback I received exceeds my expectations (e.g. the movement score MET.h)."

This opens perspectives for using this kind of feedback as an incentive to persuade people to participate in future studies on health and well-being instead of a monetary incentive.

Although more experiments would be welcome, a tentative conclusion can be reached that for pilots where people are asked to wear an accelerometer for a week, other incentive mechanisms may be at work than are seen in the survey literature on incentives (e.g., Singer & Ye, 2013). When offering a (monetary) incentive for surveys, unconditional incentives in surveys without interviewers (web or mail surveys) generally result in about 15 percentage points higher response rates above situations where no incentive is offered. And when offering conditional incentives, higher incentives generally result in higher response rates. We have seen neither mechanism convincingly at work here. Small differences were seen when higher incentives were offered, but it is as yes unclear if the higher response rates warrant the extra costs. When asked what amount of incentive would induce unwilling people, the amounts mentioned are so high that they are not feasible in practice.

Also the promise of feedback on physical activity does nothing to win people over. On the other hand, the persons who do participate are interested in the feedback they receive and expect to receive it. It would appear then, that people who participate seem intrinsically motivated by physical movement.

A hypothesis that is supported by findings that people who participate are generally more active than people who do not (Toepoel, Luiten & Zandvoort, 2021; Kraakman, Luiten & Toepoel, 2022).

This does however imply that we need to find other means to convince people to participate. In how far it also implies that we do not have to offer incentives is not clear. We have offered monetary incentives standardly in all three Dutch pilots among the general public. In the Polish and Belgian pilots no incentives were offered, but these were convenience samples among colleagues, who for a large part did it to humour their colleague, see appendix 5.

4.4.7 Barriers to participation

In none of the four pilots a direct question was asked why people were not willing to wear an activPAL. However, some indirect information is available about what would convince people. In pilot 1 a question was asked what would convince unwilling people who to change their mind. Table 11 shows the results. It is notable that most of these reasons can be addressed by the survey organization. Apparently, before asking the recruitment question we had not made it sufficiently clear what the purpose of the question and the study was, that the research would be completely anonymous, that it would not be burdensome, etc. This is also tricky: you don't want people to have to read a very long introduction to the recruitment question; many people will skip or gloss over such an introduction to a question. In pilot 3, an elegant solution to this dilemma was found: people were given 3 response options after the recruitment question: 'yes', 'no', and 'I am in doubt, please give me more information'. This gives the opportunity to provide more explanation, without burdening respondents who do not want this information. And it works: of the 165 people who answered 'I doubt', 125 gave their consent in the second instance.

Table 11. Circumstances to change nonparticipants to participants.

	n mentioned	%
if totally convinced of purpose	159	22
if totally anonimous	131	18
if higher incentive	69	9
if given advice on physical activity	68	9
if given insight in physical activity	51	7
if chance of high lottery incentive	27	4
other, of which:	230	31
if not burdensome	29	4
if I don't have to do anything else	27	4
if I would move like I normally do (now corona, ill, at home…)	27	4
am too old, move too little	24	3
if I can combine with work	16	2
other	78	11

In Pilot 3, a similar question was asked of people who indicated that they were not willing to copy data from their own tracker into a questionnaire. The mechanism there is slightly different: Most people indicated that this would cost them too much time. Privacy was also mentioned relatively often, which was never used as an argument for not wanting to wear the accelerometer.

Table 12. Reasons not to copy data from own tracker into questionnaire

Reasons	n
Takes too much time	150
Privacy	110
Too difficult	61
Other, of which:	28
Don't wear it all the time	11
not interested / don't feel like it	4
Incentive needed	2
Don't move enough (at the moment)	4
Other	5

Quite a few of the conditions that people mention prior to considering participating could be addressed in the recruitment process: more information could be given on the purpose and the way anonymity is guaranteed, for example. Toepoel, Luiten and Zandvliet (2021) show, however, that although this could increase response rates, this would do nothing for the bias in the data. Even the people who hesitate are far more active than the people who unequivocally say that nothing can convince them.

4.4.8 Gauging participants' compliance with instructions

In order to collect data of high quality, the measurement devices need to measure uninterrupted, on the intended body part, in the designated direction, during the designated amount of time. Respondents need to comply with the instructions, tolerate wearing a device that may be intrusive and may invade their privacy to a potentially higher extent than regular survey questions.

There are two sources of information for this question: participants' behaviour (did they mount the activPAL according to our specifications, how long did they wear the meter) and the evaluation questionnaire that was asked at the end of the pilots.

Participant behaviour

Table 1 already showed how many people started wearing the activPAL and how many people stopped before at least 4 (complete 24 hours) days of wear. Many wore the activPALs at least part of the day. activPAL only counts a day where the device is worn for 24 hours as a 'valid' day. That means that if people take the device off to go swimming for example, or to take a bath, the day is no longer counted as valid. In the first part of the 2nd Dutch pilot for example, 76% of participants had at least four valid days, but 89% wore the activPAL at least part of four days. In the second part these numbers were 82% valid days and 91% part of days. Especially the final, 7th day is often partly measured. We need to fine-tune the protocol in order to increase the number of useful measures of potentially willing participants. And additionally, we need to study the data in order to determine which days are validly useful for analysis. The more days we measure, the higher the reliability of the measures.

Another relevant measure is the number of times the participants wore the activPAL back to front. This is not a problem when using the activPAL software, as the software corrects for this, but is something that needs to be taken into account if proprietary algorithms are developed. The occurrence of this phenomenon is limited: no participants in pilot1-NL, 1 participant in pilot 2a-NL and 7 participants in pilot 2b-NL wore the activPAL the other way around. In Belgium, three participants wore the activPAL the other way round on the first day, but corrected in the second or third day. In

Poland only one person wore it upside down but corrected it on the second day. The rest of participants wore the activPAL correctly all the time.

Finally, the alignment of the device is important. The participant is instructed to wear the device upright, see appendix 1. The alignment is important for the right calculation of the gravity on the three axes. Most participants, but certainly not all, wear the meter in the required way or at most 10° off (75% in pilot1-NL, 81% in pilot 2a-NL, 74% in pilot2b-NL, 100% in Belgium, and 100% in Poland). Again, this is something that needs to be taken into account when analysing the data.

Evaluation questionnaire.

In pilot1-NL participants were asked how they evaluated wearing the activPAL in an open question. Regretfully, only 26 participants answered the question. Van Hoek et al., (2022) note that '*Only one participant was negative, finding the adhesive tape unpleasant and itchy, and the manual to mount the device unclear. 18 participants were unequivocally positive ('you never even notice that you are wearing it'), while 7 participants were positive, but with a qualification that some aspects were less than optimal (removing the adhesive plaster from a hairy leg, and also itchiness and skin irritation). Part of the skin irritation was not caused by the adhesive plaster, but by the latex finger condom that was used to waterproof the activPAL. This can be remedied by using other materials. These findings on respondent evaluation corroborate other research, (e.g., Berendsen et al., 2014; Edwardson et al., 2017)'.* Irritation as a result of hairy legs can be alleviated by the instruction to shave that part of the leg prior to mounting the activPAL

The Belgian participants agreed that the instructions for attaching the sensor (see Appendix 1) were clear. Nevertheless, 7 out of 10 participants were of the opinion that a short instruction video explaining step-by-step how to attach the sensor would be more efficient. Most participants were able to attach the sensor without any problems. Two participants mentioned some difficulties:

"I first applied the sensor upside down." Participant BE11

"It was a bit of searching how to use the patches." Participant BE01

Somewhat worrying is that 3 participants had physical discomfort while wearing the sensor:

"After a few days, a small cut on my leg at the level of the edge of the sensor. I repositioned the sensor." Participant BE18

"Slight itching, removing the patch was a little painful." Participant BE01

"Red rash and itching after about 5 days." Participant BE04

The Belgian participants were aware of the sensor and the fact that their movements were being recorded. Most of them agreed that after some time they became less aware of the sensor; only 2 participants disagreed. One respondent claimed that wearing the sensor affected her usual movement behaviour, but she did not specify in what way her movement behaviour had been influenced.

The Belgian participants mentioned that it is important to receive information about the proper functioning of the sensor (e.g. in the form of a green/red light). The opinions differed on the necessity of a pause button, the importance of one's own validation of the observations and the possibility to delete certain observations or observation periods. Some participants clearly want to maintain control, even in passive measurements. The participants also differed in their opinions regarding the importance of providing context to the observations. Only half thought this was an important issue. However, 8 out of 10 participants indicated that they would be willing to fill in an online questionnaire or diary on a daily basis during the measurement period regarding the conditions under which the activities are undertaken.

All Belgian participants except one found it an interesting experiment. They believe that it is feasible in the long run to use these kinds of technologies to collect data – especially as a supplement to existing data collection methods within Statistics Belgium. However, two participants suspect that many people will refuse to participate in fieldwork organised on the basis of such passive measurements:

"New mode with which one is unfamiliar and may feel less in control than if one were to give one's own written or oral answer." Participant BE02

"Sensitivity about privacy; government seen as "big brother who needs to know everything"; collaborations with organizations like 'Ministry of Privacy' seem appropriate, to get ahead of any criticism." Participant BE09

To conclude, Statistics Belgium conducted a very modest pilot study that can be evaluated positively in the sense that nothing has emerged to discourage large-scale experimentation.

Also in the pilot conducted in Poland, participants were asked to complete a questionnaire after completing the ActivPAL device activity survey. Of those who had worn the device for at least 1 day 100% responded to the questions asked. When asked about wearing the device, 33% of participants responded that the device made them uncomfortable due to itching and slight allergies or visibility under clothing. This did not affect interest in another study using the devices, 67% of participants said they would be willing to take part in a similar study. However, they stated that the study duration of 14 days was a bit too long. Participants did not have many comments about the survey except for its winter timing which resulted in limited activity due to bad weather.

4.4.9 Privacy

NL did not mention anything explicitly about privacy issues in the pilots. One of the participants withdrew after a few days however, explicitly for privacy reasons. Other respondents were aware of possible privacy issues, and solved them by removing the accelerometer (mentioned in the phase 0 pilot diary as reason to remove the meter). An informed consent form was signed by the participants. In quantitative pilots 1 to 3, standard informed consent practices of the research organisations were used. In order to be able to share the data from pilot 1 with ISTAT, participants were contacted again to ask consent for the unforeseen data sharing.

In Poland, respondents were firstly informed beforehand about all the privacy and statistical confidentiality issues, namely what kind of data would be collected by the accelerometers. The information stated that the data from the accelerometers would be analysed after being aggregated. The accelerometer ID and the ID of the person assigned to it was separated from the data collected by the device. The only exception to that rule was when a person wanted to receive the results from the survey and therefore signed a consent to de-anonymize their data

The Belgian participants were asked about their perception of privacy and attitudes. Almost all participants found it no problem that the NSI was able to access rather privacy-sensitive activity aspects over which they had no control. Not surprisingly, since they work in a NSI and know about the strict legal regulations regarding privacy. At the same time, there were some major concerns among the participants about collaborations with commercial companies – which in the case of passive measurements are often done in order to keep the price at a reasonable level. So, before large-scale studies using passive measurements can be put into the field effectively, it is necessary to think about the privacy and confidentiality issues and discuss the type of consent – informed consent vs. proxy consent vs. implied consent – that is needed from a legal, ethical and practical point of view.

4.4.10 Fieldwork and logistics

The fieldwork for the large second pilot in the Netherlands was performed by a commercial external partner. The fieldwork was evaluated at the end for the clients (the regional Health Centres and CBS). These are their findings:

- There was a lot of interest in participating. For the first round of invitations, there was a response rate of 70%. People were open to suggest a measurement week and moving it when it was not feasible for us. It is possible that the corona pandemic and its associated measures contributed to the interest.
- One week between measurement weeks is on the short side to receive everything, download the data, charge the activPAL anew and prepare for another sending. Reading out the meters takes a lot of time. The field workers only just made it in time each time to have the next batch ready. However, it is not certain whether an extra week between the measuring weeks would be better, in connection with the battery life of the meters. A possible solution would be to have more spare meters.
- The data collection is highly dependent on the mail delivery. This was not always on time. Possible solution: use standard 24-hour mail, but then take into account the additional costs.
- Thinking about what start-day is convenient. In this pilot Sunday evening was used as starting time, which is difficult because no mail is delivered on Sunday and Monday. Possibly Saturday evening would be better (also beneficial that even with too short a measurement, you then have a weekend day anyway).
- Properly working exercise meters are very important. In the first batch, measurement failed completely for some participants, and they needed to be contacted if they could do it again.
- The manual was perceived as unclear by some participants in the first round. It concerned one photo where it was unclear to participants how the meter was mounted. By adding extra explanations in the manual this was solved.
- A small number of participants got a slight rash from the protective cover around the meter. These are made of latex. Possibly next time use protective covers of a different material.
- The return envelopes used for the first batches were not strong enough to send the accelerometers over the mail. Some envelopes tore in the mail, and some were torn on purpose to find out what was inside. Halfway through the project a switch was made to reinforced envelopes to prevent tearing and curious eyes.

In the other pilots we also lost meters because the envelop was torn open to reveal the contents if the envelop was not thick enough. We considered sending the activPALs insured. However, the costs of sending all activPALs insured (to and fro) would be higher than the loss of some. In total, in the four Dutch pilots, the two Polish pilots and the Belgium pilot, with a total of almost 1000 activPALs, 18 activPALs were lost in the mail or not returned, i.e., 1.8% of devices. In the RIVM pilot (Hollander et al., 2022) this percentage was 4.3%. That means that loss of measurement devices is something that needs to be incorporated in the planning and costs of fieldwork the pilot. In 2023, CBS aims to experiment with the involvement of interviewers to deliver the meters to the respondent, instead of sending them through the mail, and perhaps also in collecting them again. Even though the employment of interviewers for this purpose can be costly, these costs need to be carefully balanced against the costs of losing the meters.

5. Phase 2': Machine learning and statistics

Statistics Netherlands produces a number of statistics based on the SQUASH questionnaire: the percentage of persons who adhere to the physical activity guidelines, who do sports at least once a week, the number of minutes per day that people sit, the frequency, length and intensity of physical activity, the kinds of physical activity and the kinds of sports. And these for the total population, for various subgroups and for groups based on other indicators of life style and health.

It is by no means an easy task to determine these variables by means of accelerometers, and it is a task that we only started tackling. In phase 0, we succeeded in training machine learning algorithms to recognize laying, sitting, standing, walking, running and bicycling, the activities that were performed in the lab. See van Hoek, Windmeijer, Luiten, Bolte & Schouten (2022) for a detailed account of the development of these algorithms. However, the algorithms did not generalize well to the activities in the free living week. Although most of people's movements are made up of combinations of the trained activities, the free living activities proved to be far more complicated. Part of the problem was that the physical activity diaries that could have informed the machine learning algorithms, were on the one hand very imprecisely filled in, and on the other hand not detailed enough. For the phase2 pilots, we instead made use of the activPAL software, that likewise distinguishes these activities, and in addition adds a value for energy expenditure (the MET value). Whether people do sports, however, and especially what kind of sports, we have not yet been able to determine with the accelerometer. There is however, an international research collaboration platform (ProPASS) for which the development of algorithms for thigh-worn accelerometers is one of the core targets. The objectives of the consortium are:

- to establish a pooled data resource on physical activity, posture allocation, sleep, and health outcomes;
- to develop methods for processing, harmonising, and pooling data of existing such studies;
- to develop methods for collecting data for future studies (prospective harmonization);

The ultimate goal of the ProPASS consortium as a data resource is to promote collaborative individual participant and prospective meta-analyses on physical activity, posture, and sleep (propassconsortium.org). Another open source initiative is described in Migueles et al., (2019). We aim to link to these developments. Hoek et al., (2022) also suggest to replicate the Phase 0 pilot, but now specifically aimed at the development of an algorithms for the free living week. They suggest to have participants wear the accelerometer, while keeping a time use survey, preferably using the time use app under development.

See also Deliverable 2.8 – physical activity for more information on the machine learning aspect and the WP3 deliverable 'Treatment of sensor data for Smart Surveys through Machine Learning: a Generalized Module' (Cerasti et al., 2022) on how the activPAL data and the analysis algorithms can be integrated in a generalized module.

6. Summary and discussion

This deliverable describes the search, in Phase 0, for a sensor system that would be able to reliably measure physical activity, against reasonable costs, that respondents would accept to wear, that could be feasibly handled in terms of logistics, and that would be secure in terms of data access. We have found such a device in the activPAL, a small thigh worn accelerometer. The activPAL can be sent to the participant with an instruction, but can also be handed over by an interviewer. The latter solution has not been tried in the pilots described here, but will be piloted by CBS in research that is envisaged in the second half of 2022 and in 2023. activPAL data are stored on the device and downloaded once the participant sends the device back (or once they are retrieved by interviewers). With the activPAL, the pilots in Phase 2 were performed, plus some other pilots that were performed during the time period of this ESSnet, but not specifically for the ESSnet. Relevant findings from these extra pilots are reported here as well, for the aspects that reflect on this ESSnet.

Small scale usability pilots among convenience samples were performed in Poland and Belgium, and larger quantitative pilots in the Netherlands, among probability and non-probability samples. The purpose of these pilots was mainly to shed light on usability and feasibility aspects of making statistics on physical activity with objective measurements. The actual derivation of physical activity statistics is still in the future. Van Hoek, Windmeijer, Luiten, Bolte and Schouten (2022) report on endeavours to develop machine learning algorithms to recognize physical activity during the week. They had to conclude that the data from Phase 0 were not sufficient to develop proprietary algorithms. In the Phase 2 pilots, physical activity was determined using the activPAL algorithms.

The quantitative pilots shed light on the question to what extent sample persons are willing to wear an accelerometer and have their physical activity measured in this way. In the three Dutch pilots that are reported, the sample person was sent an invitation to fill in a web questionnaire. In all three cases, the email address of the sample persons was known, and the invitation was sent by email. After the questionnaire was filled in, a recruitment question was posed into the willingness to wear an accelerometer. In spite of the different populations in the pilots, the willingness of the respondents to wear the meter was remarkably similar: around 50% of the respondents was initially willing. Not all of them actually wore one in the end, and not all of the people wearing them wore them for the required number of days. In the end, about a quarter to one third of respondents to the questionnaire delivered valid data; counted back to the original sample, this amounted to about 12% of the sample. The fieldwork of the third pilot is still going on at the time of writing, and we don't have the final figures for that one yet; however, from the fieldwork thus far we have seen that both willingness from respondents and willingness from sample is substantially higher in this pilot. We slowly learn how to increase willingness.

In all steps, invitation to the questionnaire, invitation to wear the meter, actual participation, and participation for the required number of days, important improvements can be made to increase the participation rates. These can be simple interventions: for example, showing a picture of the device in the recruitment question increased willingness with 9 percentage points. Also, offering the answering option 'I don't know yet, give me more information', and subsequently giving relevant information, won over most of the people who chose this answering option (5% of questionnaire respondents).

In all of the qualitative field pilots, incentives were offered as a matter of course, albeit of different values. These different values do not seem to reflect in the percentage of willing persons, although this is only experimentally varied in one of the pilots. Other experiments are necessary to see if and how much incentives are necessary; would nobody be willing anymore without incentive, or would that only be the most fanatic sports addicts? A tentative conclusion is formulated that other incentive mechanisms may be at work than are seen in the survey literature on incentives (e.g., Singer & Ye, 2013). The promise of feedback on physical activity, experimentally offered in one of the pilots, did nothing to win people over. On the other hand, the persons who do participate are interested in the feedback they receive and expect to receive it, as witnessed by the answers in the qualitative pilots.

The potentially most influential intervention has not been tried yet however: involving interviewers. It is to be expected that interviewers, especially field interviewers, can play a decisive role in persuading persons to participate, and also in keeping them participating, e.g., by calling them during the measurement period. Especially important is that interviewers will have better access to persons who are prone to move less. Preliminary analyses have shown that there is a highly significant positive correlation between participating in the web survey and physical activity. Reaching that part of the population with field interviewers is therefore crucial for minimising bias, both in sample composition and substantive variables. Statistics Netherlands is planning a mixed mode pilot with interviewers for 2023.

One of the questions that we posed is on the potential of using respondents' own accelerometers. As about half of the population possess or use such a device or an app on their smartphone, this would potentially be a valuable source of information. In two of the Dutch pilots, respondents with a Fitbit or Apple watch or similar devices were asked to either copy their data into a short questionnaire at the end of the week that they also wore an activPAL. In the Dutch pilot 3, still in the field, we also asked participants to upload data from their Fitbit unto the research institute's portal. Preliminary findings show that willingness to copy data is higher than willingness to wear the activPAL, and willingness to upload data is least of all. Nevertheless, if it were possible to use these data in a mixed device set-up, this would potentially save effort to send activPALs to people who already possess the data that we are looking for. More analysis is needed however into the feasibility. The results of pilot 3 will shed light on this.

One thing we already found, not surprisingly, is that people who possess their own fitness tracker are substantially more active than people who do not. The activity tracker data cannot thus be the only source of objective information on physical activity of the population.

Both from the analyses of the data coming from the activPAL, and the evaluation questionnaires, it appeared that adherence to the instructions was no problem for the participants, although some tweaking of the instructions will make adherence even more accurate and easy. A YouTube film exists that can be dubbed with other languages. Some discomfort is however mentioned, either as a result of the patch, or of the rubber casing of the activPAL. Especially in the Polish pilot, where people wore the activPAL for two weeks, the mention of itchiness was important. Itchiness as a result of the rubber can be relieved by using other materials. Itchiness as a result of the patch is more difficult to circumvent. Irritation as a result of removing the patch can be alleviated by advising people to shave their leg beforehand. Nevertheless, the results indicate that a large majority of participants wore the activPAL for the required number of 7 days.

An important aspect of continuous objective measurement is the issue of privacy of the data. In the Phase 0 pilot, one participant withdrew from the pilot once they realised that certain private activities would be visible for the researchers, would they focus in detail on their activity profile. Other participants realised the possibility, and dealt with it by removing the meter for certain periods of time. Once people remove the meter, there is a large risk that they forget to put it on again. In how far participants in the larger pilots also were aware of this possibility, and acted upon is remains to be seen in more detailed data analysis. The difference between the small qualitative and large quantitative pilots is obviously that in the first case the participants were known to the researcher, while in the second case they are not, making the data completely anonymous. Participants in pilot 1 in the Netherlands mentioned privacy as one of the most important barriers for participation, mentioning that they would only participate if they were totally convinced of the privacy of their data. The conclusion of the Belgian pilot was that before large-scale studies using passive measurements can be put into the field, we need to discuss the type of consent – informed consent vs. proxy consent vs. implied consent – that is needed from a legal, ethical and practical point of view. Privacy in the sense of data security was a topic in deliverable 2.8 and will not be further discussed here.

The further major topic in this work package concerned the logistics of doing fieldwork. The experience taught that this is not to be underestimated. With the present approach, an activPAL needs to be charged, given an ID, and be sent to the participant, together with the instruction, several medicinal patches and perhaps the rubber casing. In the pilots in Phase 2 we chose to put the casing on ourselves. Also included in the package is a sturdy return envelope. We have found to our chagrin that if the envelope is not sturdy enough, it will either tear in the mail, or otherwise be torn by curious persons interested in the contents. We have regretfully lost several activPALs in this way. When the activPAL is returned, the data need to be downloaded. In all pilots, even the large quantitative ones, the participants received feedback on their physical activity that needed to be individually generated. In was found that it is wise to calculate three weeks of turnaround time for one measurement period of one week. However, there may be possibilities to alleviate part of the logistic difficulties. Lately, we found that new thigh-mounted measurement devices have been developed that do not need to be stuck on the leg with medicinal patches, and that do not have to be sent back, as data are transferred to a cloud environment (see SENS motion® https://sens.dk). We suggest doing some research into this possibility. If relevant findings are known before the end of the ESSnet, these findings will be incorporated in the final deliverable.

Like the Belgian participants, the researchers are of the opinion that it is feasible to use this technology to collect data for population surveillance. Whether or not instead of, or as a supplement to, existing data collection methods.

The final, and most important question remains to be answered though. We have shown that it is feasible to use this technology, although much remains to be fine-tuned. However, in the end the most important question is of the added value of objective measurement compared to questionnaire data. Only the data of the 1st Dutch pilot are analysed in terms of substantive contents, i.e., the physical activity that is measured by the activPAL compared to the subjective SQUASH questionnaire and the physical activity measures that are generated by people's own activity trackers. Those data confirmed that the subjective data showed far more physical activity than the objective measurement, both by the activPAL and people's own trackers (Kraakman et al., 2022). However, the sample of this study was very small. The data of the 2nd Dutch pilot are still being analysed, and the 3rd Dutch pilot is still in the field. The answer to this important question will become available in the near future though. Interested readers are invited to contact one of the first three authors (a.luiten@cbs.nl; v.toepoel@cbs.nl; jg.schouten@cbs.nl) to remain informed of future developments.

7. References

Berendsen, B. A., Hendriks, M. R., Meijer, K., Plasqui, G., Schaper, N. C., & Savelberg, H. H. (2014). Which activity monitor to use? Validity, reproducibility and user friendliness of three activity monitors. *BMC Public Health*, *14*(1), 749. https://doi.org/10.1186/1471-2458-14-749

Bikker, R., Heuvel van, G., Hoek van, S., Hoogland, J., Lodder, B., Windmeijer, D., Luiten, A., & Schouten, B. (2020). *The accelerometer project* [unpublished CBS manuscript]

Brondeel, R. (2016). *The relevance of transport to promote physical activity: Addressing challenges related to the measurements and the observational analysis of transport-related physical activity, and the simulation of shifts in transportation mode*. PhD thesis, Université Pierre et Marie Curie - Paris VI.

Cerasti, E., M. De Cubellis, F. De Fausti, A. Guandalini, F. Inglese, A. Pappagallo, F. Pugliese, M. D. Terribili (2022). POC on methodologies: Treatment of sensor data for Smart Surveys through Machine Learning: a Generalized Module. https://webgate.ec.europa.eu/fpfis/wikis/pages/viewpage.action?spaceKey=EstatBigData&title=ESSnet+Smart+Surveys+2020-2021

Crowley, P., Skotte, J., Stamatakis, E., Hamer, M., Aadahl, M., Stevens, M. L., Rangul, V., Mork, P. J., & Holtermann, A. (2019). Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: A proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 65. https://doi.org/10.1186/s12966-019-0835-0

Edwardson, C. L., Winkler, E. A. H., Bodicoat, D. H., Yates, T., Davies, M. J., Dunstan, D. W., & Healy, G. N. (2017). Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science*, *6*(2), 162–178. https://doi.org/10.1016/j.jshs.2016.02.002

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 42. https://doi.org/10.1186/s12966-015-0201-9

Gorzelitz, J., Farber, C., Gangnon, R., & Cadmus-Bertram, L. (2020). Accuracy of Wearable Trackers for Measuring Moderate- to Vigorous-Intensity Physical Activity: A Systematic Review and Meta-Analysis. *Journal for the Measurement of Physical Behaviour*, *3*(4), 346–357. https://doi.org/10.1123/jmpb.2019-0072

Hoek, S. van, Windmeijer, D., Luiten, A., Bolte, J., & Schouten, B. (2022). Comparing activity trackers to investigate physical activity. CBS working paper, accessible at https://www.cbs.nl/en-gb/background/2022/08/comparing-activity-trackers-to-investigate-physical-activity

Hollander, E. de, Toepoel, V., Duijvestijn, M., Wendel-Vos, W., Luiten, A., Schouten, B., and Loyen, A. (2022). Monitoring Physical Activity and Sedentary Behaviour in the General Population: a feasibility study into the Use of Accelerometers. Under review PLOS ONE.

Júdice, P. B., Santos, D. A., Hamilton, M. T., Sardinha, L. B., & Silva, A. M. (2015). Validity of GT3X and Actiheart to estimate sedentary time and breaks using ActivPAL as the reference in free-living conditions. *Gait & Posture*, *41*(4), 917–922. https://doi.org/10.1016/j.gaitpost.2015.03.326

Kraakman, R. (2021). The validity of consumer-level activity trackers in research on physical activity: A comparison with self-report and a professional accelerometer on representation and measurement (Master's thesis). Utrecht University.

Kraakman, R., Luiten, A., & Toepoel, V. (2022). The Validity of Consumer-Level Activity Trackers in Research on Physical Activity: a Comparison with Self-Report and a Professional Accelerometer on Representation and Measurement. Under review *Survey Research Methods.*

Migueles, J.H., Rowlands, A.V., Huber, F., Sabia, S., & van Hees, V.T. (2019). GGIR: A research community-driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *Journal for the Measurement of Physical Behaviour*, 2, 188-196. https://doi.org/10.1123/jmpb.2018-0063

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Wheeler, M. T., & Ashley, E. A. (2017). *Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort*. 12.

Scherpenzeel, A. (2014). Survey participation in a probability-based internet panel in the Netherlands. In Engel, U., Jann, B., Lynn, P., Scherpenzeel, A., & Sturgis, P. (Eds.). (2014). *Improving Survey Methods: Lessons from Recent Research* (1st ed.). Routledge.

Singer, E., and Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, 645¸ 112-141.

Tarp, J., Andersen, L., & Østergaard, L. (2015). Quantification of underestimation of physical activity during cycling to school when using accelerometry. Journal of Physical Activity and Health, 12, 701–707.

Toepoel, V., Lugtig, P., Loyen, A., Duijvestijn, M., Luiten, A., & de Hollander, E. (2022). Monitoring Physical Activity in the General Population with Accelerometers: a Nonresponse Analysis on the Use of Incentives. Under review Journal of Physical Activity and Health.

Toepoel, V., Luiten, A., & Zandvliet, R. (2021). Response, willingness, and data donation in a study on accelerometer possession in the general population. Survey Practice, 14(1).

van Kuppevelt, D., Heywood, J., Hamer, M., Sabia, S., Fitzsimons, E., & van Hees, V. (2018). *Segmenting accelerometer data from daily life with unsupervised machine learning* [Preprint]. Epidemiology. https://doi.org/10.1101/263046

Wu, Y., Johns, J. A., Poitras, J., Kimmerly, D. S., & O'Brien, M. W. (2020). Improving the criterion validity of the activPAL in determining physical activity intensity during laboratory and free-living conditions. *Journal of Sports Sciences*, 1–9. https://doi.org/10.1080/02640414.2020.1847503

Appendix 1a. Instruction how to mount the activPAL (NL)

User Guide Motion Meters

You have received an envelope from us with the following contents:

- a bubble envelope containing a motion meter in a protective pouch

- some 'Tegaderm' stickers to attach the motion meter

- a return envelope to send the motion meter back to I&O

This guide explains how to attach and wear the motion meter.

ATTENTION!

Put on the exercise meters before going to sleep on Sunday evening …. They are programmed to start measuring at midnight.

If you have questions, please contact

Naam:

Telefoon:

E-mail:

Note

We ask you to wear the exercise meter for 7 consecutive days and nights, from the evening of … to the evening of …

ATTENTION!

The exercise monitor can be left on when showering.

Take the exercise meters off temporarily for water related activities such as swimming, surfing and during a sauna visit. You should also remove the exercise meters during medical examinations, such as an MRI scan or an X-ray, or at customs at the airport.

Sunday evening, …, the exercise meter may be taken off again. Please post the meter the next day. We need it for the next round of the survey.

When we receive the meter, you will receive the Bol.com link of €20.

Explanation how to attach

This is the motion meter: the *activPal.*

It is wrapped in a protective rubber.

Take a sheet of *3M Tegaderm*. Lay it on the table with the white site up (where it shows 1) and remove the white side (there is a small loose strip).

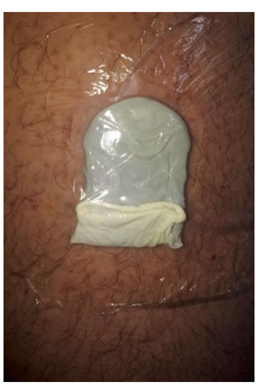
Now put the *activPAL* with the **orange site down** on the sticking site of the *3M Tegaderm*.

Fold the rubber, to carefully enfold the motion meter.

Sit down and stick the *activPAL* in the middle of your

d tip

Now remove the other side of the *3M Tegaderm*.



The a*ctivPal* is now ready for measurement.

You can leave the a*ctivPal* on during the entire week, also during showering. But take it of when you go swimming or in the sauna. There are additional theraderm plasters in case you need to take of the activPAL.

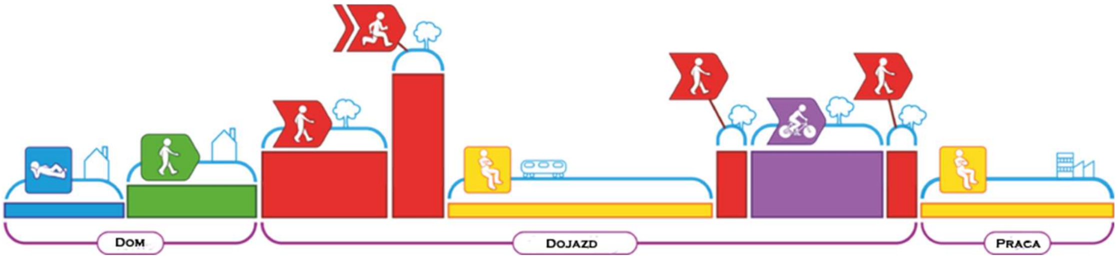
Good Luck!

Appendix 1b. Instruction PL

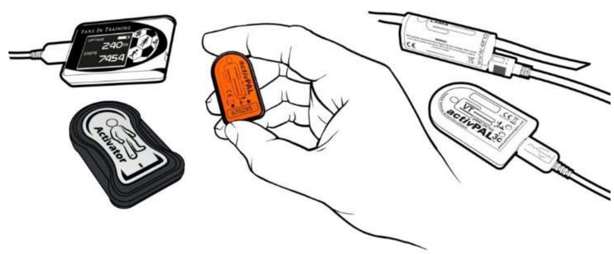
What is the survey about?

In the Smart Survey project, we are investigating the use of smart devices that inform us about health issues. The experimental study involves using accelerometers to examine people's daily physical activity both at home and while commuting to and from work.

What is activPAL™?



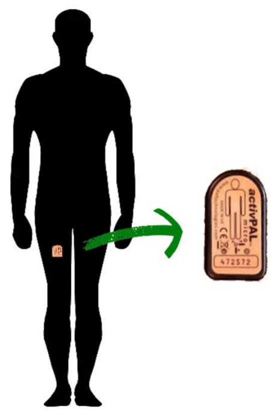
The activPAL™ is a miniature electronic recorder designed to quantify daily physical activity. The device contains a microprocessor, a sensor and recording element as well as a associated electronics and a power supply. The microprocessor controls the processing and recording of the sensor signal, with minimal burden on the user, using complex analysis routines to produce an activity classification profile.



What kind of data is collected?

The activPAL™ device measures physical activity. When you move, the device generates data on time spent lying down, sitting, standing and moving every second of the day.

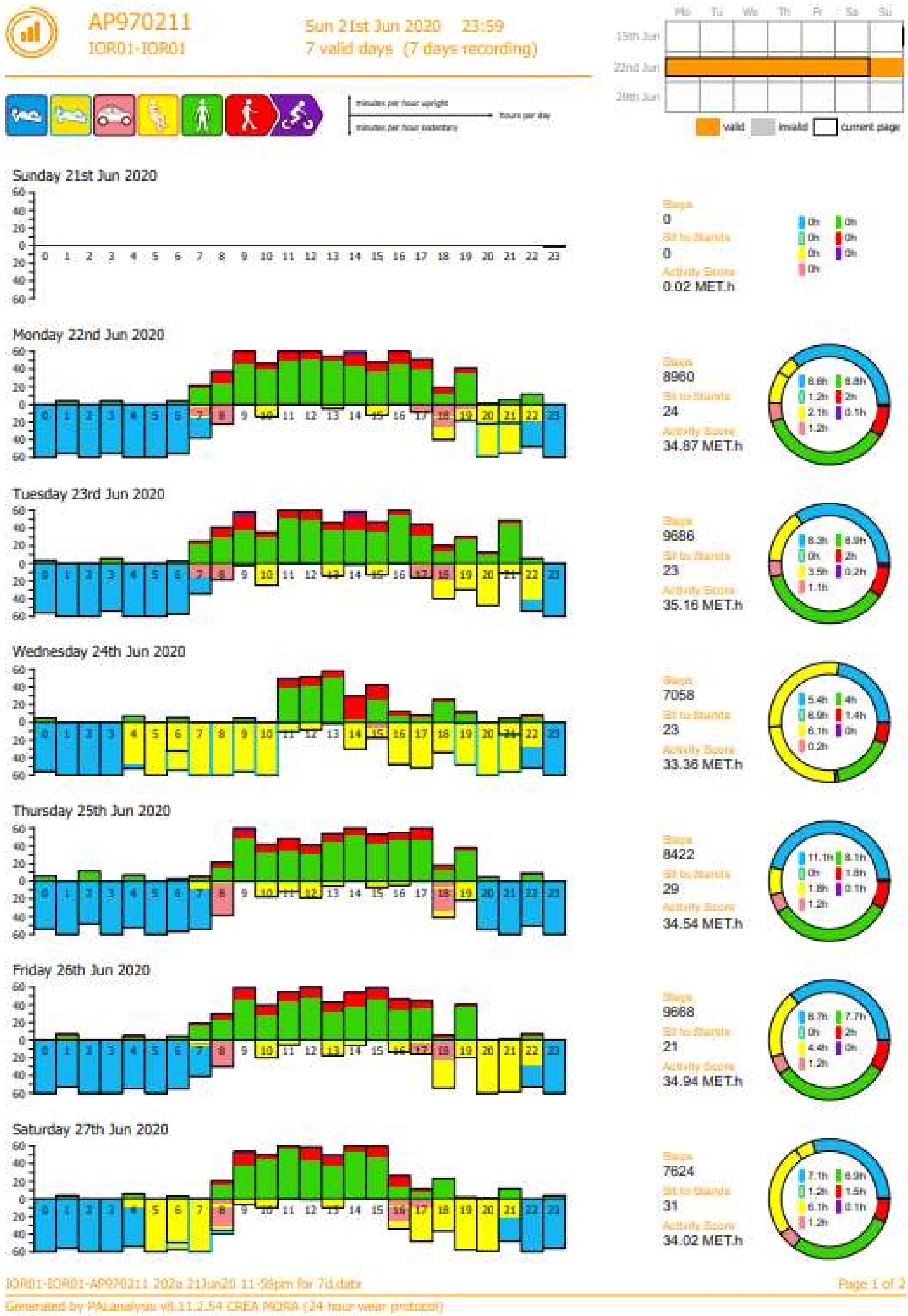
How to wear and use activPAL™?



The activPAL™ device is inserted into the protective rubber finger. Place the device on the thigh in such a way that the graphic element and diodes are visible, and the side without graphic elements adheres to the leg. Adhere with the supplied waterproof bandage.

Wear the device 24 hours a day, including sleep and bathing/showering (the bandage is waterproof). If the bandage comes off, a new one should be applied.

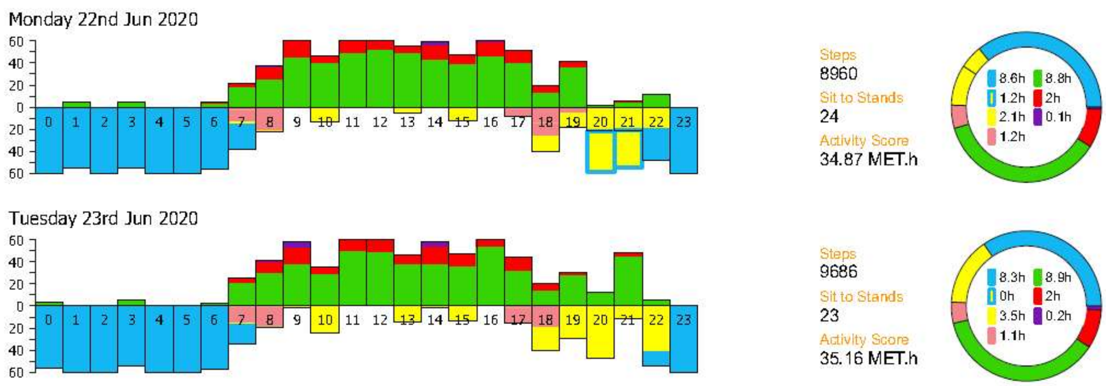
Appendix 2. Example of participant feedback



Explanation of your exercise profile.

Here you find an overview of the activities measured by the meter on your upper leg. Below is a short example to illustrate. You see per day how many minutes you have spent on consecutively: sleeping, lying down/sleeping, driving, sitting, standing, walking and cycling. Driving a car is measured separately because the meter could otherwise wrongly conclude that you were moving yourself.

Below the 0 line you see how many minutes you were at rest (sedentary: lying down, sitting and sleeping) and above the line the other activities.



You also see for each day how many steps you took (Steps, where cycling is also converted to steps), how many times you stood up (Sit to Stands), and what your Activity Score is: the average MET value per measured hour (MET.h). The MET value indicates how much energy a particular effort takes compared to the amount of energy needed at rest. One MET corresponds to the amount of energy consumed while sitting still. The MET value of physical activity ranges from 0.9 MET (when sleeping) to 18 MET (strenuous exercise). If you do an activity of 4 MET for 30 minutes, for example cycling, that is 4 x 30 = 120 MET minutes, or 2.0 MET hours (MET.h).

If you were to sit for an entire day, the MET.h value would be 24 (24 x 60 x 1 MET minutes). In both days of this example, about 11 MET hours more were moved. Research has shown that significant health gains are achieved with as little as 11 MET hours of increased exercise.

The Dutch exercise guideline for adults and the elderly is as follows:

- Do at least 150 minutes per week of moderately intensive exercise, such as walking and cycling, spread over several days.

- Do at least twice a week muscle and bone strengthening activities, for older people combined with balance exercises.

- Avoid sitting still a lot.

To see if you meet the first guideline, you can add up the number of "red" and "purple" minutes per day for the total. In the example above that is over 2 hours a day (total already 240 minutes). The 150 minutes should also be spread over a number of days. A long bike ride on the weekend is of course fine, but should not be the only exercise.

Appendix 3. Start and evaluation questionnaire Poland

Urząd Statystyczny
w Rzeszowie

General physical activity survey

Device code:

The questionnaire is part of the activPAL™ smart device physical activity survey and should be completed before the survey begins. The survey is anonymous and the results will be used to prepare the final report of the project.

For questions 1 - 5, please put an 'x' in the box next to the one answer you chose.

1. Sex ☐ Female ☐ Male
2. Age
   a. ☐ 17 – 25
   b. ☐ 26 – 35
   c. ☐ 36 – 49
   d. ☐ 50+
3. How often do you engage in sport or physical activity?
   a. ☐ less than twice a week
   b. ☐ 2 – 4 times a week
   c. ☐ 5 – 6 times a week
   d. ☐ everyday
4. How would you rate your overall physical activity throughout the day?
   a. ☐ lack of physical activity
   b. ☐ small
   c. ☐ medium
   d. ☐ high
5. How many hours a week on average do you engage in sport or physical activity?
   a. ☐ 1 hour or less
   b. ☐ 2 – 3 hours
   c. ☐ 4 – 5 hours
   d. ☐ 6 hours and more

6. What sport do you practise most often? Please specify ….……………………………………………

Urząd Statystyczny w Rzeszowie

Questionnaire to be filled in after the activPAL™ physical activity survey

The survey is anonymous and should be completed after the survey has been completed.

1. Has wearing the device caused you any discomfort or problems??
   a. ☐ yes (what kind of?)
   ……………………………………………………………………………………
   b. ☐ no
   c. ☐ I don’t know

2. Has awareness of the data your activPAL™ device collects increased your physical activity?
   a. ☐ yes
   b. ☐ no
   c. ☐ I don’t know

3. Was the duration of the study (2 weeks) appropriate in your opinion?
   a. ☐ too long
   b. ☐ too short
   c. ☐ adequate
   d. ☐ I don't have an opinion

4. Would you be interested in participating in other studies using this type of smart device?
   a. ☐ yes
   b. ☐ no
   c. ☐ I don’t know

5. What form of motivation would encourage you to take part in research using smart devices?
   a. ☐ financial motives
   b. ☐ report on my activities
   c. ☐ I do not need any form of motivation
   d. ☐ other (what kind of?)
   ……………………………………………………………………………………

6. Do you have any suggestions or comments on the survey?

……..

……..

Appendix 4. Evaluation questionnaire – Belgium

Survey questionnaire

Thank you once again for taking part in the experiment with the posture and movement monitor activPAL where you wore a sensor on your upper leg for one week.
The main purpose of the experiment is to determine the extent to which our NSI can use this type of technology to collect data in the future. It is important to gain insight into how respondents react to such data collection methods. That's why we're launching a short questionnaire to gauge your experiences during the experiment as well as your opinions on the use of such technologies. Of course, we realise that as a NSI employee, you have a coloured view. Still, we are convinced that we can learn a lot from your answers. May we therefore ask you to answer as openly as possible. Thanks in advance!

Before we start, we would like to ask you again for your permission to use the data collected through this questionnaire for our research.

1. **I agree that the data provided by me in this questionnaire will be further processed to produce global and anonymous statistics on the subject of this study.**
   ( ) Yes
   ( ) No

The following questions are about the use of the posture and movement monitor activPAL. To what extent do you agree with the following statements?

2. **The instructions for attaching the sensor were clear.**
   ( ) Strongly agree (continu to question 4)
   ( ) Agree (continu to question 4)
   ( ) Disagree
   ( ) Strongly disagree
3. **What was unclear to you?**
4. **A short instruction video explaining step-by-step how to attach the sensor would be more efficient.**
   ( ) Strongly agree
   ( ) Agree
   ( ) Disagree
   ( ) Strongly disagree
5. **I was able to attach the sensor without any problems.**
   ( ) Strongly agree
   ( ) Agree
   ( ) Disagree (continu to question 7)
   ( ) Strongly disagree (continu to question 7)
6. **What problems did you encounter?**
7. **I had physical discomfort while wearing the sensor.**
   ( ) Strongly agree
   ( ) Agree
   ( ) Disagree (continu to question 7)

( ) Strongly disagree (continu to question 7)

8. **What physical discomforts did you experience?**
9. **I was aware of the sensor and the fact that my movements were being recorded.**

   ( ) Very often

   ( ) Often

   ( ) Sometimes

   ( ) Rarely

   ( ) Never (continu to question 11)
10. **After some time I became less aware of the sensor.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Disagree

    ( ) Strongly disagree
11. **Wearing the sensor affected my usual movement behaviour.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Disagree (continu to question 13)

    ( ) Strongly disagree (continu to question 13)
12. **In what way was your movement behaviour influenced? Did wearing the sensor make you move more? Or did you move less because for example the sensor was not suitable for swimming?**

Passive data collection, as in this case, creates a whole new setting that is very different from the methods we have been using up to now within our NSI. To what extent do you agree with the following statements?

13. **It is important for me to receive information about the proper functioning of the sensor (e.g. in the form of a green/red light).**

    ( ) Strongly agree

    ( ) Agree

    ( ) Neither agree nor disagree

    ( ) Disagree

    ( ) Strongly disagree
14. **It is important to me that the sensor is extended with a pause button which allows me to interrupt the observation period.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Neither agree nor disagree

    ( ) Disagree

    ( ) Strongly disagree
15. **It is important to me that I can validate the observations myself.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Neither agree nor disagree

    ( ) Disagree

    ( ) Strongly disagree

16. **It is important to me that I can delete certain observations or observation periods myself (because they are incorrect or because I do not want them to be shared).**
    ( ) Strongly agree
    ( ) Agree
    ( ) Neither agree nor disagree
    ( ) Disagree
    ( ) Strongly disagree
17. **It is important to me that I can provide context to the observations.**
    ( ) Strongly agree
    ( ) Agree
    ( ) Neither agree nor disagree
    ( ) Disagree
    ( ) Strongly disagree
18. **It is important to me that I can get feedback on my results and how they relate to an average.**
    ( ) Strongly agree
    ( ) Agree
    ( ) Neither agree nor disagree
    ( ) Disagree
    ( ) Strongly disagree
19. **Do you find it a problem that our NSI is able to access rather privacy-sensitive health aspects such as the number of steps you have taken, information about your sleeping pattern, ... over which you have no control?**
    ( ) Definitely yes
    ( ) Rather yes
    ( ) Rather no
    ( ) Definitely no
20. **Do you find it a problem that other NSI's (such as CBS in this experiment) are able to access rather privacy-sensitive health aspects such as the number of steps you have taken, information about your sleeping pattern, ... over which you have no control?**
    ( ) Definitely yes
    ( ) Rather yes
    ( ) Rather no
    ( ) Definitely no
21. **Do you think it is a problem that commercial organisations are able to access rather privacy-sensitive health aspects such as the number of steps you have taken, information about your sleeping pattern, ... over which you have no control?**
    ( ) Definitely yes
    ( ) Rather yes
    ( ) Rather no
    ( ) Definitely no
22. **Privacy is much more of an issue in passive measurements with sensors (as in this case) than in interviews with an interviewer.**
    ( ) Strongly agree

( ) Agree

( ) Neither agree nor disagree

( ) Disagree

( ) Strongly disagree

23. **Privacy is much more of an issue in passive measurements with sensors because of the cooperation with commercial organisations.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Neither agree nor disagree

    ( ) Disagree

    ( ) Strongly disagree

Finally, we would like to gauge how valuable you found this experience and to what extent you think this type of technology can be used to produce statistics.

24. **I found it interesting to take part in the experiment with the posture and movement monitor activPAL.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Disagree

    ( ) Strongly disagree

25. **The feedback received regarding my movement profile meets my expectations.**

    ( ) Strongly agree (continu to question 27)

    ( ) Agree (continu to question 27)

    ( ) Disagree

    ( ) Strongly disagree

26. **What additional feedback did you expect regarding your movement profile?**
27. **I would be willing to fill in an online questionnaire or diary on a daily basis during the measurement period regarding the conditions under which the activities are undertaken.**

    ( ) Strongly agree

    ( ) Agree

    ( ) Disagree

    ( ) Strongly disagree

28. **I think it is feasible in the long run to use these kinds of technologies to collect data -- especially as a supplement to existing data collection methods within our NSI.**

    ( ) Strongly agree (continu to question 30)

    ( ) Agree (continu to question 30)

    ( ) Neither agree nor disagree (continu to question 30)

    ( ) Disagree

    ( ) Strongly disagree

29. **Why do you think it is not feasible to use these types of technologies within our NSI?**
30. **I suspect that many people will refuse to participate in fieldwork organised on the basis of such passive measurements.**

    ( ) Strongly agree

( ) Agree

( ) Neither agree nor disagree (continu to question 32)

( ) Disagree (continu to question 32)

( ) Strongly disagree (continu to question 32)

31. Why do you think people will be reluctant?

To end, two more questions.

32. Do you already use health apps and wearable health meters, such as a smartwatch, a pedometer, etc.?

( ) Yes

( ) No

33. I took part in the experiment with the posture and movement monitor activPAL ...

( ) because I found the experiment interesting

( ) because I wanted to gain an insight into my movement profile

( ) because I wanted to do [redacted] a favour

( ) because I think this kind of innovation is important for our NSI

Thank you for your cooperation!

Appendix 5 Survey results Belgium

Survey results

Profile of the participants

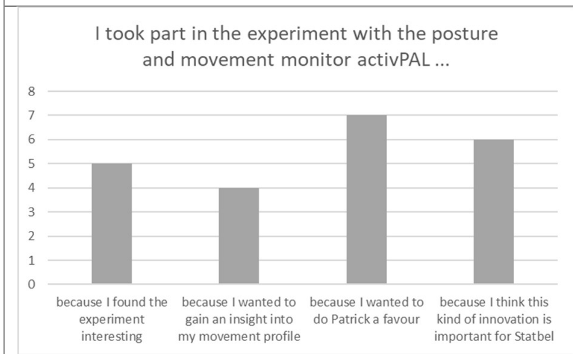


Figure 1

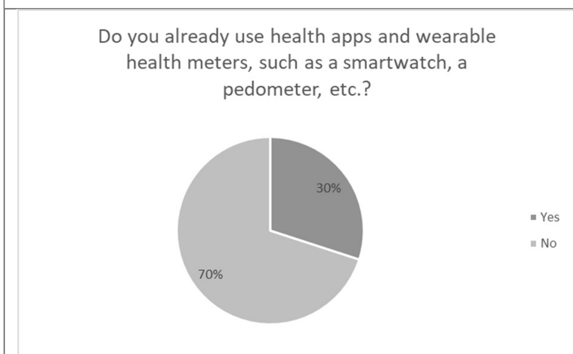


Figure 2

Usage and experiences

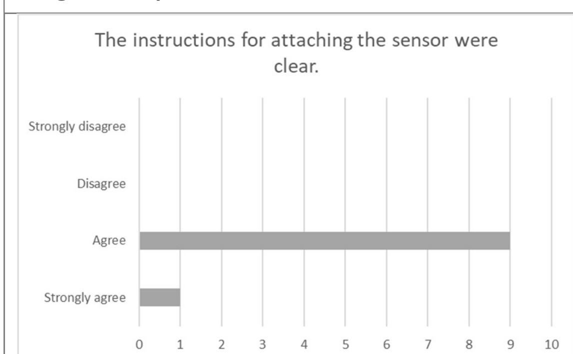


Figure 3

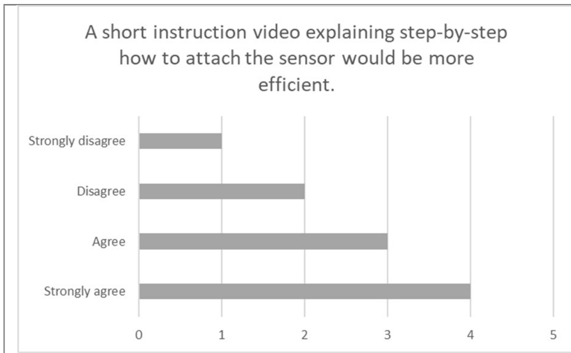


Figure 4

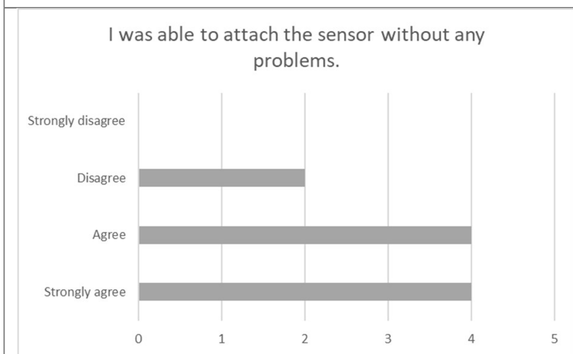


Figure 5

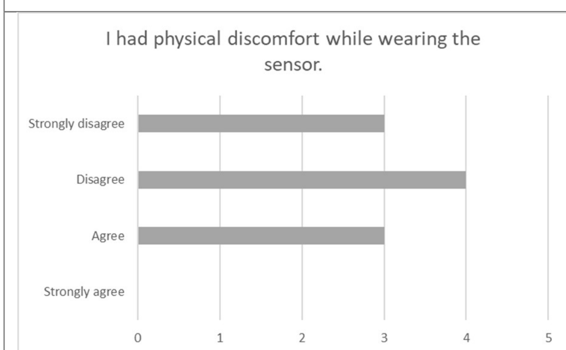


Figure 6

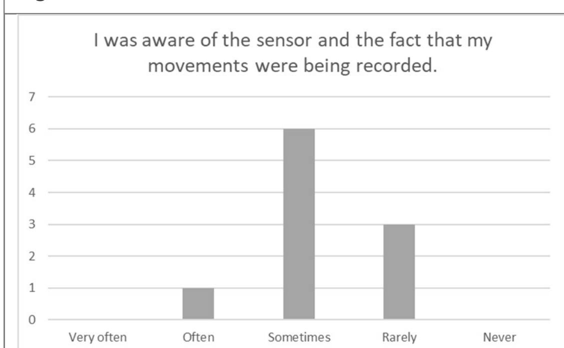


Figure 7

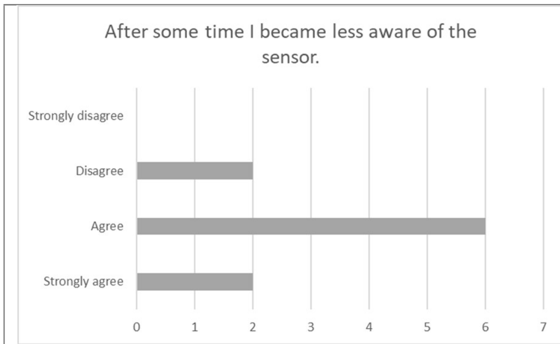


Figure 8

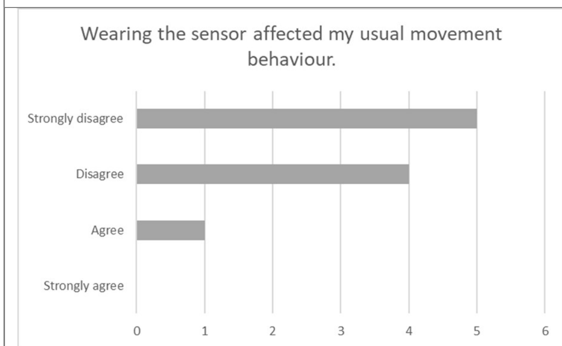


Figure 9

Passive measurement perception

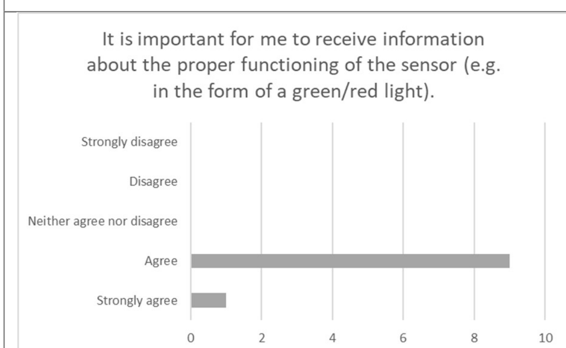


Figure 10

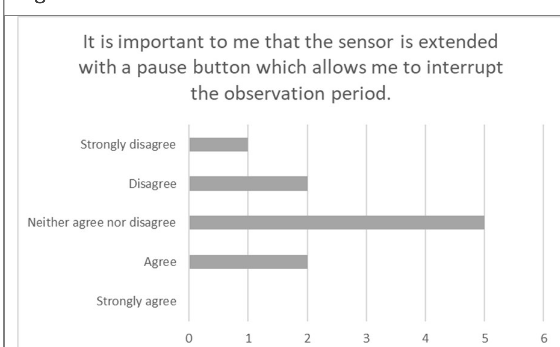


Figure 11

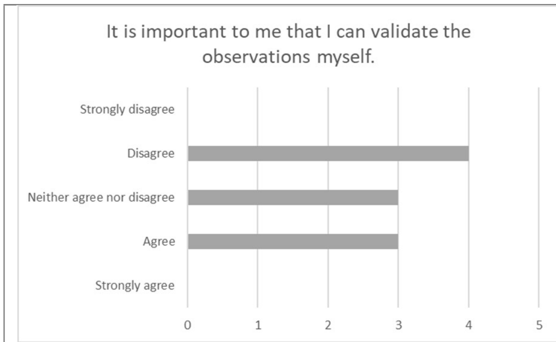


Figure 12

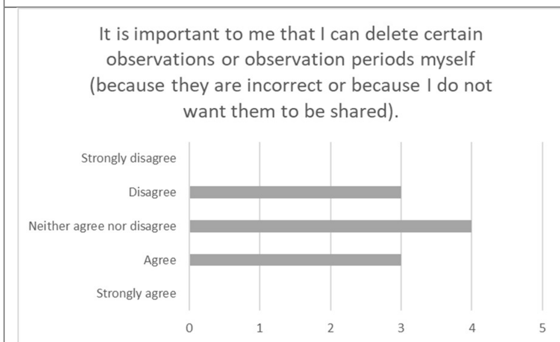


Figure 13

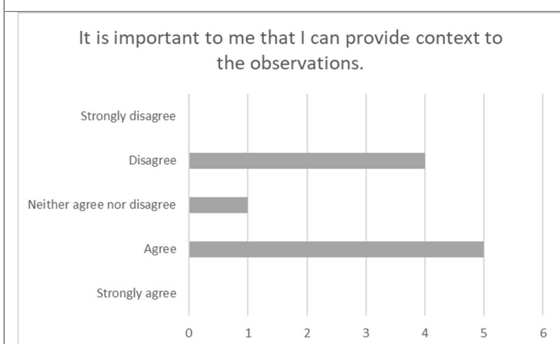


Figure 14

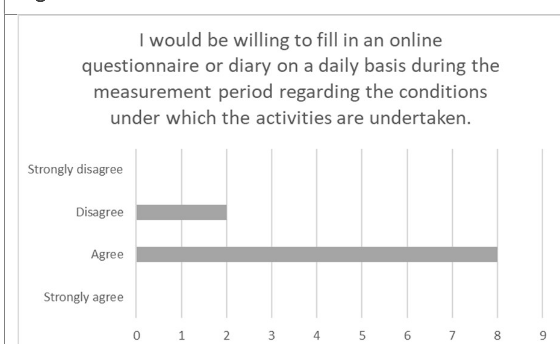


Figure 15

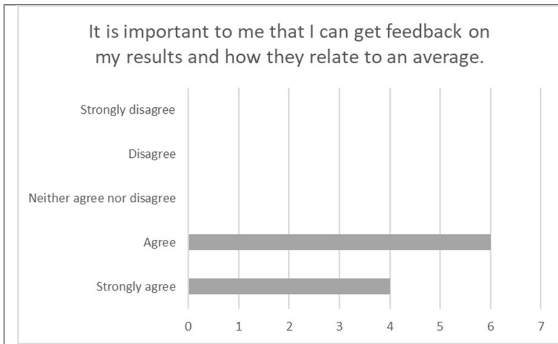


Figure 16

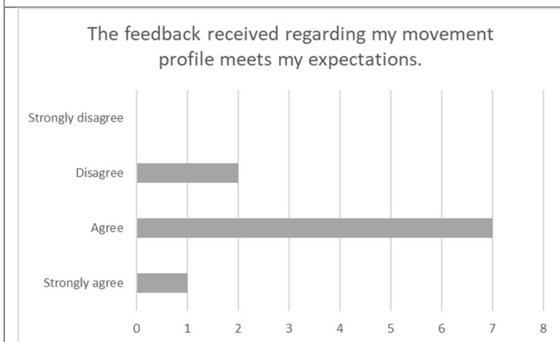


Figure 17

Privacy perception and attitudes

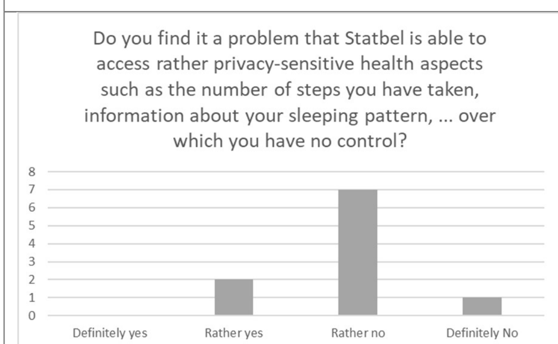


Figure 18

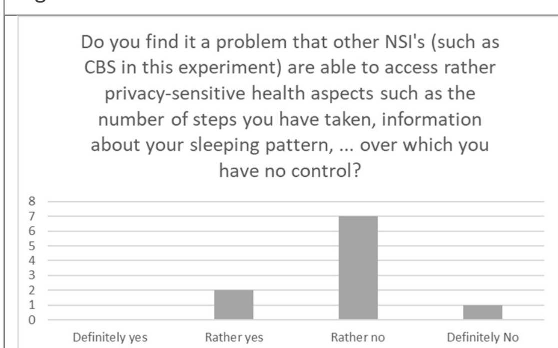


Figure 19

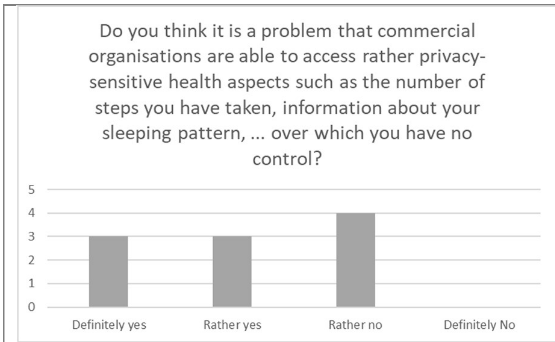


Figure 20

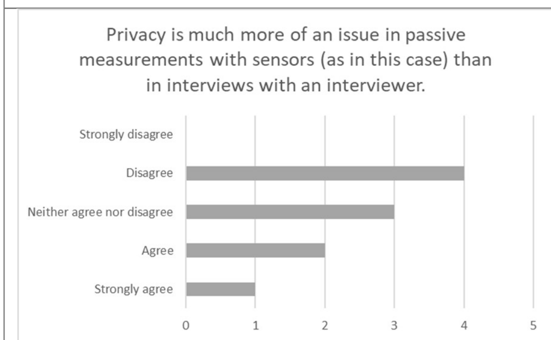


Figure 21

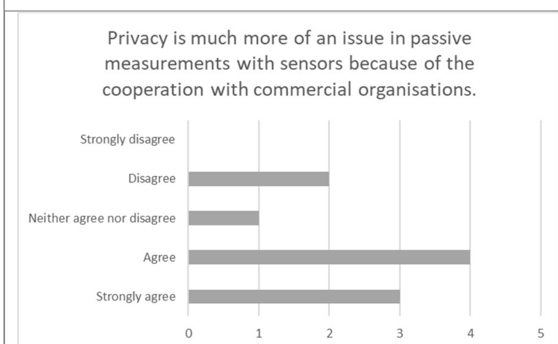


Figure 22

Value and statistical value perception

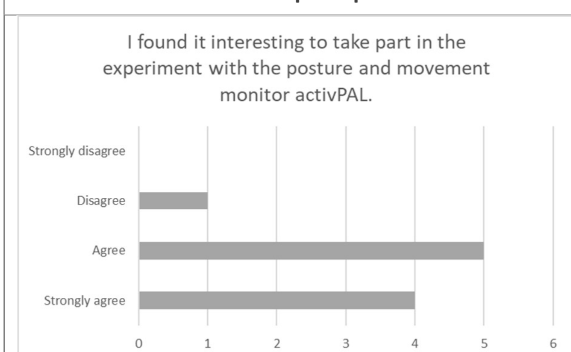


Figure 23

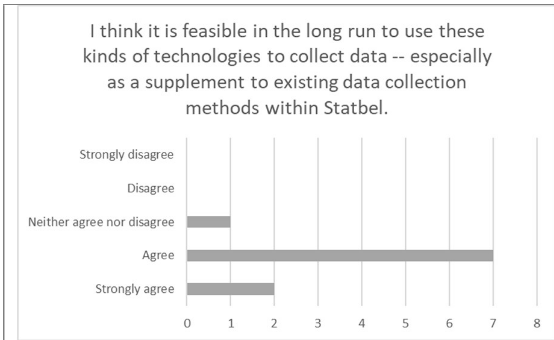


Figure 24

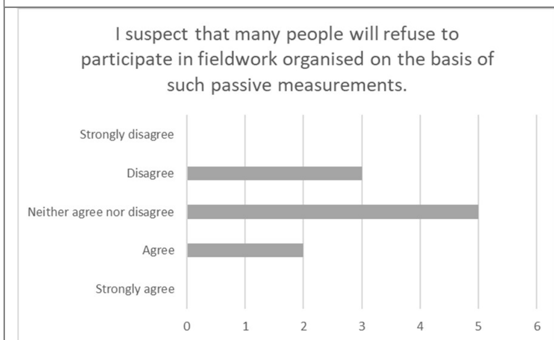


Figure 25

Appendix 6. Evaluation questionnaire pilot 1 – NL

You have been wearing our exercise monitor for a week now. One of the aims of our research is to find out whether we can also use the measurements from devices that people already own.

1. Would you please fill in the measurements of your own device for each day of the past week? It is about the score at the end of that day.
   Your own activity meter can be a smartwatch, an activity tracker or an app on your phone. If you have several devices, you can use the one you use the most.

	Ma 6 juli	Di 7 juli	Wo 8 juli	Do 9 juli	Vr 10 juli	Za 11 juli	Zo 12 juli	Not applicable
Number of steps								
Distance (traveled)								
Active minutes								
Calories burned								
Heart rate								
Sleep duration								
Speed (current and/or mean)								
Stairs climbed								

We would also like to know how you experienced wearing the activity tracker you received from us on loan.

2. Has your exercise behaviour (movement and sports) during the past week been different to normal?

   a. Yes, I did move more than usual during the past week
   b. Yes, I moved less than usual during the past week
   c. No, I have moved (about) as much as usual during the past week

3. How did you experience wearing the activity tracker?
   You can say here whether wearing the activity tracker was a success or a failure, and in which situations. You can also write down your suggestions for future research with activity trackers here.
   [open question]

These were our questions. Thank you for participating in our experiment.

If you have not yet returned the activity tracker to us, would you please do so? You can use the reply envelope that you received earlier.