

# The impact of a smart survey approach on participation and data quality

## - Tentative results and insights from the Norwegian HBS 2022

Nina Berg, Statistics Norway, [Nina.Berg@ssb.no](mailto:Nina.Berg@ssb.no)

Gezim Seferi, Statistics Norway, [Gezim.Seferi@ssb.no](mailto:Gezim.Seferi@ssb.no)

Contributors: Aina Holmøy, Kristin Egge-Hoveid, and Kari-Anne Lund, Statistics Norway

Oslo, 27.02.2022

### 1. Abstract

---

*Statistics Norway are one of the front runners in taking advantage of new, advanced technology to collect data on household expenditures for the Household Budget Survey (HBS). We have recently received the first batch of data from the survey's 1st quarter in field. We are excited to share the preliminary results for the Norwegian HBS 2022. In this paper we discuss the impact of using a smart survey approach for participation and data quality.*

*Overall, our data show that the HBS app is well received. It does not seem to have a negative impact on recruitment and dropouts, except in the older age groups.*

*We find respondents unexpectedly choose manual registration more often than scanning. This does not seem to be linked to digital skills, but that many respondents do not have paper receipts available.*

*Scanning helps the respondents' memory and reduces the response burden, and therefore also, improves the data quality. The bigger volume of data points, even with some errors, benefits the data quality. But it also increases our editing job.*

*With access to transaction data, we can keep improving data quality through machine learning and continue to develop a smart survey approach to reduce response burden and better future statistics.*

**Keywords:** *Smart surveys, new technology, new data collection tools, respondent perspective, questionnaire design, survey design, data quality, representativity*

## 2. Table of content

---

<b>1. Abstract</b> .....	<b>1</b>
<b>2. Table of content</b> .....	<b>2</b>
<b>3. Introduction</b> .....	<b>3</b>
<b>4. About the study</b> .....	<b>4</b>
4.1 Our research questions .....	4
<b>5. Description of survey setup and app</b> .....	<b>5</b>
5.1 HBS 2022 survey design.....	5
5.2 Survey communication .....	6
5.3 Development of a mobile app for data collection .....	8
<b>6. Results</b> .....	<b>9</b>
6.1 Participation .....	9
6.2 Use of smart features .....	16
6.3 Usability and user experience.....	22
6.4 Data quality .....	26
<b>7. Discussion of findings</b> .....	<b>37</b>
7.1 How does use of a smart survey approach impact participation? .....	37
7.2 How does use of a smart survey approach impact data quality? .....	38
7.3 Going forward .....	40
<b>8. References</b> .....	<b>42</b>
<b>9. Appendix</b> .....	<b>44</b>
9.1 Essnet projects on new data and data collection tools.....	45
9.2 Digital development in Norway .....	45
9.3 HBS survey setup in Norway .....	45
9.4 Response rates.....	47
9.5 Non-response.....	50
9.6 Qualitative data HBS .....	51
9.6 Short about user testing.....	52
9.7 Topic of contact reasons at Helpdesk support HBS.....	52
9.1 Tables .....	54

### 3. Introduction

---

The Norwegian Household Budget Survey 2022 (HBS 2022) is the first survey at Statistics Norway (SSB) to use mobile app technology in our data collection. With mobile apps like the @HBS app, MOTUS in the European (ESSnet)<sup>1</sup>, and the new Time Use Survey app we are developing, SSB is at the forefront of using Smart Survey tools in data collection. Smart Surveys refer to surveys based on data collection using smart personal devices, typically the smartphone (Mussmann, 2019).

In this paper, we will share our experience from developing and collecting data for the first household survey in Norway that uses an app with optical character recognition (OCR) to scan receipts on smartphones. We assess our preliminary results to review what impact the use of smart survey devices and features, have on survey participation and data quality.

We use the preliminary survey results from HBS 2022, the first quarter (Q1) in Norway (N=903), as well as qualitative user tests and in-depth interviews (N=126) with respondents for our analysis. See description in appendix 9.4 and 9.6. Our data and insights are not final as the field period is not completed.

Our approach focuses on the user perspective of our survey respondents. In sharing our first experiences with a smart survey approach in data collection, we want to add to the collective knowledge of best practice for data collection in official statistics.

---

<sup>1</sup> The ESSnet (European Statistical System) is the partnership between the EU statistical authority, which is the Commission (Eurostat), the 'National Statistical Institutes' (NSIs) and 'Other National Authorities' (ONAs) in each EU Member State that are responsible for the development, production and dissemination of European statistics.

## 4. About the study

---

Key to the concept of smart surveys is smart devices with sensor technologies such as accelerometers, GPS, OCR scanning, or similar. Smart surveys use sensor technology in devices such as respondents' smartphone to measure objective features together with survey data from respondents. The idea is to let sensors assist or passively measure instead of respondents having to actively register or answer questions etc, to reduce response burden and increase data quality. Smart devices and technology also provide the opportunity to use personalised feedback depending on respondent's interaction with instrument. More intelligent lists of pre-coded answers can be used, such as product lists and user defined answer options using machine learning. Both sensors and interactive technology can alleviate response burden and hence attempt to counteract the declining response rates we generally see in surveys.

### 4.1 Our research questions

The question is if data collection with smart devices and features will reduce the response burden and halt declining response rates. And, if "smarter" data measures more accurately and correct. Or is a smart survey approach a cause for concern for participation, instrument effect, and or data quality? Presently, Statistics Norway does not have alternative solutions, such as a paper diary or computer assisted telephone interviewing for respondents who can't use a smartphone, and a key question is whether we need one.

We will assess both quantitative and qualitative HBS data for first quarter 2022 and try to answer our two research questions:

1. How does use of a smart survey approach impact participation?
2. How does use of a smart survey approach impact data quality?

## 5. Description of survey setup and app

---

### 5.1 HBS 2022 survey design

During the planning of a new HBS in Norway, we discussed the best way to administrate data collection to get an adequate response rate in times where it is falling, achieve good data quality, and comparable results with earlier surveys. In 2017, we started piloting a web version of HBS that was never launched full scale, mainly due to a lack of funding and a wish to explore transaction data<sup>2</sup> as an alternative data source. And for 2022, we aspired to setup a smart survey solution using a mobile app with OCR scanning and possible use of machine learning from transaction data. Use of transaction data have not been possible to realise for 2022, but an app with scanning is used in field.

It is obvious that we have to use modern, or digital, technology to reduce response burden and meet today's user expectations to surveys and reporting systems. As smartphones are the preferred device to answer surveys on, it was a natural choice to design the new HBS for mobile phones. Further, we chose a progressive web app (PWA), which is an application that can run on all online platforms not only on mobiles, to support older respondents that might preferer working on larger screens and keyboards. For the user, a PWA is just like opening a web link form and does not require downloading from App Store or Goggle Play. Along with this decision, we discussed the need for alternative solutions for respondents that do not perceive themselves as "digitally" proficient.

#### *Design*

The HBS is designed with a sample of 12 000 households, and contact persons are normally between 18-84 years (see comment on contact persons age sample composition 10). Survey consists of a questionnaire and a diary to register all

---

<sup>2</sup> This could be 1) bank transaction data from Nets that can provide total amount of purchase and 2) receipts data from the big grocery stores with product name and price etc. These two data sources can be matched and compared against survey data to access correct product category, values, amounts etc for expenses in Norwegian households. This can be done to either improve or replace existing survey data. This could be a step forward in reducing response burden and improving data quality, but it might be overstepping what respondents see as acceptable in respect to privacy. Statics Norway hope to be able to explore these kinds of data and find ways to move forward within a system that protects both respondents' privacy, data security, and development of new data collection methods.

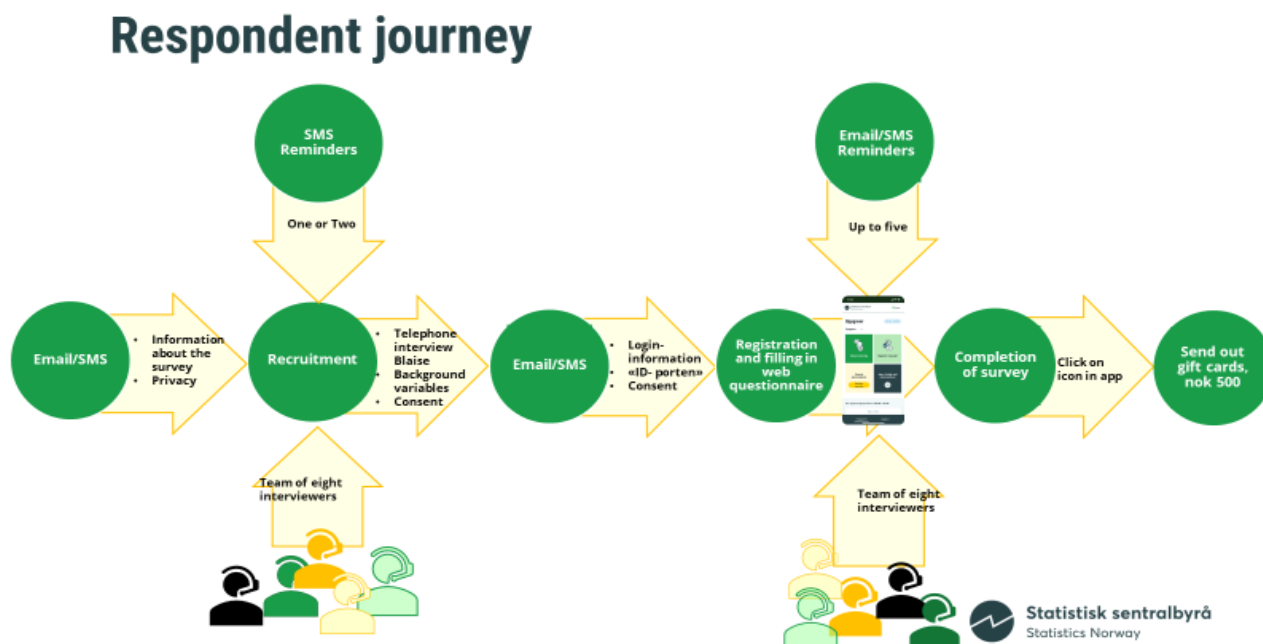
household expenses for a specified week. The selected sample is contacted by telephone for a recruitment interview (10 min). Then contact persons receive a text message with a link to the app where both the main questionnaire (approximately 20 minutes) and the registrations of household expenses are to be done by self-completion. When the contact person for the household confirms the completion of both the questionnaire and the registration week, they receive a gift certificate of € 50. See appendix 9.3 for further specifics about the HBS setup.

## 5.2 Survey communication

The survey communication for HBS is extensive.

From the household sample we recruited a contact person to report on behalf of the household. Every contact person that cooperated received several contacts from our recruitment personnel, via e-mail, SMS, and phone call. The communication plan for the respondent journey through the survey is shown in Figure 5.1.

Figure 5.1 Communication plan respondent journey



The first contact point in the user journey is an e-mail from SSB with information about the survey and that the household of the receiver is selected to participate in the survey. The email is sent three days prior to the recruitment call. The e-mail is

structured with the information of highest interest of the respondent is at the top of the email. This includes a € 50 incentive, privacy assurance, and information about the duration of the task/survey.

The registration period for the survey is one week. It starts on a Monday and ends on a Sunday. On the first working day of the registration-week, the contact persons will receive an SMS text with a link to our web app as a registration reminder. We decided to send the webpage link by SMS in order to “nudge” the respondent to use their mobile phone as intended.

All contacts are designed to act as reminders or push notifications on a smartphone, either by SMS or in-app. The main purpose is to motivate the respondent to register every purchase or expense of their entire household. If a respondent stops registering, an interviewer calls to assist. At this point the role of an interviewer shifts from the traditional interviewer role to providing service to motivate and help the respondent to complete the survey.

Our communication strategy for non-recruited or non-responders diverges from the strategy for the ones recruited. For those that have been recruited but have not yet started in app the number of contacts via both SMS and phone calls increases throughout the week. On Tuesday, an interviewer calls all contact persons with no registered expenses. Further, respondents who have registered expenses receive an SMS with tips to ease their participation and registration. The day after (Thursday) respondents receive another SMS with a reminder regarding the questionnaire. On the same day respondents with few or no registered expenses will receive a phone call. Respondents with no registered expenses will be offered to move their registration week to a maximum of 5 weeks forward, while respondents with 1-5 registered expenses will be assisted with potential issues regarding registration.

Respondents who complete the registration of expenses and the questionnaire by the end of their registration week will receive a “thank you” e-mail and their gift card certificate.

### 5.3 Development of a mobile app for data collection

SSB decided to run their own instrument development of the mobile app for HBS to be able to have full control of the process and to achieve spillover effects for other surveys and our data collections system. Our system for receiving and storing data was old, and we needed to upgrade to a cloud solution, and wanted to use HBS resources to break the ground for this.

#### *Requirements*

An important condition for our development was that the solution should run on all platforms, meaning all internet browsers and all types of devices. It should use scanning with automatic text recognition (minimum requirement); automatic product coding; log in with two-factor identification or authentication (that all official services use); communicate with Blaise 5 and SSB's survey systems; and be flexible and easy to supplement with machine learning modules. Ideally, it should be open source and possible to reuse within SSB and by others. And finally, be intuitive and easy to use for all respondents.

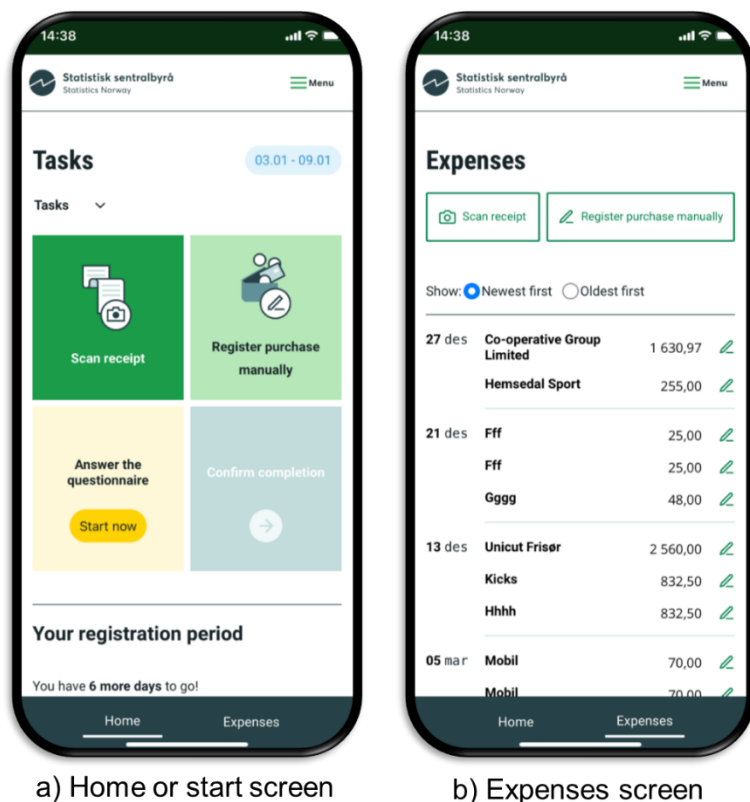
#### *Web app*

We chose to develop a progressive web app (PWA) to ensure a responsive app that is adaptable and optimized for any type of device, screen size, and resolution. It has an app-like interface and looks and acts like an app but is in practice a web-link that self-updates and is only accessible to authorised users.

First time logon is with two-factor authentication and on-boarding. For later logon, users go straight to the home screen and chose what task they are about to do. The home screen has access to four major functions: 1) Scan receipt, 2) register purchases manually, both for running expenses for registration period, and 3) a questionnaire regarding large expenses, yearly expenses and monthly fixed expenses. The 4) task, confirming completion, is not selectable until the registration period is over. Under the menu, users can find information about the survey, help and contact information, and their user profile.



Figure 5.2 Home and Expense screen in the HBS app



## 6. Results

We will present and assess both quantitative field and survey data and qualitative usability tests, in-depth interviews, and focus groups, in this section. First, we will look at participation and use, and then data quality.

### 6.1 Participation

#### *Response rate*

A crucial question is whether the HBS survey will achieve a response rate-adequate for official statistics. Q1 indicates that we can achieve a response rate of about 30 percent (see attachment 9.4, page 47). Our best estimate while planning was 40 per cent, like what is expected for a demanding online social survey today. The HBS is a more demanding survey, including also a one-week diary and cannot reach this level of respondent cooperation, and not at all the 48,9 per cent from 2012. But even if the response rate is lower than we hoped for, it is an acceptable achievement compared

to the figures coming in from European HBS pilots, which are around 20 per cent (de Groot, 2022), though the setup is not identical.

### *Sample bias*

To study sample biases, we have looked at indicators such as gender, age, education, household size, household type, and region for selected household contacts Q1/2022. Table 6.1 presents the difference between the gross sample before recruitment, which represents the proportional population size, and the net sample after recruitment in per cent for household or contact person depending on variable.

*Table 6.1 Demographic differences gross and net sample (per cent)*

	<b>Gross sample (per cent)</b>	<b>Net sample (per cent)</b>	<b>Difference</b>
<b>Total (N)</b>	<b>2 986</b>	<b>903</b>	
<b>Gender:</b>			
<b>Man</b>	35,6	37,2	1,6
<b>Woman</b>	64,2	62,8	-1,3
<b>Age:</b>			
<b>18-24 years</b>	4,2	4,8	0,5
<b>25-44</b>	44,8	46,9	2,1
<b>45-66</b>	37,8	39,0	1,2
<b>&gt; 67 years</b>	13,0	9,4	-3,6
<b>Education:</b>			
<b>Elementary School</b>	18,1	13,1	-5,0
<b>High school</b>	35,5	32,7	-2,8
<b>Higher education</b>	42,1	52,6	10,5
<b>Unknown</b>	4,0	1,6	-2,4
<b>Household size:</b>			
<b>1 person</b>	22,0	18,5	-3,4
<b>2</b>	34,5	33,4	-1,2

<b>3</b>	16,6	17,5	1,0
<b>4</b>	19,4	22,1	2,7
<b>&gt; 5 persons</b>	7,3	8,5	1,2
<b>Household type:</b>			
<b>Single</b>	18,5	17,4	-1,1
<b>Couples without children</b>	18,9	19,4	0,5
<b>Couples with children 0-17 years</b>	36,5	36,4	-0,1
<b>Single parents</b>	12,8	13,3	0,5
<b>Other</b>	13,3	13,5	0,2
<b>Region:</b>			
<b>Oslo and Viken</b>	36,6	38,6	2,0
<b>Innlandet</b>	6,8	6,9	0,0
<b>Agder and Sør-Østlandet</b>	13,4	12,8	-0,7
<b>Vestlandet</b>	25,1	23,7	-1,4
<b>Trøndelag</b>	8,8	9,5	0,7
<b>Nord-Norge</b>	9,0	8,6	-0,4

In the net sample composition, we see small biases for most indicators we controlled for. However, we see a clear underrepresentation for subgroups with lower education, people aged 67 years and older, and one-person households, and an overrepresentation for higher education and larger households (3 persons or more). Note that at present we have not yet uploaded household income from register to sample. Still, we know that education is a strong proxy variable for household income. Overrepresentation by higher education is a recurrent pattern in all our surveys independent of data collection method.

Note that selection of contact person in the households will have affected the gender and perhaps the age of the net sample composition as well. The selection of contact persons was random, with a few exceptions: In single-family households with children living at home, we chose the youngest of the adults, which in most cases are the women. For households where no one is under 70 years, we also chose the youngest,

which more often is the women. During recruitment, the household could choose who to be the contact person.

### *Motivation for participation*

In our qualitative work, we have tried to go beyond response rate statistics to understand attitudes and motivations among respondents. Interviewers report that a lack of motivation and understanding of the importance of participation are the primary reasons for refusing to participate. Many respondents say they do not have time; some do not want telephone sales or to answer opinion polls. (See description Table 9.3 for refusals, page 20.)

Privacy concerns does not rank high on our list of reasons for refusals. But we want to mention that SSB's access transaction data is under dispute. We have not detected any questions from recruited or respondents but note that the media have addressed whether transaction data is a proportional invasion of privacy. In connection to this, we see a few sceptical comments and anti-government attitudes surface online, from an article on NRKbeta<sup>3</sup>. As far as we can analyse, this group is very small, but probably most dominant among refusers – a group we have not been able to interview. If we go forward with use of transaction data, it is fair to assume that it will impact our relations to refusers. And what kind of data we use how for what purpose will be very important if we are to defend use of transaction data.

### *Participants*

For HBS, as for other social surveys, “community spirit”<sup>4</sup> is the most important motivation for participation. The respondents explain that they want to contribute to national statistics on household consumption and its use. They trust in the sender of information (Statistics Norway) and understanding of the benefits of official statistics. We observe that community spirit and trust in SSB is stronger among completers than

---

<sup>3</sup> <https://nrkbeta.no/2022/05/28/ssb-krever-a-fa-vite-noyaktig-hva-nordmenn-kjoper-i-matbutikken/>

<sup>4</sup> By “community spirit” we mean “dugnadsånd”, which could be defined as voluntary, unpaid work that is done together for the common, or national good in this case.

dropouts and non-starters. Community spirit and trusting SSB increase with age (40 years+).

Incentives motivate younger respondents (below 30 years), respondents with a lower educational level, and respondents with a lower household income. While secure handling of personal data (GDPR) is an important argument during recruitment, it is rarely mentioned as a reason for refusal, non-start, or dropout, but it is likely a concern for non-contact and non-response (that we have no data on).

### *Non-starters*

Non-starters agreed to participate during the recruitment interview; received logon info; but never started. They may or may not have attempted to open the link and log on. In this group we find that a lack of time; personal reasons, like family or job commitments; vacation; and sudden illness are hindering completion. Not unlike the refusal group. And more crucial for the HBS; they forgot to collect receipts.

It seems that the task has grown in their perception after they said yes to participate, and they never opened the link or started collecting receipts. We have the impression that some non-started think that the interviewers oversold the survey and made it sound easier than it is. We also experience that the sense of community spirit is not as strong in this group as among respondents that completed the survey.

Presently, we do not have figures whether non-starters attempted to logon or not. From qualitative interviews it does not seem so. Still, we can't rule out usability issues with logging on as a barrier for non-starters, because we know these have been issues for starters.

We do not have an impression from in-depth interviews that access to the internet and/or a device, the use of a web app, or privacy concerns are barriers for non-starters. Information on topics from our Helpdesk suggests that understanding of the survey theme and technical issues could be causes of concern for starters, hence it might affect non-starters as well. From our interviews, we detect that non-starters, more than refusers, we assume, have a positive attitude towards SSB and official statistics.

We seem to detect that there are more non-starters below 45 years, and more respondents with technical issues above 60 years+, but we do not have quantitative data to support this at this point.

### *Dropouts*

Dropouts are respondents that were recruited and started using the app but did not complete the survey. As for non-starters, lack of time due to sudden personal matters hindered completion. In addition, usability issues with logon, functionality issues in the app, not having receipts available when trying app, or not remembering to keep the receipts, are leading reasons for dropping out. Also, the fact that the survey must be confirmed completed add to incompletes.

From in-depths interviews, the burden of the task is notably an issue for wavering motivation. It is evident that several respondents struggle to keep up with registration throughout the week of registration. And the first impression of the questionnaire is for some overwhelming with a series of questions about regular and large expenses respondents do not know the correct answer to.

In regard to usability and user experiences, technical issues with the app or trouble registering expenses; lack of access to internet or equipment is mentioned, but do not rank high as reasons for non-completion. Neither do dropouts have second thoughts about GDPR, or trust of sender and data security and does not state this as a cause.

Compared to refusals and non-starters, active dropouts have higher motivations and community spirit. In the in-depth interviews they are often embarrassed for not having followed through and contributed to what they perceive as an important commitment to society. As such, it is obvious that if we can iron out functionality issues and ease the response burden by emphasising the advantages of using smartphone and scanning, we could help more dropouts complete.

Looking at socio demographics we have an impression from qualitative interviews that young people and young families more often drop out while older respondents try to follow through on their commitment. Our interviewers and our qualitative work do not experience device or use of app as refusals or dropout reason for older age groups. Still, our impression from our interviewers is that refusal increases for people in their

mid-70s and seems to be a real barrier from mid-80s. And from our Helpdesk, we see that people outside of work and people from 60 years and up, need more technical support. So even if technical issues do not rank high on causes for refusals and drop-out, we can't rule it out for the older age groups.

### *Completers*

Completers has per definition completed the questionnaire and saved at least one receipt or purchase. From demographic analysis, we know that there is somewhat a higher share of people with higher education and people aged 45-60 years in households with 3 persons+ among completes. Compared to refusals and non-completes, this group seem to have higher aspiration to serve. We find completers to be more understanding of the purpose of the survey and patient with the details required to complete the survey task. They often know Statistic Norway and want to support our work. They are not stopped when they do not know an answer or how to use the app. They often have a partner or network to ask for assistance, or maybe they feel confident or careless (?) enough to carry on with estimates. While non-completers seem to be bit more afraid of doing something incorrect, or they tire more easily of the task required and give up.

## 6.2 Use of smart features

We have also looked at data for use of smart features. And checked to what degree smartphone and scanning is used, and what respondents' attitude are towards this.

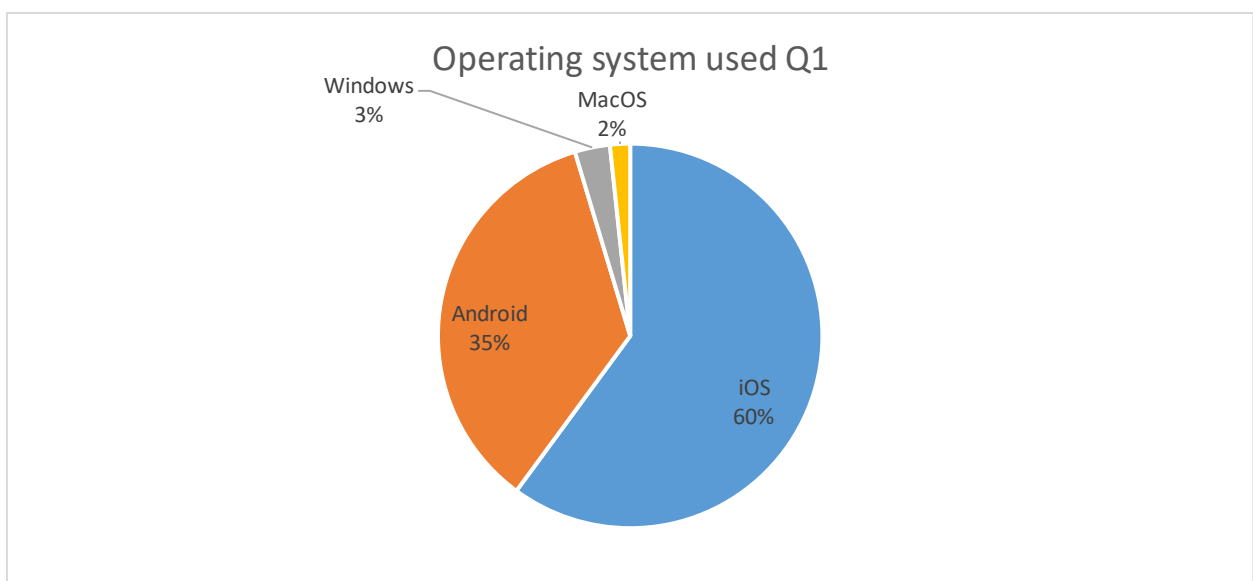
### *Main device*

One of our concerns was the large number of different devices and how we could deliver an instrument with a consistent user experience regardless of device, browser, or operating system. Our data from Q1 reflects the distribution of operating systems in the population and doesn't indicate a large diversity.

Most respondents use phones with either Android or iOS operating systems. This is in part caused by our communication and recruiting strategy where we strongly recommend the use of mobile phone because added functionality such as scanning.

The same divide between operating systems is evident regarding browsers as well. It seems that our respondents use the default browser included in the operating system which gives us predictability regarding compatibility between our web app and web browser.

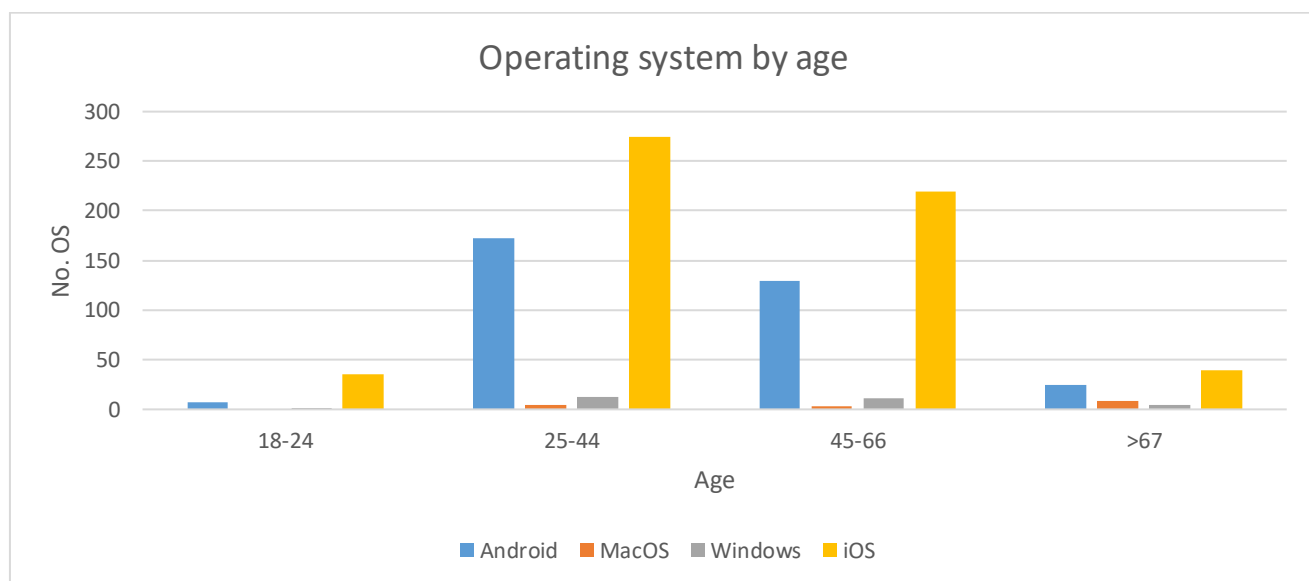
*Figure 6.1 Operating system used Q1*





We know from our Helpdesk that respondents with iOS seemed more prone to technical difficulties. We expected that age and technical experience could explain this, but looking at the figure below, we see that iOS is dominating in all age groups and does not support this thesis.

Figure 6.2 Operating system by age Q1



From our quantitative data we have seen those 67 years and older is underrepresented in the sample. Unfortunately, we do not have a finer breakdown of age in this group, which we should have to find the breaking point. As we know from qualitative work, reservation due to lack of device or ability to use smartphone does not seem to start that early. Rather we seem to detect a change for people early 70s and a real problem from 80 years or older.<sup>5</sup>

Neither in recruitment nor in our qualitative work do we detect much reservation due to the use of an app. For respondents below pension age, the smartphone is the expect and preferred device for registration and answering the survey. The smartphone is “always” at hand, it is mobile, and it is a service provide we expect all services available to work on it. Some people in the age group 60 and older open the app on a tablet or a pc because they prefer larger screens and keyboard, and some

<sup>5</sup> We have no completes 80 years or older in Q1.

use several screens/devices simultaneously. Habits can account for some of this as older age groups are more likely to do their banking on a larger screen than younger age groups. Moreover, our survey communication may also explain this because it has emphasised that the web app can work on all devices or screen sizes.

### *Scanning*

We had high aspirations for our new app, and at the start of data collection, we expected that most of our respondents would use the scanning feature because it reduces the response burden. When we started looking at the registrations, we noticed quite contrary that respondents more often used manual registration, see Table 6.2. Note that in our survey communication we mentioned the advantage of scanning, but the app design presents scanning and manual registrations as equal choices (see Figure 5.2).

Out of all the receipts registered in Q1, by 903 respondents, 52,2 per cent were manual and 47,5 per cent scanning. However, when we later will look at the number of product items, as opposed to the number of receipts, the distribution changes, and most products are registered using scanning (83,1 per cent), see Table 6.3.

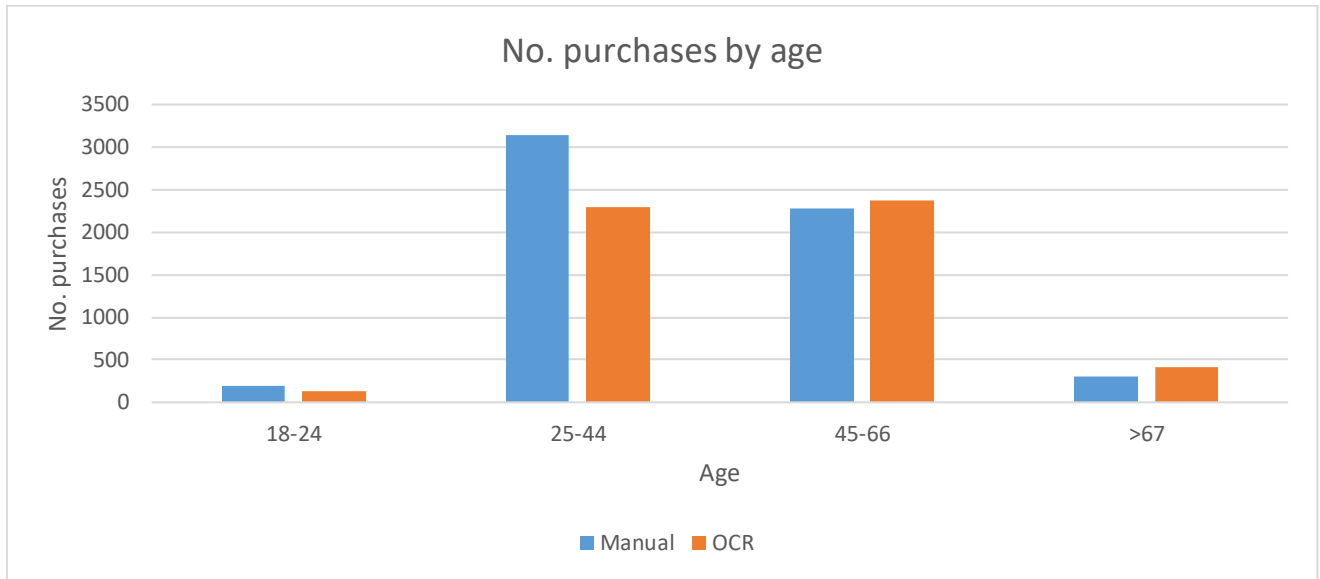
*Table 6.2 Total number of receipts by scanning or manual registrations*

	Count	Per cent
Total number of receipts	14 499	100,0
Manual	7 574	52,2
OCR scanning	6 925	47,8

*At the time of writing this report, we identified bugs in our data systems. Thus, the numbers here may not be exact but should not be off by a large margin.*

Further, we expected younger respondents, that often have more technical experience, to use scanning more frequent than older, but Figure 6.3 shows no such pattern.

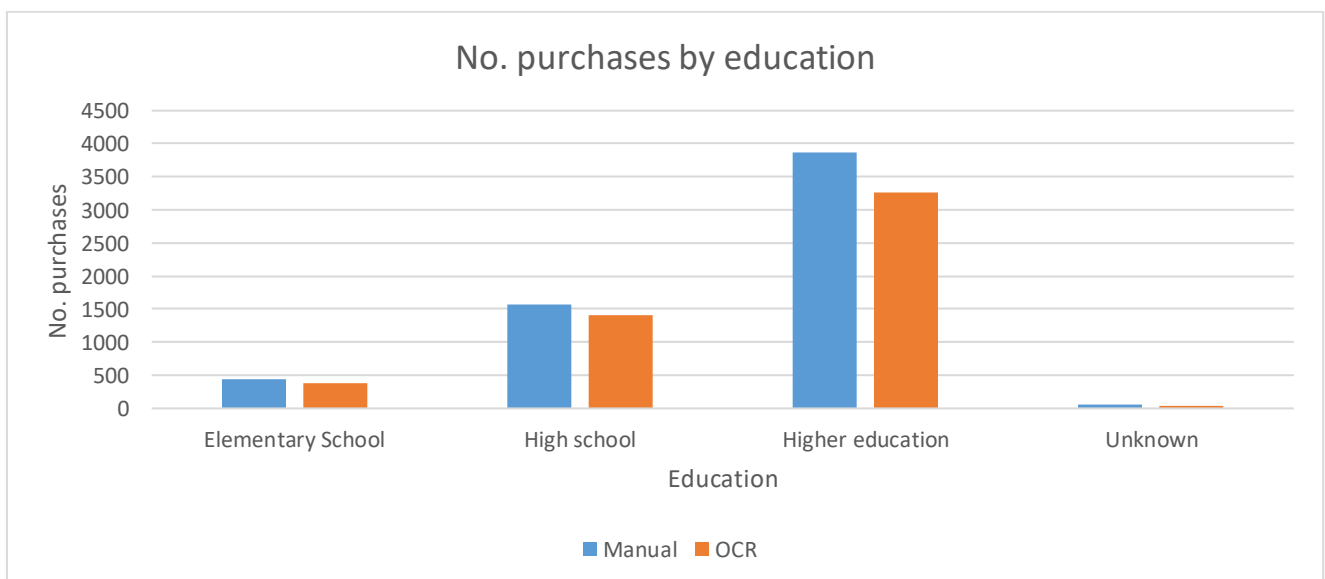
Figure 6.3 Number of purchases by age



Respondents up to 44 years more often used manual registration, whereas 45 year+ used manual and scanning in about equal amounts, with a slight overweight for scanning.

In relation to education, the share of manual registrations increases with the level of education, Figure 6.4.

Figure 6.4 Number of purchases by education



These findings are interesting since scanning is a vital feature for our app and smart survey. We regarded the scanning feature as deciding element regarding recruitment and essential to reduce the response burden. We need to study further the relationship between the share of scanning and household size and age, to see if scanning drops with high age, as we would assume it does.

### *Respondents' attitudes*

From qualitative interviews and focus groups with older respondents, we detect a reluctance to use new technology and low trust in their own digital skills. But, as we have seen, it is not a substantial refusal or dropout reason. Neither do we find less use of scanning for 45 years and above (when we split sample in two only), as we would have expected. Quite the contrary, respondents 45 and older use scanning slightly more often than those in the age group 45 years and below. This leads us to think that older people are more motivated and likely to follow instructions conscientiously, but a higher scanning rate in this group could also be due to household size and expenses, but this has yet to be studied. We hypothesize that respondents who scan more often might live in larger households and thus have many receipts with long lists of products, making scanning the best solution. We also know that respondents 45 and older follow instructions more conscientious and more often collect or are ready to collect receipts.

### *Lack of receipt*

Lack of receipts is the main reason for not using scanning. Collecting receipts, is for most respondents a long-forgotten habit<sup>6</sup>. More receipts are electronic, and many participants in the survey forget to ask for paper receipts. This seems to coincide with the type of expenses. Namely, smaller shops and small expenses such as public transportation, parking, coffee on the go or purchases from a booth do not provide (paper) receipts and several larger electronic- or sports equipment stores only provide

---

<sup>6</sup> From 2019 shops in Norway are no longer required to print out receipts automatically, but they are required to ask if their customers desire a paper receipt.

electronic receipts. While grocery stores are required to always ask if customers want the receipt.

From in-depth interviews we know that some respondents attempt to take screen shots of electronic receipts, but not all respondents are that explorative and test this option out. Presently, this is an option that works for the biggest grocery shops and some other stores, but not for most electronic bills. And as our machine learning becomes smarter, we will be able to recognise more and more of such scans.

### 6.3 Usability and user experience

Through user testing (see page 52) we try to understand the user, or respondent of the survey. And we attempt to understand the respondents' user journey<sup>7</sup> when they participate in our survey and use the app. We study their attitudes and behaviours to understand the cognitive process in how the respondent comprehend the survey and the survey questions, how they retrieve the information they need to answer, how they evaluate and calculate how to respond, and how they interact (formulate and select) their response or answer in the app (ref: The psychology of survey response, Tourangeau, 2000).

#### *Understanding the survey task*

From survey dropouts, we have learned about possible obstacles to the user journey. One important issue is that expectation to the survey is not met. Many dropouts find the survey, or task, more complex and time-consuming than they expected. Requiring data from all the household members and all types of expenses.

Very few of our respondents know or remember the answers to all questions in the questionnaire. In households with two incomes, the necessary info is most often shared between the two partners, and still, between them they don't always know the exact value asked. From in-depth interviews we find that respondents check with their partner, but only a minority check bank statements or agreements from service providers to get the exact sum that we are asking for.

It is also questionable whether all expenses for all of the household members are documented properly. The respondents we have talked to say that they understand this is part of their task, but they believe that many other respondents would need to have this emphasized. The latter being a signal that they might not have followed our instruction to the letter or fully been aware of it themselves.

We have seen that lack of paper receipts can cause dropout due to perceived response burden. Further, missing receipts when testing the app for first time is a problem, because then respondents do not immediately see how the scanning feature can reduce their response burden.

---

<sup>7</sup> In web design and usability testing we talk about the «user journey», which is the user experience a person has when interacting with a software like a mobile app.

### *Memory*

In our recruitment communication, we encourage immediate registration to avoid respondents forgetting to register the receipts. We also encourage collecting receipts, but for many respondents, it is hard to change a habit and remember this. Registration in-store rarely happens. Some respondents take pictures of receipts with their phone (not using the app) and wait with registration until they are outside the store, or they collect a few receipts and register them at a quieter time in the car, at home, or similar.

A few respondents attempt to scan screenshots of electronic receipts and invoices, but as stated earlier, few seem to be that explorative and test this option.

### *Labelling and grouping products together*

Registration of receipts is overall intuitive to understand how to do for most, but what to write when registering products manually is not. We have seen that respondents tend to group things together and report in categories such as lunch, clothes, and gas station, when they can, instead of registering each product. This has been detected as a major challenge to communicate to respondents through instructions, design or interaction design from the start of development and we see we still have not mastered it.

### *Confirmation of completion*

As stated earlier, confirmation of completion in the app is sometimes forgotten and a dropout reason. It has not been possible to automate, or set up push-warnings outside the app, if respondents don't actively confirm within a given time. But they receive a text message as a reminder that it is required to confirm. Respondents seem to think that SSB receive all data immediately and understand when they are done. Incompletes often state in interviews that they have done all they were asked and entered all required. Therefore, it is evident that the user journey in the app is not illustrated clearly enough for respondents to understand that they actively have to confirm completion.

### *Deadline for completion*

We note that 22,7 per cent of respondents changed their registration period in Q1, as the selected period did not suit them. Dropouts often say that if registration period could more easily be changed or moved forward, and if last date for completion could be

extended further, they would complete. But not all respondents start or complete even if the registration period is changed.

### *Usability*

#### *Logon and web app (PWA)*

When respondents receive SMS with link to the web app, most immediately try the link on their smartphones and start to test the app. They do not fully read the onboarding introductions, or the help texts. They go through fast and click around based on expectation from earlier experiences and intuition to see if they can “figure out” what they are supposed to do. To avoid dropout at this stage it is important that they do not meet any technical errors or cognitive barriers.

Unfortunately, both first logon required the first time, by Bank-ID or ID-porten<sup>8</sup>, and downloading the web app are obstacles. Logon is standard procedure for all bank and public services and works fine for most, but not all. Downloading PWA to the start screen is becoming more and more used over the last few years by online services in both public and private sector for services, shopping, and media sites in Norway. Still, the terminology and functionality seem unknown to many and how to handle it is not immediately intuitive to all, even if they have used/done it before.

#### *Registration and questionnaire*

That the app and questionnaire both work well with no errors, time delays, and problems is something the user takes for granted and is considered hygiene factors the app must meet. At present, the app has no major problems with registrations – neither in regard to interaction nor technical errors, as perceived by user. But that said, we know that it can be challenging to scan long receipts and that saving the receipts sometimes takes too long. We also know that there are more problems with old smartphones and some operative systems, but we did not register this as major problems.

---

<sup>8</sup> Both are a two-step certification login solution for public services on the internet. It is like an ID or passport and internet services can't be used without one. All citizens have one (if they want to) and using this gives the user a security that their personal data are handled according to law.



Quantitatively, we have also seen that we have problems with scanning reading of rebate, untrue minus values, quantity and volume, and correct sum. But so far, we have not registered many complains or obstacles for users because of this. Most respondents check briefly if the read-in sum seems plausible and leave it at that. We have not registered many attempts to edit or comment on scanning errors.

The questionnaire also works well technically and in use. We know that a number of users believe they are done with the survey when they have completed the questionnaire, and this is on our list for possible improvements as it drains interviewer resources to follow up the ones that forget to confirm completions.

### *Universal design*

SSB aspires to be sustainable and comply with universal design (UD) requirements for the public sector. Throughout development and testing, we have done user testing with a number of different target groups that we suspected would have either a lower cooperation and response rate, or that they might struggle with the app. We have tried to cover many subgroups in respect to age, gender, household size, household income, education, employment, not-native speakers, and urbanisation. A few subgroups, such as young men that are hard to reach, not-native speakers, elderly, people of limited digital skills, and visually impaired have been focused on in particular. Other groups on our list that we would have liked to study, is native speakers with weak reading and writing ability. We hope designing our data collection instrument according to SSB's design handbook and good practice for questionnaire design of survey instruments that many challenges these groups may have managed.

## 6.4 Data quality

The aspiration in official statistics is to provide survey data that is as accurate as possible. This is a demanding task, and it can be evaluated in many ways, both quantitatively and qualitatively. For the HBS we hope to compare our survey data with transaction data<sup>9</sup>. This is because we know the HBS survey is prone to many possible inaccuracies. SSB had hoped to get access to transaction data to evaluate the data quality, but access is yet to be fully granted.

Errors can be caused both by the respondent (Tourangeau, 2000), the survey instrument or the technical solution, and in the interaction between the two. For optical character recognition, poor image quality and misinterpretation of the receipt cause problems. For instance, we see that poor image quality can cause data to disappear; the image may be good, but the product line disappears; rebates and discounts can be misinterpreted; and other misinterpretation by scanning.

### *Scanning versus manual registration*

In the following section, we will assess plausibility and quality of the registrations for Q1 (N=903). We will look at the number of registered receipts, items, and total expenses to identify limitations and challenges for manual and scanned registrations.

We have observed that respondents who choose to use the scanning feature uses it in many ways and in ways not foreseen. For instance, in Table 6.2, we saw earlier that manual registration of receipts (52,2 Per cent) are slightly higher (4,5 percent) than scanning (47,8 per cent), while we would have expected the opposite. Looking at the total number of products registered, the numbers shift, and scanning (83,1 per cent) clearly outweighs manual registration (16,9 per cent) in Table 6.3 (also ref Larsson, 2022).

---

<sup>9</sup> We are thinking of receipt data from main store owners and bank transaction data from Nets. These kinds of data could be used as a valuable data training set for machine learning to improve data quality and to develop future data collection set ups.

*Table 6.3 Total number of products by scanning or manual registrations*

	Count	Per cent
Total number of products	51 436	100,0
OCR scanning	42 758	83,1
Manual	8 678	16,9

### *Negative values*

In Table 6.4 we see a very small per cent (0,11 per cent) of the receipts had a negative total value. 11 of 12 were registered through our scanning feature. We have seen that this could be true due to rebates, but it could also be untrue indicating limitations in image recognizing of receipt correct.

*Table 6.4 Number of receipts with total amount positive or negative <sup>10</sup>*

	Total		Manual		Scanning	
	Count	Per cent	Count	Per cent	Count	Per cent
No receipts	10 741	100,00 %	5 999	100,00 %	4 742	100,00 %
No positive total sum	10 729	99,89 %	5 998	99,98 %	4 731	99,77 %
No negative total sum	12	0,11 %	1	0,02 %	11	0,23 %

### *Untrue total amount*

Some of the datapoints that the scanning feature recognizes on a receipt are the price per product, the quantity and units per product, and total amount of the receipt. These variables allow us to estimate the correct price per product to compare the calculated total amount of the receipt with the amount from the scanning solution.

We checked the plausibility of the total amount of the receipt by multiplying the registered price per product by the quantity.

*Figure 6.5 Control check of the total amount recognized*

---

<sup>10</sup> Notice that the total number decreased compared to the Table 6.2 showing Total registered receipts. The reason for the difference is caused by a null value generated by the system. This stops some of the receipts from fully completing the recognition process, this bug should be fixed after publishing this article.

Price per item (x) quantify = Y  
 OCR registered total amount per item = X  
 X = Y = No difference

When comparing the scanned amount with the calculated total amount per receipts, 82,5 percent of the receipts were identical, but 17,5 per cent do not match.

*Table 6.5 Number of receipts differing between scanned and calculated total amount*

	Count	Per cent
Total number of receipts	10 741	100,0
No difference	8 862	82,5
Difference	1 879	17,5

The total amounts of the receipts differ between scanning and our calculation because the recognition of quantity, units and price per product is mixed up. Therefore, the system calculates an incorrect price per item as shown in Table 6.6.

In this example the variable 'amount1' is the registered total price for the receipt, "name\_items" refers to the product name, "qty1" the product quantity, "units" the unit of the quantity, "amount\_items" the item or product price, and the last column is the price per product we have calculated.

Table 6.6 Example of scanned receipt with amount\_item and calculated price per product

	amount1	name_items	qty1	units	amount_items	calculatedPricePrProduct
64891	676.0	ENTREC-TE S-LARSEN	1.00		99	99.0
64892	676.0	ENTRECETE S-LARSEN	1.00		99	99.0
64893	676.0	FISKEBURGER PURRE	1.00		59	59.0
64894	676.0	L. MELK 0,5%	16.90	L	1.75	29.6
64895	676.0	L. MELK 0,5%	16.90	L	1.75	29.6
64896	676.0	HJERTEFISKEKAKER	1.00		55	55.0
64897	676.0	HELMELK 3,5	1.00		19.4	19.4
64898	676.0	POMMES FRITES	1.00		27.4	27.4
64899	676.0	L. MELK 0,5%	16.90	L	1.75	29.6
64900	676.0	LA MIA SALAME	1.00		65.4	65.4
64901	676.0	LA MIA SALAME	1.00		65.4	65.4
64902	676.0	LA MIA PROSCIUTTO	1.00		69	69.0
64903	676.0	EPLER RØDE	0.91	kg	19.9	18.1
64904	676.0	FRØ- & KJERNEBRØD	1.00		40	40.0
64905	676.0	KETCHUP ØKOLOGISK	1.00		29.9	29.9

Looking at the products “L. MELK 0,5%” we can see that the values within “qty1” and “amount\_items” are switched and therefore the scanned total price for the receipt, amount1 (676 NOK), is wrong. Product price or amount\_item should be 16,90 and not 1,75. The last column shows the correct product price, which is 29,60 NOK, and not 1,75. This receipt failed our control check, as the scan mixed quantity, units, and price per unit or product. This is presently a source of error. In Q1, 17,5 per cent of scanned receipts possibly have errors in the recognized amount scanned (see Table 6.5), but in the future, we expect this will be reduced through machine learning.

Further, our control check of total amount of receipt has its limits, as it assumes registration or recognition of the quantity is correct. This is not always true as we can see in our next example from a manual receipt in Table 6.7.

Table 6.7 Example of manual receipt passing the check despite negative values

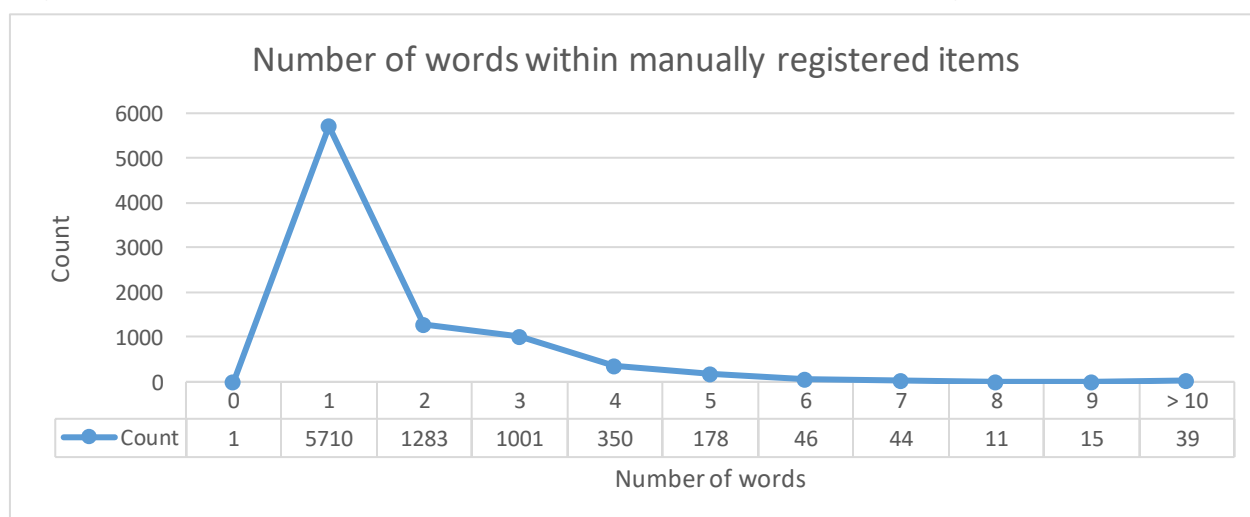
	amount1	name_items	qty1	units	amount_items	calculatedPricePrProduct
59625	-679.0	2 l lettmeik, 2 store youghurt, 6 egg, tomater...	1.0	stk	-529.00	-529.0
59626	-679.0	1 bok, 3 kopper 3 skåler, 3 asjetter	1.0	stk	360.00	360.0
59627	-679.0	Fysioterapi	1.0	stk	-510.00	-510.0

In this case, the respondent manually registered multiple products in one entry with a total amount for all the products, instead of registering each product separately. This case passed the check, but since our calculated price per product assume quantity is correct our calculated product price (-529 NOK) does not match total amount (-679,0 NOK) like we expect it to do.

*More than one product entry in one item*

The example just discussed (Table 6.7) brings us to another challenge regarding the data quality, namely, the registration of more than one product in the same entry. When registrations are done manually, we expect text entries with up to three to four words. We have checked all manual registration Q1 and found several registrations with more than five words within a product entry or item.

Figure 6.6 Count of words within one product entry for manually registered receipts



Manually entered items containing more than five words are difficult to classify into COICOP codes<sup>11</sup> which is the main purpose of the HBS. Some texts with five words or more could be an actual item, but most of the registrations with five or more words are for more than one product where the respondent has listed several products and registered the total amount paid for the products. Table 6.8 shows examples where the respondents have manually registered several items in one entry (see column with long texts Table 6.8).

*Table 6.8 Example of manual entries with several products in one item*

FBU_COMPLETE	MANUAL	1001036	...	NaN	0	appelsinjuice, vorterøl, rundstykker, suppebul...	1	NaN	stk	1.0	160.5	1.0	160.5
FBU_COMPLETE	MANUAL	1001038	...	NaN	0	Keflex(resept), Paracet,lbux, Micropore, Gum p...	1	NaN	stk	1.0	488.2	1.0	488.2
FBU_COMPLETE	MANUAL	1001044	...	NaN	0	Ansiktwipes, Store lomper, Wienerpølser, Smoot...	1	NaN	stk	1.0	199.5	1.0	199.5

To reduce these types of incorrect entries in the future, we plan to implement push warnings and focus on improving survey communication.

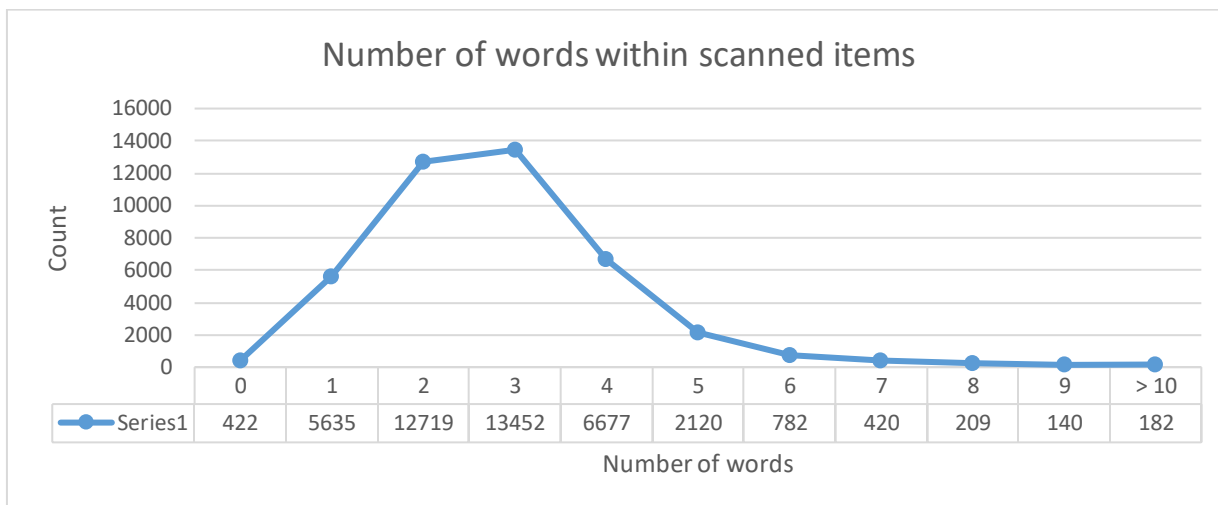
Entries with only one word are also likely to be entries with poor data quality. Entries registered with “miscellaneous” as product name and only the total sum of the receipt is in this category. These types of entries show us only the price and not product details, and they are therefore difficult to code.

We see the same pattern when it comes to items that are recognized through our scanning feature. Here, we also find a larger word count for one time. This is, foreseen since the system gives us more data regarding product name and it will more often be the correct name for one product.

---

<sup>11</sup> COICOP is a classification of individual consumption by purpose. It is developed by the United Nations Statistics Division to classify and analyse individual consumption expenditures by households. It includes categories such as clothing and footwear, housing, water, electricity, and gas and other fuels.

Figure 6.7 Count of words within one product entry for scanned receipts



Items containing more than seven words in a receipt are not successfully recognized by the OCR scans, example shown below in Table 6.9.

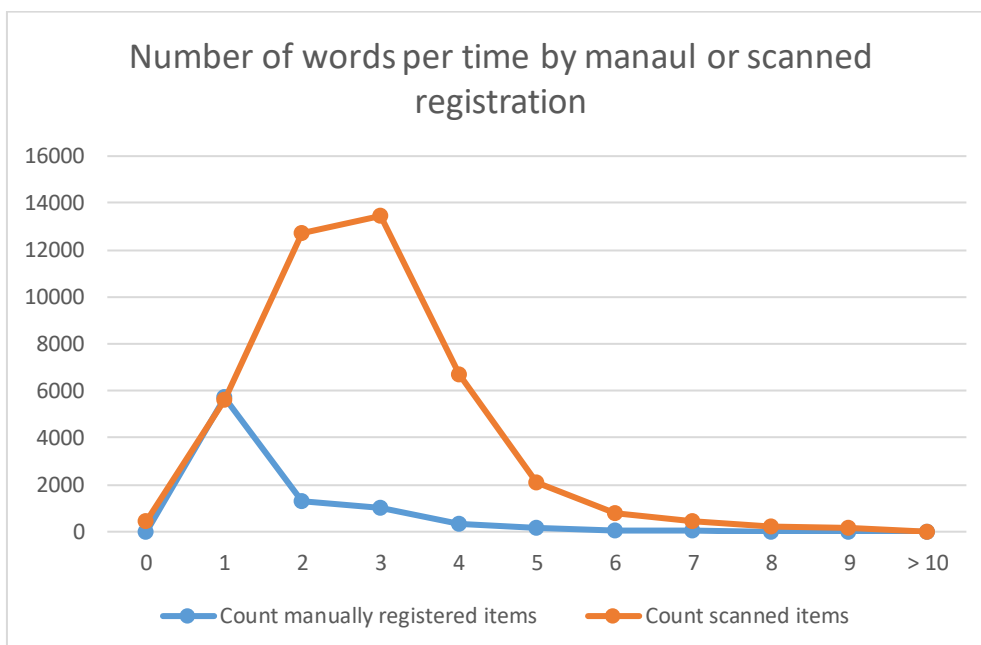
Table 6.9 Example of scanned entries with several words in one entry

ocrIndicator	items_id	...	id_items	idx	name_items	qty	searchTermId	units	referencePeriod1	amount_items1	qty1	amount1
OCR	1000241	...	146084750.0	7	GENERAL G3 NOS VOLT SUPER STRONG\ n990000000000...	1	NaN		2.0	79.9	1.00	260.80
OCR	1000254	...	170584671.0	0	Havregryn lettkokte økologisk 1kg axa\ n- 2 % T...	1	NaN	stk	8.0	36.9	1.00	192.07

In total, there are fewer items with potentially poor registration quality in receipts that are registered manually. However, there might be multiple or only one misleading item on a receipt that can distort the price per product and the total amount on the receipt. While there is a higher chance that scanned items has multiple words that refer to only one item, Figure 6.8, the high number of words per entry is correct, and the scanned registration is likely to have the right expense per item.



Figure 6.8 Count of words per item by manual or scanned registration



### Summary of quality check

In Table 6.5 we saw a low number of receipts with differing total amounts. This share is higher when we look at the accuracy of the registered products. The factors that might distort quality is the scanned reading of product name of various units, quantity, and for manually registered receipts respondents entering more than one product under a miscellaneous label.

In summary, there are drawbacks with both scanning and manual registration. But one should note that scanning is a system where the recognition capabilities increase with the number of scanned receipts, and we have yet only focused on the early stages of the data collection – Q1. Neither should we forget that number of incorrect receipts represents only a minimal per cent of the total.

The focus on data quality should also include the respondent since they are the users of the instrument. As mentioned before, the HBS is one of the few surveys where we give our data collection instrument directly to the respondent for self-completion. Thus, we guide the respondent through the week of registration with tips and help to ease the response burden. Our continuous communication with the respondent also improves the quality of the scanned and manually registered receipts.

### *User perspective on registration quality*

From our qualitative interviews and feedback to our interviewers and our Helpdesk we have some observation about quality of the data registered.

We have seen that several respondents group products together and provide categories instead of the product names when registering receipts manually. Therefore, labelling of fields for registration and interactive controls based on programmed logics; for instance, if more than X number of words or commas are used, app will ask if this is correct and remind about instructions for registrations to assist the user to register and label correctly.

### *Active editing by user*

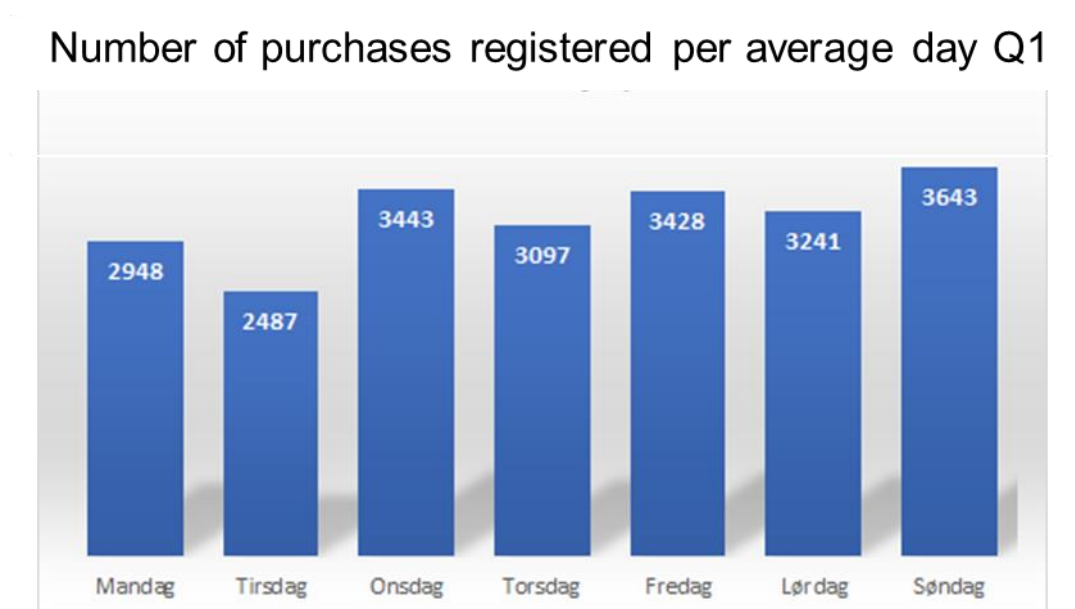
Users do not use the edit option much. When they take the picture of the receipt, they are asked to confirm that they want to use the scanned image. After the app has controlled the scan (a task the user can see happening), the user has to save the scanned receipt. We have seen from tests that most users look at what the scan has registered of products, amounts, price and total amount, the first few times, but do they not put much effort into control and editing to make sure it is correct. They assume it is not necessary. That means that if we want respondents to control and edit, logic controls have to be added to reinforce control and confirmation of correct product names, product numbers and weight/volume, product value, and sum of value of buy). We have not added logic to force editing control, as this would increase the response burden and not necessarily improve data quality. We believe working in-house with big data analysis, control checks, and editing is a better way to improve data quality.

### *Self-evaluation of registration*

From in-depth interviews we have learned that the respondents tell us they collect expenses and register when they have a break in their day. This break is usually before or after work, on public transportation, at home, after they have put their children to bed or when the day is slowing down. Respondents do not register receipts continuously throughout the day or when they are shopping. They wait for a quiet and undisturbed moment in their day.

From similar surveys, we know it is important to register receipts soon after the purchases to avoid forgetting smaller or more tedious expenses. From our data, we see that the respondents do a fairly good job registering purchases every day, see Figure 6.9 below. We see that they are more active towards the end of the week. This pattern can be a reflection of more purchase being done by the end of the week or respondents trying to get finished. Be reminded that we have poured interviewer resources by calling and sending SMSs to get this result, see chapter 5 page 5. It might not be sustainable going forward.

Figure 6.9 Number of purchases per average Monday - Sunday, Q1



*Impact of interactive tool?*

An additional important topic is what effect the instrument and the interaction between respondent and instrument has on data quality. What kind of interactive instrument effects do we see? We are thinking both about use of pre-coded lists with product names and categories, personalised feedback in app based on machine learning and or interaction between the respondent and the instrument/app. Unfortunately, we can't say much about the impact of "smart" or interactive tools such as a mobile app after Q1 except that our qualitative work shows that respondents

expect smart solutions used when using an app. We hope to study this topic further later.

## 7. Discussion of findings

---

We will now proceed to summarise and try to answer our two main research questions.

### 7.1 How does use of a smart survey approach impact participation?

#### *Participation*

A key question is if use of a smart survey approach or app will exclude or include subgroups we generally struggle to recruit, namely, young men, non-native speakers, and people with lower education. We are also concerned that a smart survey approach will exclude groups with limited digital skills<sup>12</sup> and disabilities. Is the use of a smartphone and image scanning an additional contributing factor causing them to refuse or drop out?

We do not have many quantitative indicators for the subgroups mentioned earlier (except for age). Nevertheless, we have gained qualitative insights indicating that a smart survey approach is not a problem in the recruitment phase, nor a reason for respondents to drop out, except among the oldest age group.

From assessing the net sample, we know that there is an underrepresentation of contact persons aged 67 years and older. This imbalance could be due to the recruitment procedure that selects for the youngest in senior-households, but we can't rule out that it is because contact persons are required to use an app. In our qualitative investigations, we have spoken to several 70 years and older that wanted to participate and had no problem using an app, but this might not be the case for most people above 70 years. Therefore, we suggest taking a better look at the older age group, 67 years and older, and investigate more closely at what age the drop in participation and dropout occurs. Consequently, we can consider whether we need an alternative data collection method for them, as it would be considered controversial to exclude older people from the target group.

---

<sup>12</sup> Only 3 % of the population are "non-users" of internet. The share of non-users' rise from retirement age and is 30 per cent for 80 years+. We also should consider that not everybody feels skilled in using new digital solutions. 15 per cent says that their digital skills are lacking with increasing shares from retirement. We know that some of the oldest cannot cope themselves with tax report, banking, health care appointments etc and have their relatives or friends sign in with their bank ID to handle necessary cores. This is hard to track, and we do not have figures for this.

### *Response burden*

Usability tests show that using smartphones and scanning ease the response burden for most respondents. The biggest challenge for initial recruitment and later dropouts is a demanding survey task, not the app. This is something we have detected throughout our qualitative testing. We see that groups we struggle to recruit, such as groups with lower education, often do not know SSB and what we do and hence are not as cooperative as other respondents. Thus, they often refuse to participate or complete the tasks. Whereas the very oldest often know us but excuse themselves due to age or personal reasons.

All things considered we conclude that the use of a smart survey approach does not seem to be a problem for participation. However, the response burden seems to be the main barrier to participation and will probably be so regardless of the data collection approach.

Going forward we want to emphasise the advantages of smartphone scanning to respondents during our recruitment. And suggest to the respondents that they have a receipt available when they first try the app to see the full potential of scanning. In addition, interviewers will encourage the respondents to open the app while on the phone. We have also considered a short instruction video, but our impression is that respondents tend to skip instructions.

## 7.2 How does use of a smart survey approach impact data quality?

Use of smartphone features, OCR scanning, and interaction between respondent and instrument could be of great help to respondents' memory and ability to report household consumption as accurate as possible, both in respect to number of purchases, value of purchases, and product categories. Therefore, ensuring better data quality than a traditional questionnaire or diary solution has been able to do. This alone speaks for the advantage of scanned data. The question is if scanned data provides higher or better data quality.

In our first quality checks, we identified some problems with both scanning and manual registration that required editing. We have seen in the data that scanning can create

incorrect rebates, unit entities, and duplicates. Our most considerable worry regarding manual registration is that the respondents tend to group together items in one entry. Grouped entries are difficult to code and are therefore an issue for data quality as they have to be resolved and edited before we can produce statistics.

Presently we do not have a complete overview of which registration method has the most sources of errors or more severe errors. But due to the volume of products (i.e., “big data”) registered by scanning, it is likely that scanning has the biggest source of error, and it will be a large job to edit manually. But with access to transaction data and machine learning quality it can be drastically improved. And if so, quality of manual registration is no match for scanning.

We know that most users find self-completion<sup>13</sup> in an app on a smartphone the best way to collect diary-based surveys. And we know that using the scanning feature simplifies and reduces the task and enhances the amount and quality of data. Unfortunately, it is not used as much by respondents as we expected. And surprisingly, less in the younger groups that we assumed would possess the necessary digital skills. This is because they lack paper receipts. In this way, the smart survey approach does not distinguish itself from other data collection methods for younger groups but consolidates existing patterns.

### *Usability and quality*

We have seen that some respondents struggled more with logon than expected. They did not understand how to handle a web app. We have also seen that we have not communicated well enough the advantages of using scanning on the smartphone. Except for this, the app is easy to use for registration of expenses and questionnaire completion.

The survey task, on the other hand, is a challenge, as it is perceived by respondents, particularly younger ones, as a demanding task. Several respondents do not know from the top of their mind the correct answer to questions, such as the amount of fixed

---

<sup>13</sup> Survey questions about income, ownership and purchase can be sensitive and therefore considered to be better suited for self-completion. Without the interviewer it is expected that respondents provide more true data and less socially acceptable answers – at least if the respondent know the answer without assistance.

invoices and large expenses, and therefore tend to estimate or satisfice<sup>14</sup>. This is a qualitative issue for data quality.

We had hoped that using smartphones would help respondents register sooner after their purchases and therefore they would depend less on memory. We also expected that since the mobile phone is always available, it would make it easier to record all household purchases and not miss small purchases. This happened to a limited degree, but not to the extent we expected. Too many seem to have chosen an easy way out and grouped products together and not been industrious in getting all the purchases by all members of the household included.

We know from our qualitative work that the respondents find the app, personalised feedback, and smart interaction in the app useful, and something they expect when using a mobile app. Presently we do not have quantitative data can tell us anything about the possible instrument effect this might have, but it is something we would like to follow up in later studies.

We also could need more analysis to conclude whether the use of scanning and smartphone improve the data quality. It increases data points with errors that need manual editing. But evidently, the bigger number of data reduces the impact of errors in recognition. And with machine learning, we expect manual editing can be reduced and the quality to keep improving, but we need respondents to use scanning more.

Today we use a high number of interviewers and hours supporting respondents to complete the survey, mainly with logon and completion issues in the app. In this work pushing for scanning should be added. The move from interviewing to support in data collection was expected, but we do not think the necessary volume is sustainable going forward. We can improve survey communication, app design, and SMS nudges to promote a better experience for respondents through the survey journey.

### 7.3 Going forward

In conclusion, a smart survey approach can alleviate response burden and possibly improve data quality. By continuing to study new data sources and ways of utilizing

---

<sup>14</sup> Satisficing is when respondents don't put as much effort into answering a survey as they should.



new data, like use of transaction data and machine learning, we may reduce the need for survey data and reduce the response burden even further.

For response rate and representativity, it does not seem to matter much one way or the other, except for the oldest. For respondents, their perception of what is the easiest is the key, and for most, it is to use the smartphone. Whether scanning or manual registration is easiest depends. If the volume of purchases is high, like for bigger households, scanning seems to be preferred, but for single homes with fewer purchases, manual registration often seem to be preferred from what we have learned from in-depth interviews.

With or without the app, our biggest challenge continues to be the high response burden and convincing groups that are hard to recruit to participate. With access to transaction data, we can validate our HBS data and continue to develop the smart survey approach to reduce response burden and better future statistics. To gain this opportunity we need to move forward with care and protect respondent's privacy as well.

## 8. References

---

Berg, N. Egge-Hoveid, K.; Holmøy, A.; Sandnes, T. (2019). "Explore the reuse of new digitalised multi-mode design developed for the Household Budget Survey in a new and modernised Time Use Survey." Deliverable for Eurostat grant agreement no.: 07141.2017.001-2017.358, Action I: INNOVATIVE TOOLS AND SOURCES FOR TIME USE SURVEYS.

de Groot, J., Oerlemans, T., Rodenburg, E., Schouten, B., Mariscal, A. C. L., Bernia, E. M., Poch, J., Garcia, T., H., Criado, V., C., Gauche, C., Osier, G., (2022): "Deliverable 2.1: Thematic paper on the Consumption pilot (WP2.1)", ESSnet Smart Survey.

de leeuw, E., et al. (2018). "International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data under a Creative Commons Attribution 4.0 International License (CC BY 4.0) International License (CC BY 4.0).".

Jäckle, A., et al. (2022). "Increasing Participation in a Mobile App Study: The Effects of a Sequential Mixed-Mode Design and In-Interview Invitation." Journal of Survey Statistics and Methodology.

Larsson, M. R. (2022). «En rask titt på varelinjene fra dagboken», Internal paper Statistics Norway.

MIMOD project 2018: [https://ec.europa.eu/eurostat/cros/content/essnet-project-data-collection-social-surveys-using-mixed-modes-dcss-annemieke-luiten-and\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-project-data-collection-social-surveys-using-mixed-modes-dcss-annemieke-luiten-and_en)

Mussmann, O. and Schouten, B (2019). "Final methodological report discussing the use of mobile device sensors in ESS surveys. Mussmann and Schouten, 2019 and description of framework for smart survey.

Ricciato, F., et al. (2020). "Trusted Smart Statistics: How new data will change official statistics." Data & Policy 2: e7.

Sensor data for ESS surveys: a first inventory." Eurostat: MIXED MODE DESIGNS FOR SOCIAL SURVEYS - MIMOD. GRANT AGREEMENT FOR AN ACTION WITH MULTIPLE BENEFICIARIES. AGREEMENT NUMBER – 07112.2017.010-2017.786 WP5 – Deliverable 3.

Smart Survey deliverables, [https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys_en)

Struminskaya, B., et al. (2021). "Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Nonparticipation Bias." Public Opinion Quarterly 85(S1): 423-462.

## 9. Appendix

---

### Attachments:

- 9.1 ESSnet projects on new data and data collection tools
- 9.2 Digital development in Norway
- 9.3 HBS survey setup in Norway
- 9.4 Response rates
- 9.5 Non-response
- 9.6 Qualitative data HBS
- 9.6 Short about user testing
- 9.7 Topic of contact reasons at Helpdesk support HBS

## 9.1 Essnet projects on new data and data collection tools

In the last decade the EU has financed a number of grants to explore use of new tools and data sources and modernise data collection in official statistics to meet declining response rates and secure data quality. SSB has participated in several of these grants. For this paper deliverables from the mixed mode data collection in social surveys – MIMOD<sup>15</sup>; Innovative tools and sources for diary-based data collection for the Household Budget Survey (HBS) and the Time Use Survey (TUS); and the Smart Survey grant<sup>16</sup> is important, together with the knowledge sharing within the network of national statistical institutes in Europe.

## 9.2 Digital development in Norway

Over the last two decades in Norway almost all public services have been transformed to electronic or digital services. Public sector has been leading the way and few services has an option to digital solutions. Today Norway has close to full internet access, and 9 of 10 use internet on an average day in Norway<sup>17</sup>. Almost all citizens have their own smart phone and use it daily<sup>18</sup>. And this is the same throughout age groups up till end of work life (mid-60s). For 65-74 years daily use is 84 per cent, dropping to 67 per cent for 75-79 years.

## 9.3 HBS survey setup in Norway

A short description of set specifics:

---

<sup>15</sup> “Mixed mode” in data collection is combining different ways (modes) of collecting data for a single project. Like presently a social survey can use both telephone interviewing and web survey for the same questionnaire. Often data collection start with web survey (the cheap data collection method) and end up with telephone interviewing (the expensive method) for the respondents without cooperation online.

<sup>16</sup> See Trusted Smart Survey: [https://ec.europa.eu/eurostat/cros/content/trusted-smart-statistics-towards-european-platform-trusted-smart-surveys\\_en](https://ec.europa.eu/eurostat/cros/content/trusted-smart-statistics-towards-european-platform-trusted-smart-surveys_en) and ESSnet Smart Survey deliverables, [https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys_en).

<sup>17</sup> 93 percent use internet on an average day in Norway, 9-79 years, Norwegian media barometer, SSB 2021.

<sup>18</sup> 99 percent has their own smartphones, 16-79 years. And 93 percent use smartphones on an average day, Norwegian media barometer, SSB 2021.

### *Survey task*

Selected households must complete three tasks:

- Start interview, approx. 10 minutes
- Registration of purchases and other expenses for one week in the consumption app
- Completion of the questionnaire in the app, approx. 20 minutes
- Gift certificate of € 50 to each household who completes all the tasks

### *Registration period*

- The registration period is assigned in the initial interview
- This coming Monday automatically comes up as the first choice
- The period can be moved up to five weeks in advance
- The registration period can be changed after recruitment if the household does not start registration as agreed

### *Household sample*

- 12,000 households where at least one person is aged 18 to 84 years
- The sample is divided into 52 equal periods (note: no registration periods)
- The sample is drawn from Statistics Norway's diet register
- Reference person is selected according to given criteria (the person contacted for recruitment)
- It is the reference person's household that must participate

### *Contact person (CP)*

Determined in the initial interview. This must not be the reference person, but should be a person with an overview of purchases and expenses in the household

The contact person has the main responsibility for registering all expenses in the household

#### *Team of interviewers*

- Eight interviewers, including super user
- Recruitment and follow-up during the registration period
- Weekly meetings with planner
- Close dialogue on Teams

## 9.4 Response rates

### *Response rates*

We have achieved a net sample, or a response rate, of 30,2 per cent and 69,8 per cent non-response (including partial non-response) <sup>19</sup> for HBS 2022/Q1, see

---

<sup>19</sup> Response or completion rate is the share of gross sample that have registered or stored at least one receipt and answered the questionnaire. We divide non-response into four categories: Partial non-response, refusals, other non-response, and no- contact. Partial non-response completed the start interview but did not start using the app or start registrations in app and/or answering the questionnaire but dropped out before they met completion requirement. Refusals are those reporting to us that they are not willing to participate, either to an interviewer or through Statistics Norway's information service. The no-contacts are the respondents that we have not been able to get in touch with. They have neither picked up the telephone, nor clicked on the link to the app, or been in contact with the information service. The category other non-response are not direct refusals, but households that for various reasons that can or should not participate, for example if contact person is ill or temporarily unavailable.

Table 9.1. Non-response made up of 24,3 per cent refusal, 28,2 per cent no-answer, 7,7 per cent other non-response, and 9,5 per cent partial non-response.



*Table 9.1 Response rates HBS 2022/Q1 in counts and per cent*

<b>Response</b>	<b>Count</b>	<b>Per cent</b>
Sample	2 993	
Natural fallout	7	
<b>Gross sample</b>	<b>2 986</b>	<b>100,0</b>
Refusal	726	<b>24,3</b>
No-answer	843	<b>28,2</b>
Other non-response	230	<b>7,7</b>
Partial non-response (did not complete)	284	<b>9,5</b>
<b>Net sample or response rate</b>	<b>903</b>	<b>30,2</b>

We are not comparing response rates with 2012 as data collection method, definition of completion, and conditions have changed and are not comparable. But Table 9.2 indicates a drop of almost 20 per cent. This is no surprise as we know that response rates from a decade ago is no longer achievable for sample survey, no matter setup and data collection method. Still, a response rate of 30,2 per cent is lower than we had hoped for and what we normally could expect to achieve on similar web or phone surveys, right enough without the registration task.

*Table 9.2 Comparing response rates HBS 2012 and 2022/Q1 in per cent*

<b>Response</b>	<b>2012</b>	<b>2022/Q1</b>
<b>Gross sample (N)</b>	<b>7 000</b>	<b>2 986</b>
Total non-response	42,7	<b>60,2</b>
Partial non-response	8,4	<b>9,5</b>
<b>Net sample or response rate</b>	<b>48,9</b>	<b>30,2</b>

Partial non-response, recruited and completed the start interview but did not start using the app or started, but did not complete, is 9,5 per cent of gross sample. We also note that of a total of 1 340 recruited 206 (15,4 per cent) never started. This dropout is an indicator of response burden. In 2012 partial non-response in total was 8,4 per cent. Requirements for completion is not identical but at the same level.

## 9.5 Non-response

Non-response, total and partial, is 2 083. 40,5 per cent of non-respondents were no-answers, 34 per cent is direct refusal and 13,6 percent partial non-response. The most important other reasons given, but without important frequency, is language trouble (4,1)<sup>20</sup>, illness (2,2) and no internet access or device (2,1), see Table 9.3.

*Table 9.3 Reasons for non-response HBS 2022/Q1*

<b>Reasons for non-response</b>	<b>Count</b>	<b>Per cent</b>
Non-response (total and partial)	2 083	<b>100</b>
Partial non-response	284	13,6
No answer/no telephone-number	843	40,5
Do not want to participate	709	34,0
Lack of time/availability	15	0,7
Language	86	4,1
Ill (long & short term)	46	2,2
No internet access or device	44	2,1
Other fallout	45	2,2
Temporarily absent	9	0,4
No consent (GDPR)	2	0,1

Language naturally is a barrier for not native speakers (immigrant population is about 15 per cent population), but we have no population figures on language problems. From second quarter English was added to web app and help to reduce this. Lack of time is more often given by people in mid-life with family or work commitments. And lack of device or internet access occurs more often in groups outside work life

---

<sup>20</sup>

(retirement age starts at 62, but traditionally it is 67 year) and lower educated. Illness is given as reason by 45 years+ (note that 2022/Q1, had high occurrence of covid-19 in Norway).

## 9.6 Qualitative data HBS

To develop our web app, survey set up, recruitment strategy, and overall survey communication we have done 101 user tests, cognitive interviews, in-depth interviews and focus groups with different target groups and topics since we started development of the HBS app fall 2020. And after 2022/Q1 we have done an additional of 25 follow-up interviews (see table below). We also have insights from interviewer feedback and our Helpdesk.

*Table 9.4 Overview qualitative tests HBS*

BRUKERTESTER - H2020-V2021							
Test nr	Når	Type test	Tema/fokus	Målgrupper	Antall	Gjennomført	Kommentar
1.	Uke 45	Hurtigtest	- Sammenlignet to konsept (prototype)	Kolleger utenfor teamet med mix av: - HS-str (m/u barn) - Alder/ livsfase (jobb/ikke jobb - permisjon)	8-10 intervju	9 intervju	Kom i tillegg til plan opprinnelig plan
2.	Uke 50_51	Hurtigtest	- Reg av kjøp i prototype (funksjonalitet)	Snowballing eksternt	9-12 intervju	9 intervju	Færre tester enn planlagt (Esben var syk)
3.	Uke 04	Brukertest	- Test av app v. 1 av (mobil-device)	Rekruttere eksternt: - Ikke-digitale/godt voksen - Ikke norsk som morsmål	10-15 intervju	14 intervju	
4.	Uke 11-12	Brukertest/fokusgrupper	- Oppgaveforståelse, unge & eldre grupper	Sneballrulling på SSBs FB 19-73 år (unge/unge voksne/voksne/godt voksne)	4 fokusgrupper	4 FG, 19 intervju	
5.	Uke 15 14-15.04.21	Hurtigtest	- Prototypetest av 2 alt. hjemskjermer	Sneball, kollegaer utenfor teamet, familie eller venner	8-10 intervju	11 intervju	Pluss ekspertevaluering fra Dag og Kari Anne
6.	Uke 25-28/2021	Gerilja-grupper pilot	- Brukervennlighet - Motivasjon  - Brukerreisen	Deltakere m/ulik status fra piloten	4 fokusgrupper	17 intervju	U25-33. (Ferie 28-32) på Teams U25: 8 intervju (6 individuelle + 1 par), startet ikke fullført U26: 4 indiv intervju, flerpers/60 år + - fikk ingen lav utd, ingen under 30 år, fullførte/påbegynt U27: 5 indiv intervju, under 30 år, flerpers, fullført
						<b>70 intervju og 4 små FG = 79 personer totalt</b>	

BRUKERTESTER - H2021							
Test nr	Når	Type test	Tema/fokus	Målgrupper	Antall	Gjennomført	Kommentar
1.	Uke 39-41	Brukertest brukerreisen	- Test av brukerreisen i nytt design med ny onboarding og bruk av tips fra datafangst og i appen. Testet med prototype	Rekruttere eksternt fra ssb.no-panel: - Tversnitt fra 19 år - Over 60 år	10-15 intervju	12 intervju	Viste seg at flere av deltakerne var eksperter som jobber med kommunikasjonsarbeid (f eks info.sjefer ol)
01.jan	Uke 41	Ekspertevaluering	Frode Larsen, komm Maciej Warchal, komm Anne Marti Pettersen, datafangst	Fokus på brukerreisen og på interaksjon			Finns to rapporter og opptak av debrief
2.	Uke 48	Siste brukertest	Forståelse og funksjonalitet med forberedende hjemmeoppgave	Ikke norsk som morsmål, 70 år+, Nedsatt syn, behov for hj.midler utover vanlig brille, 30-50 år, flerpers-HH (kontrollgruppen)		10 intervju	
						<b>22 totalt</b>	
BRUKERTESTER - V2022							
Test nr	Når	Type test	Tema/fokus	Målgrupper	Antall	Gjennomført	Kommentar
1.	Uke 18-20	Oppfølgingsintervju	Frafallsgrunner	Rekruttert, men ikke startet Startet, men ikke fullført Fullført, men mye manuelt	20-30 intervju (grupper greide vi ikke å rekruttere til)	25 intervju	
						<b>25 totalt</b>	

## 9.6 Short about user testing

In user testing we test functionality, user experience, and cognition process. We test both the usability of the app itself (with registration of expenses and the questionnaire) and the cognitive process of participating in the survey from recruitment to completions. Our technique is focused on the respondent, as we call the “user” of a survey, or user-centred as tech developers and designers in user interaction (UI<sup>21</sup>) and user experience (UX<sup>22</sup>) would say. Through the respondents we aim to gain insight in the user experience.

## 9.7 Topic of contact reasons at Helpdesk support HBS

What does topic of contact from respondents with Helpdesk tell us about respondents' experience with the app and HBS survey?

From the Helpdesk we see that the general number of contacts for HBS 2022 are somewhat higher than for other web surveys, and in particular for technical issues, see Table 9.5 below. We do not find this level alarming, as it has to be expected when

<sup>21</sup> User interaction is about reflecting the real user experience after interaction with the user interface.

<sup>22</sup> User experience (UX) is how a user interacts with and experiences a service or a product. It includes the user's ease of use and perception of usability, efficiency, and utility. User experience is subjective quality.

using a web app/smart survey. We note that type of technical issues is mainly log-in problems and will add insight to this topic under section for usability.

We see that most of the contacts is about participation in one or another form and the task at hand; the survey in general; wish to change registration period; and explanation why respondents cannot participate. This contact reflects that many respondents are committed to SSB and have ready to contribute.

We do not have many direct complaints in general about SSB or the survey, but indirectly non-corporation and -response rates tell us something about how well the survey is received. But as data collection is pouring in more contacts and resources in in this survey than normal, comparing corporation and response rates is difficult.

We note that we do not have many direct contacts about GDPR or about safety of opening link from us, but we see confirmation of sender as a related issue. It is not a topic of high contact numbers at the help line, but still ranked forth on our list, and always a topic our survey communication wants to reassure our respondents about. That is about who sender is, how we got their (register) data, and how we treat their data to maintain the high trust we have in public opinion.

## 9.1 Tables

*Table 9.5 Main reason for contact with Helpdesk support HBS 2022/Q1*

<b>Main reason for contact with Helpdesk support</b>	<b>Count Contacts</b>
Want an appointment	208
Do not want to participate	146
Questions about the survey	127
Help to logon	33
Refusal	30
Help to register/answer questions	30
Illness	22
Questions about the questionnaire	18
Confirmation that sender is SSB	15
General complaint	13
Change of period	12
Lack of time/Availability	9
Want to do the interview now	8
Feedback on questionnaire quality	4
Language barriers	3
Privacy/data source for person data IO	2
IO live in institution	2
Delete interview/information on IO	1
No device or internet access	1
IO migrated/Live abroad	1
Dead	1
Complaint number of contacts	1
Contacted, but already answered	1
Complicated case (difficult to log one main reason)	4
Other	31
<b>SUM</b>	<b>723</b>