ESSnet Smart Surveys

Grant Agreement Number: 899365 - 2019-DE-SmartStat

Link to our CROS website

Workpackage 2

Smart Survey Pilots

Deliverable 2.1: Thematic paper on the Consumption pilot (WP2.1)

Version 3.0, 13-04-2022

Prepared by:

Jelmer de Groot, Tom Oerlemans, Evelien Rodenburg, Barry Schouten (CBS, Netherlands)
Ana Clara Lope Mariscal, Encarnacion Martin Bernia, Jara Poch, Tamara Horcajo Garcia, Victor Balsa Criado, (INE, Spain)
Catherine Gauche, Guillaume Osier (STATEC, Luxemburg)

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

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 the phase 2 (functional and usability) or, if applicable, the phase 3 tests (methodology and infrastructure field tests) for each pilot. This deliverable 2.1a is about the consumption pilot with the Household Budget Survey app in three countries, ES, LU and NL, from September to December 2021. The focus in this report is on the plausibility and richness of purchase data for HBS statistics using an app-assisted smart survey approach. In deliverables 2.3 and 2.4, the role of, respectively, incentives/gamification and interviewer-assistance in the HBS phase 3 test are discussed in detail. Our main conclusion is that the app-assisted HBS provides diverse and plausible purchase data.

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 3 pilots are focussed on understanding how to further smart surveys to effective recruitment and motivation strategies and how to find the right balance in active-passive engagement of respondents. Phase 2 functional and usability tests form the stepping stone to such field tests and are a guarantee that technically and conceptually smart surveys should deliver adequate data. Phase 3 tests inform on how to operationalize methodological and logistical choices. For WP2.1, Consumption, a phase 3 test was conducted using the Household Budget Survey app.

The HBS is a survey that clearly fits smart survey eligibility criteria; it is a time-demanding survey which requires detailed recall over a longer period of time from two weeks up to a month. It is considered time-demanding and response rates are the lowest of all official surveys. Candidate smart features are advanced product search algorithms, receipt scanning, in-device receipt scan text recognition and processing, uploading of digital receipts, linkage to public points-of-interest data, and donation of bank transactions data. In the HBS app the first three features have been implemented and extensively tested.

The Household Budget Survey app has been developed within Eurostat-funded projects @HBS (Jan 2019 to February 2020) and @HBS2 (March 2021 to April 2022). In ESSnet Smart Surveys the app has been modified in two stages to six new countries: BE, DE, ES, LU, NO, and PL. The app already existed in country versions for FI, NL, SI and UK. The six countries performed functional tests within the ESSnet during May 2020 to December 2020. Revisions of the app, based on these phase 2 tests, are reported in deliverable 2.9. on WP2 specifications for WP3. Country differences and commonalities are reported in deliverable 2.5 on shareability. Three countries, ES, LU and NL, participated in the phase 3 field test and for these countries the app went through another modification stage between Jan 2021 and September 2021.

Underlying to the phase 3 HBS field test are the following questions as listed in the ESSnet project proposal:

- How does an app-assisted HBS affect recruitment rates?
- How does an app-assisted HBS affect drop-out/completion rates?
- How does an app-assisted HBS affect data quality?
- How does an app-assisted HBS affect substantive plausibility?

These four questions are further elaborated for the role of interviewer assistance and respondent statistics feedback in deliverables 2.3 and 2.4, respectively.

This report is structured as follows: First a description is given of the phase 3 field test design. Next, the two main purchase data entry options are elaborated and linked to the field test design. Subsequently, criteria and metrics are defined for evaluation of data diversity and richness. Fieldwork results are then presented. The report is completed with a discussion, a summary of lessons learned and recommended next steps.

2. DESIGN OF THE HBS FIELD TEST

2.1 The HBS app

The phase 3 field test employed version 2.1.15 of the Household Budget Survey app available in Apple and Google app stores. The app could be used on smart phones and on tablets in portrait mode, but not on desktops and laptops. Technical performance was guaranteed for operating systems from 2016 on. The app version 2.1.15 included a number of user interface revisions suggested by functional/usability phase 2 tests within ESSnet Smart Surveys as well as a number of extensions prepared within Eurostat-funded project @HBS2. The functional/usability tests were accompanied by a short questionnaire to all WP2.1 countries (BE, DE, ES, LU, NL, NO, PL) on the desired frontend and backend specifications. The questionnaire and country answers are presented in Annex A. Country specifications were grouped into three categories: shared by all countries, shared by the majority of countries, and shared by a minority of countries. Details are reported in upcoming deliverable 2.5 on shareability.

The revised app features are:

- In-app tutorial movie to instruct respondents on how to scan receipts

- Adjustment of parameters in the product search algorithm (using a Jaro-Winkler distance function)
- Revision of discount and rebate data entry options
- Option to turn on daily reminders
- Textual changes in some of the UI screens

The new app features are:

- Automatic receipt scan detection
- The option to crop receipt scan images
- In-app text recognition including editing options by respondents
- Inclusion of paradata on in-app navigation behaviour
- A configurable in-app intro questionnaire

During the preparations for the field test, it was decided not to display in-house text recognition and COICOP[1] classification results in-app to respondents. The main reasons were the complex backend database structure that would be needed to keep track of unchecked, checked and revised receipts and the burdensome and complicated additional respondent classification task.

2.2 Experimental conditions in the field test

Three experimental conditions were included in order to evaluate push-to-app strategies and active-passive data collection trade-offs:

- Yes/no instant feedback of individual household expenditure statistics; this feature is discussed in detail in deliverable 2.3
- Yes/no interviewer-assistance in recruiting and motivating respondents; this feature is evaluated in detail in deliverable 2.4
- Yes/no in-app respondent editing of text recognition

The three conditions were randomized independently, i.e. the potential interaction between the conditions was not deemed influential enough to evaluate and is ignored. It would also imply a larger sample size to reach acceptable statistical power The in-app edit option implied that only the edited data was submitted to the backend. However, through paradata logging it can be deduced afterwards whether respondents actually performed editing. In all cases the scanned receipts were submitted to the backend, i.e. regardless of in-app editing. To safeguard privacy, respondents were instructed to crop images to the product-price section on receipts.

2.3 The field test sample design

For various reasons, BE, DE, NO and PL decided not to participate in the phase 3 field test. The anticipated sample size of 4000 households was, therefore, allocated to the three remaining countries, ES, LU and NL. The original anticipated sample size per country was 800, assuming five participating countries. LU and NL agreed to double their sample sizes in order to meet statistical power requirements. The resulting sample sizes were 1600 for LU and NL and 800 for ES. Ultimately the

[1] COICOP is short for Classification of Individual Consumption by Purpose. It is published by the United Nations Statistics Division

sample size for LU was slightly larger, because the amount of administrative nonresponse was smaller than anticipated. The sample size for NL was slightly larger as the workload for face-to-face interviewers had to be slightly reduced because of COVID-19 procedures.

Table 1 presents the sample sizes per country and condition.

The samples were randomly split to the feedback condition. One half got instant feedback. Under this condition, feedback was explicitly mentioned in invitation letters and interviewers could use it as an argument at first contact. The other half got feedback on the last reporting day and no mention was made in invitation letters or by interviewers. For these households the app Insights screen displayed a so-called empty state layout indicating that feedback would only be delivered later during the reporting period.

The samples were also randomly split to the interviewer condition. However, in LU the interviewer condition implied that interviewers were mentioned as contact persons in invitations. Interviewers did not actually visit or call households. The non-interviewer condition in LU meant that no mention was made of interviewers. In ES and NL, one half of the sample received only a written invitation including a brochure and instruction material. The other half was visited at the door by an interviewer and a motivational call was made a few days later. Annex B presents recruitment material for ES, LU and NL.

The in-app condition was only in part randomized, since it was not planned in the original design and was not enforced.. In ES, in-app editing was always enabled, while in LU it was always disabled. In NL, in-app editing was enabled for the interviewer sample and randomized for the non-interviewer sample. The choice not to randomize within the interviewer sample was made in order to not confuse interviewers when interacting with households. While not randomized within countries, the total samples with and without in-app editing were equal size.

Table 1: Anticipated sample sizes per country and experimental condition. Actual sample sizes are given in brackets.

	ES	LU	NL
Feedback instant	400 (433)	800 (882)	800 (748)
Feedback delayed	400 (433)	800 (884)	800 (737)
Interviewer	400 (433)	800 (881)	800 (685)
No interviewer	400 (433)	800 (884)	800 (800)
In-app editing	800 (866)	-	1200 (1085)
No in-app editing	-	1600 (1766)	400 (400)

The sample designs varied per country. In ES and NL simple random samples were drawn within selected regions of the country. Since around 20 households per interviewer was deemed optimal in terms of training and experience, samples were drawn in specific interviewer regions. These regions were spread across the countries from rural to urban areas. In LU samples were randomly drawn from the SILC panel, however. Since interviewers were not directly involved in recruiting respondents, no clustering in geographical regions was needed. Samples were not stratified based on household

characteristics such as age, income or household composition, as is done in some countries such as NL, in order to optimize statistical power in evaluations.

It is important to remark that the data reporting period varies between LU-NL and ES. LU and NL both invited households to report all expenditures for two weeks, while ES asked full reporting only for one week. This choice was made in order to conform to the regular HBS design in ES and LU. In NL, the regular full reporting period is one week, but a two week period is desired; the HBS phase 3 field test thus also implicitly allows for an evaluation of length of the reporting period. The one week/two weeks reporting period started on the day the household registered the app, i.e. logged in for the first time. In the app calendar screen the reporting days were highlighted and a completed day changed colour from blue to green.

3. DATA ENTRY OPTIONS

The HBS app allowed for two types of data entry: manual data entry assisted by product lists and scan data entry assisted by text recognition. Households wanting to enter purchases from digital receipts were instructed to use two devices, one to display the digital receipts and the other to scan the receipt using the app. In the future, uploading of digital receipt will be a third data entry option. In-app paradata was logged for both types of data entry in order to facilitate evaluation of potential barriers or issues. Both data entry options are explained.

3.1 Manual data entry

Manual data entry consists of five steps:

1. Choose the manual option in the calendar or purchase overview screen;
2. Indicate basic characteristics of the purchase: name of the shop or type of shop, yes/no a purchase abroad, and yes/no a purchase online;
3. Add products and corresponding prices one by one;
4. Add discounts and/or rebates
5. Confirm the purchase

A purchase can be edited at any time, even when a reporting day is flagged as complete. When a purchase is revised, the reporting day needs to be flagged again as complete. Steps 2, 3 and 4 require extra explanation. Screenshots of the app UI are provided in deliverables 2.8a and 2.9.

The HBS app makes use of so-called shop lists. These are country-specific and consist of two columns. The first columns lists names of shops (or to be more correct of service providers) and types of shops and the second column a fixed classification into 52 shop categories (see Annex D). When respondents start typing the name of the shop, then a search is performed in parallel within the list of shops. The algorithm employs the Jaro-Winkler distance, i.e. it calculates distances to terms in the list based on the number of character and position edits needed to transform the one to the other. Names/types are suggested to the respondent and he/she can pick a name/type. If a shop name or type is not found, then the respondent is asked to classify the type of shop/service. The respondent classification is added to the in-device database so that a next time this shop is chosen the classification is applied again. The shop names and classifications in manual data entry are collected in order to remain aligned

with scan data entry (see subsection 3.2). Deliverable 2.11 will provide an overview of the country product lists.

Product data entry, like shop name entry, makes use of lists, so-called product lists. These list can be viewed as one level more detailed than COICOP, and, most importantly, use common, daily language and brand names. The product lists consist of two columns, one with product names/descriptions and one with COICOP. Once a product is selected by a respondent, the corresponding COICOP becomes available. The product list for NL consists of roughly 30000 records, all linked to one unique COICOP. When typing a product name, a product search is started which again employs the Jaro-Winkler distance. However, for product search more options can be tuned, namely the maximum number of 'hits', a lower threshold to the distance, and the number of hits shown per COICOP. If a product is not found (or a respondent does not perceive a hit as a correct match), then he/she is invited to classify the product him/herself. The new product is then added to the in-device product list. Product lists are also provided in deliverable 2.11. The product search algorithm was tested and evaluated within Eurostat-funded project @HBS2. The following hyperlink points to a website where different tuning options can be evaluated

Product search algorithm

Entry of product discounts and rebates is known to be a complex respondent task as options vary per shop. Discounts can be added in three ways: 1) per individual product, 2) as a percentage of the sum of the entered product prices, and 3) as an amount on the sum of the entered product prices. Rebates can added as individual products with an amount that automatically gets a negative sign.

The product COICOP categories are added instantly to the household statistics overview in case a household is in the instant feedback condition. In the delayed feedback condition all entered products are added to the statistics, but display is postponed until the final day of reporting.

3.2 Receipt data entry

Receipt scan data entry consists of seven steps:

1. Choose the scan option in the calendar or purchase overview;
2. Wait for the automatic receipt detection to find the edges of the receipt;
3. If needed, crop the image to the section where products and prices are printed and confirm or retake the picture;
4. Add basic characteristics of the purchase: shop name/type, yes/no purchase abroad, yes/no purchase online and total amount;
5. If in-app editing is turned on: Wait for the app to display the products and prices recognized by the app;
6. If in-app editing is turned on: Edit products and prices and confirm;
7. If in-app editing is disabled: Confirm the receipt or retake the picture;

A receipt can be deleted at any time, also when a day is flagged as complete. When in-app editing is turned on, a respondent can also edit products and prices at any time. In both cases, the reporting day needs to be indicated as complete again. Since uploading of digital receipts is not supported in version 2.1.15, respondents were instructed to use a second device to display those receipts and make

a scan on the device on which the HBS app is installed. It must be remarked that in-app language processing of receipts assumed the format and structure and format of printed receipts, so that performance for digital receipts was suboptimal. Steps 2, 3 and 6 demand for extra explanation. Step 4 is the same as step 2 for manual data entry, except that also total amount was asked. The total amount was asked as a backup/validation option for text recognition and in order to include purchase statistics to the in-app feedback statistics. Step 5 took at most five to ten seconds depending on the length of the receipt and the type of OS/phone. Screenshots of the various steps are given in deliverables 2.8a and 2.9.

In-app receipt detection means that the app searches for a white/grey rectangle that is supposed to be the receipt. The respondent can follow this search on the screen, but can also overrule and take the picture directly. When found, a green box appears on the screen. The respondent can then accept or retake the picture.

Once a bounding box is selected, the respondent can move the four edges of the box in order to select a section of the receipt. This is recommended in tutorial movies and in instructions, especially for longer receipts. Once the bounding box is adjusted or deemed correct, the respondent can confirm or if needed retake the picture.

In-app editing of receipts employed the exact same editing screen as for manual data entry. This was done in order to remain consistent, but also turned about to have the best usability in preceding tests. Respondents could remove spurious products, add missing products, add discounts or rebates and check the total amount computed automatically with the amount on the receipt and the amount provided by them in step 4. When editing products, again the product lists were enabled so that classification was applied. However, when a receipt product text was not edited and accepted, then no classification took place. Product classification was done in-house only and results were not fed back to respondents.

In case the in-app editing was turned on, both the receipt and the extracted, potentially edited, products remained available in the app. Both were also sent to the HBS backend.

Regardless of in-app editing, the in-app OCR routines produces an OCR confidence score per anticipated product line. These scores were averaged over all lines and submitted to the HBS backend. A configurable lower threshold is embedded in the app that initiates warnings to the respondent whenever the confidence score is deemed too low. In the phase 3 field test, this threshold was set relatively low in order to allow for post-survey evaluation.

4. DATA QUALITY AND PLAUSIBILITY

The focus in this deliverable is on the substantive outcomes of the HBS phase 3 field test, i.e. the plausibility and richness of HBS main statistics. In section 5, the app-assisted HBS will be compared to the regular HBS 2020 to have a direct benchmark. However, since true values for household expenditures are unknown and the very reason for going ‘smart’ is data quality improvement, multiple quality metrics are evaluated.

Three general quality metrics are considered:

- Recruitment rates, both height and balance across relevant subgroups
- Completion rates, both height and balance across relevant subgroups
- Data richness

Of these three metrics, the recruitment rate is the most straightforward to define. It is defined as the proportion of eligible households that registered the app on at least one device and completed the introduction questionnaire (if applicable). For the interviewer-assisted recruitment, the causes for non-registration can be detailed and will be part of deliverable 2.4. For the self-administered recruitment, such an elaboration is not possible.

The completion rate is harder to define. It can be defined as the proportion of eligible households that completed all days of the diary within the specified time period of one week (ES) or two weeks (LU and NL). Households can complete a day by closing it explicitly in the app calendar screen. Marking a day as completed does, however, not necessarily mean that the household has reported all purchases for that day. So a threshold may be set to a minimal level of reporting of purchases in terms of numbers of receipts or amounts spent. Here, completion is defined as the household has shown in-app activity until the final reporting day.

Completion of the HBS diary overlaps with data richness. The lower level as defined for completion above may be seen as a minimally expected richness, despite it being possible, though unlikely, that a household has fewer or no purchases at all. Data richness is the hardest metric to define as any reported pattern may strictly be true. Data richness can be defined in a number of sub-metrics:

- Number of purchases reported per household member per day
- Variance of number of products per purchase within the household
- Variance of total purchase amounts within the household
- Variance of total purchase amounts across main COICOP categories within the household
- Distance of the proportion of total food and drink purchases relative to 10% of the household monthly income
- Variance in the amounts of individual purchases reported by household members

The first and last sub-metrics are not considered here as household members did not register individually in the field test. In the current HBS app, it is known how many devices are used by a household and which purchases are made on which device. However, it is unknown who is the owner of the device, multiple persons may enter data on the same device and the same respondent may switch devices during data collection. The number of devices used by households will be reported in section 5.1, but merely is an indication on how many different household members have been active.

In order to evaluate the sub-metrics, the following data can be used:

- Interviewer call/visit data
- Interviewer observations during recruitment and motivation
- General analytics by Apple/Google on app usage
- Type of smart devices
- Paradata on in-app navigation including time stamps, devices and clicks

- Helpdesk emails
- Linked auxiliary data from the sampling frame and admin data
- Original receipts and receipts processed in-device and in-house including performance metrics

The following household characteristics are available per country for the entire sample:

- ES
    - Age of all household members
    - Urbanization degree of the area of residence
    - Size of the household
    - Highest attained educational level of the household reference person in four classes
- LU
    - Household income
    - Age of the household reference person
    - Gender of the household reference person
    - Nationality of the household reference person
    - Highest attained educational level of the household reference person in three categories
    - Size of the household
- NL
    - Household income
    - Household total value of assets
    - Ownership of the house and registered house value
    - Urbanization degree of the area of residence
    - Type and size of household
    - Age of the household reference person
    - Gender of the household reference person
    - Ethnicity of the household reference person
    - Marital status of the household reference person
    - Highest attained educational level in five categories

Unfortunately, the overlap in variables between the three countries is limited to household size/type and age. Educational level is available for all countries but definitions vary per country. For this reason, data are not pooled across the countries in evaluations relative to auxiliary variables. It must be stressed again that the focus here is on HBS data; detailed elaborations of recruitment and completion rates will be part of deliverables 2.3 on respondent feedback and 2.4 on interviewer assistance.

5. FIELDWORK RESULTS

Discussion of HBS field test results are split into four parts: general results on recruitment and completion, results on data richness, results on data richness as a function of day in field, and comparisons to the Dutch HBS 2020.

5.1 Fieldwork results overall

In general, the context for the HBS phase 3 field test was unfavourable. In all countries, COVID-19 measures were in place and it is was uncertain at the start whether the face-to-face part would be

affected by restrictions at the door. The fieldwork was planned from September 1st to November 30th in NL within two batches of households with a month time lag. In ES and LU, fieldwork was planned from November 1st to December 31st. LU released the sample in a single step, but applied several reminders. ES fielded the sample in four batches with one week time lags. Table 1 also presents actually fielded sample sizes. In NL, the face-to-face sample faced restrictions from the second half of October onwards to the end of data collection. For this reason, 115 of the face-to-face households were never fielded and are left out of the analysis. In ES, face-to-face fieldwork had to be stopped towards early December in most of the regions. A total of 143 of the households that were in the field were moved to telephone recruitment. In LU, there were no restrictions as interviewers were only employed as contact persons. Interviewers prepared detailed doorstep reports that will be discussed in deliverable 2.4.

The incentive strategy varied per country. In ES, no incentive was given to respondents. In NL, an unconditional incentive of 5 Euro was sent/given at the invitation and a conditional incentive of 20 Euro was given upon completion. In LU, the incentive was dependent on the size of the household and ranged from 40 Euro for a single person household to 60 Euro for a household with children.

Table 2 provides the recruitment, activity and completion rates in the three countries. Activity is defined as submitting at least one purchase. Completion is defined as showing in-app activity from day one until the last day of data collection. Strictly speaking, households can thus complete the HBS diary without being active. However, this combination occurred only a few times in each country. For ES, rates are given with and without the households that were moved to telephone recruitment.

Table 2: Registration, activity and completion rates per country per mode condition including standard errors by normal approximation.

Country	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
ES	F2F + tel: 25.4 (2.1) F2F only: 29.4 (2.8) No-F2F: 11.3 (1.5)	F2F + tel: 21.5 (2.0) F2F only: 26.5 (2.7) No-F2F: 10.9 (1.5)	F2F + tel: 18.8 (1.9) F2F only: 22.4 (2.5) No-F2F: 8.3 (1.3)
LU	F2F contact: 30.1 (1.6) No F2F contact: 28.0 (1.6)	F2F contact: 22.9 (1.5) No F2F contact: 21.3 (1.4)	F2F contact: 18.3 (1.4) No F2F contact: 17.6 (1.3)
NL	F2F: 25.6 (1.7) No-F2F: 15.9 (1.3)	F2F: 23.8 (1.7) No-F2F: 11.6 (1.1)	F2F: 20.3 (1.0) No-F2F: 9.6 (1.0)

Table 2 shows that the three countries have relatively similar registration rates when interviewers are employed and when not. For ES and NL also activity rates and completion rates are quite similar. However, for LU activity rates and completion rates show a stronger decline with respect to registration. In deliverable 2.4 this country difference will be evaluated in detail using interviewer field reports.

In Annex C, registration rates, activity rates and completion rates are detailed for a selection of the available auxiliary variables in which response across modes is pooled. Although sample sizes for some of the strata are relatively small and, consequently standard errors fairly large, the general picture is consistent. There is little relation between participation and urbanization degree and household size, while there is a clear relation with educational level and age. Higher educational level goes hand in

hand with higher participation rates. Younger respondents and older respondents have lower registration and completion rates with reference persons of 75 years and older standing out very clearly. The picture of income is more mixed. In LU, there is no strong relation while in NL higher incomes participate more often. In NL, there is no clear relation between total household assets and participation.
For almost all variables, activity rates and completion rates drop in a comparable fashion along the strata.

Table 3 displays the results of logistic regressions for registration and activity employing the two variables, age and household size, that are shared by the three countries. Country is included as well as main effect. The results are in line with the univariate results; there are significantly lower rates for older age groups across all countries, but little variation in household size.

Table 3: Logistic regression models for predicting registration and activity for all data combined using household characteristics and with country as a main effect. Average Mean Effect (AME), regression coefficients (B), standard errors (SE) and p-values (p) are provided.

	Registration				Activity			
	AME	B	SE	*p*	AME	B	SE	*p*
Intercept		-1.78	0.23	***<.001***		-2.03	0.26	***<.001***
Age (ref = 18-24)								
25-34	0.12	0.70	0.24	***.003***	0.11	0.78	0.27	***.003***
35-44	0.12	0.70	0.23	***.003***	0.11	0.77	0.26	***.003***
45-54	0.07	0.43	0.23	*.065*	0.06	0.49	0.26	*.065*
55-64	0.04	0.26	0.24	*.277*	0.04	0.31	0.27	*.243*
65-74	-0.01	-0.05	0.25	*.834*	0.01	0.07	0.28	*.890*
≥75	-0.03	-0.25	0.28	*.380*	-0.03	-0.36	0.32	*.268*
Household size (ref = 1 person)								
2 persons	0.02	0.10	0.11	*.346*	0.01	0.06	0.12	*.630*
3 persons	0.04	0.22	0.12	*.070*	0.03	0.20	0.13	*.113*
4 or more persons	0.03	0.17	0.11	*.130*	0.02	0.12	0.12	*.330*
Country (ref = NL)								
Luxembourg	0.06	0.30	0.09	***<.001***	0.02	0.14	0.10	*.160*
Spain	-0.04	-0.26	0.11	***.025***	-0.03	-0.20	0.12	*.095*

Note. Significant p-values (<.05) in bold.

Table 4: Number of devices used by households in the three countries.

# devices	ES	LU	NL
1	90%	86%	88%
2	7%	12%	10%
3	1%	1%	2%
4+	2%	1%	0%

To finish, table 4 shows the number of devices being used by households in the three countries. In all countries, the majority of households reported purchases for HBS with a single device. Around 10% of households used two devices and a very small proportion of a few percentage points used more than two devices.

In summary, there is indication of selection bias on relevant auxiliary variables such as age, educational level and income. However, selection bias does not seem to increase during data collection.

5.2 Data quality and data diversity overall

In section 4, various data quality metrics are introduced related to numbers of purchased products and amounts paid by households. In this subsection, part of these metrics are explored for the three countries. A distinction is made between active households (i.e. submitting at least one purchase) and complete households (i.e. displaying in-app activity until the final reporting day). However, it turns out that the observed patterns for both types of households are similar. It must be stressed that the reporting period in ES was one week and in LU and NL two weeks. Some of the metrics are dependent on the number of reporting days.

Table 5: Household average total number of purchases, average number of manually entered purchases and average number of scanned receipts for ES, LU and NL. A distinction is made between active and complete households.

	ES	LU	NL
Total	Act = 11.8 Com = 12.6	Act = 22.7 Com = 25.5	Act = 19.7 Com = 21.8
Manual	Act = 8.0 Com = 8.5	Act = 17.5 Com = 19.8	Act = 15.0 Com = 16.8
Scan	Act = 3.8 Com = 4.1	Act = 5.2 Com = 5.7	Act = 4.8 Com = 5.1

Table 5 gives the average total numbers of purchases by households and the average numbers submitted manually and by scan for the three countries. The number of purchases for ES indeed is roughly half of those in LU and NL. The proportion of scans in ES is slightly higher. In ES, around a third of the purchases is reported through scans whereas in LU and NL this is around a quarter. As expected the numbers are slightly higher for complete households.

In tables 6 to 8, amounts submitted by active and complete households are described. Table 5 presents medians for total amounts per household, medians for average amounts per purchase per household and medians for minimum and maximum purchase amounts per household. The total amounts for NL are roughly twice as large as for ES, conforming to the double length of the reporting period. For LU, amounts are much higher. When looking at average purchase amounts it becomes clear that in LU these are much higher. As yet, it is unclear why these differences are so huge. Obviously, price levels are different in the three countries, with LU being more expensive than NL and NL in turn being more expensive than ES. A future option, may be to standardize by purchasing power parity (PPP) in order

to remove country differences in income. In section 5.3, it will be shown that LU households reported more larger purchases.

In table 7 and 8, descriptives are detailed to size of the household and income quintiles. As income quintiles are not available for ES, results are only given for LU and NL. For ES and NL, the size of the household clearly matters, as expected; households with two persons spent almost twice as much as single-person households. Again this is a sign of data richness as households seem to report expenditures for all household members. For LU, the pattern is less clear; larger households reported smaller amounts than households with two persons. In-depth explorations are needed to get more intuition to why this is, i.e. larger households had inactive household members and/or larger households really spent less. Looking at income quintiles in table 8, the reported amounts increase with income, as expected. For LU, there is an exception for the highest quintile. Again in-depth analysis is needed to understand this unexpected drop.

Table 6: Household median total amounts, median purchase amounts and median minimum and maximum purchase amount in Euros for the three countries. A distinction is made between active and complete households.

	ES	LU	NL
Median total	Act = 229 Com = 255	Act = 977 Com = 1120	Act = 487 Com = 580
Median purchase	Act = 21.2 Com = 21.3	Act = 48.1 Com = 51.1	Act = 28.2 Com = 28.5
Minimum purchase	Act = 1.5 Com = 1.5	Act = 3.2 Com = 3.0	Act = 1.7 Com = 2.1
Maximum purchase	Act = 72.5 Com = 76.9	Act = 202 Com = 228	Act = 97.7 Com = 105

Table 7: Household median total amounts in Euros as a function of household size for the three countries. A distinction is made between active and complete households. A distinction is made between active and complete households.

	ES	LU	NL
1	Act = 94 Com = 127	Act = 800 Com = 896	Act = 294 Com = 394
2	Act = 219 Com = 223	Act = 1166 Com = 1235	Act = 667 Com = 738
3+	Act = 256 Com = 290	Act = 971 Com = 1084	Act = 680 Com = 747

Table 8: Household median total amounts in Euros as a function of income quintile for LU and NL. A distinction is made between active and complete households.

	LU	NL
0-20%	Act = 789 Com = 1117	Act = 262 Com = 262

20-40%	Act = 887 Com = 933	Act = 419 Com = 447
40-60%	Act = 1080 Com = 1276	Act = 454 Com = 530
60-80%	Act = 1104 Com = 1299	Act = 535 Com = 650
80-100%	Act = 931 Com = 1120	Act = 657 Com = 774

Table 9: Household median total numbers of products, median purchase number of products and purchase number of product standard deviation for the three countries. A distinction is made between active and complete households.

	ES	LU	NL
Median #products	Act = 43 Com = 46	Act = 47 Com = 56	Act = 70 Com = 81
Md #prod purchase	Act = 3.8 Com = 3.8	Act = 2.1 Com = 2.3	Act = 4.7 Com = 4.4
Minimum #prod	Act = 1.0 Com = 1.0	Act = 1.0 Com = 1.0	Act = 1.0 Com = 1.0
Maximum #prod	Act = 15.5 Com = 16.0	Act = 16.2 Com = 12.0	Act = 19.0 Com = 21.0

Table 10: Household median number of products as a function of household size for the three countries. A distinction is made between active and complete households.

	ES	LU	NL
1	Act = 29 Com = 29	Act = 46 Com = 61	Act = 58 Com = 62
2	Act = 48 Com = 53	Act = 44 Com = 55	Act = 71 Com = 80
3+	Act = 46 Com = 49	Act = 47 Com = 52	Act = 102 Com = 111

Tables 9 and 10 provide similar information on household reporting but now looking at numbers of products rather than amounts. In table 9 medians are given for the household total number of products reported, the household average number of products per submitted purchase, and household minima and maxima purchase number of products. As anticipated, the total numbers of products for NL are almost twice as high as for ES. However, remarkably, the median number of products for LU is much smaller. This holds for both the total numbers and the average numbers. It turns out that LU households reported more purchases with smaller numbers of products but larger purchase amounts. A closer look at the type of purchases is needed to understand this difference to ES and NL. The minimum and maximum numbers again give a strong indication that households reported both purchases with small and large numbers of products. In the regular HBS, reporting of smaller purchases is conjectured to be underreported.

Table 10 displays the same descriptives as function of registered size of the household. The same pattern is observed as for purchase amounts. For ES and NL, there are increasing numbers of products with increasing household size. For LU, there is no such relation and numbers are stable across household size.

Table 11: Household average amount per category for ES, LU and NL. A distinction is made between active and complete households.

Purchase category	ES	LU	NL
Petrol station	Act = 30 Com = 33	Act = 157 Com = 189	Act = 40 Com = 46
Superstores	Act = 9 Com = 11	Act = 39 Com = 43	Act = 13 Com = 15
Supermarkets	Act = 111 Com = 100	Act = 302 Com = 344	Act = 172 Com = 190
Specialised food-drink	Act = 18 Com = 20	Act = 51 Com = 59	Act = 24 Com = 28
Clothing and shoes	Act = 25 Com = 29	Act = 67 Com = 78	Act = 37 Com = 42
Drug store/perfumery	Act = 3 Com = 3	Act = 15 Com = 16	Act = 15 Com = 17
Home interior	Act = 2 Com = 2	Act = 24 Com = 28	Act = 6 Com = 7
Garden and house	Act = 16 Com = 21	Act = 6 Com = 7	Act = 16 Com = 18
Lunching and dining	Act = 37 Com = 43	Act = 110 Com = 154	Act = 68 Com = 77
Media and books	Act = 12 Com = 13	Act = 52 Com = 61	Act = 14 Com = 17
All other types	Act = 56 Com = 60	Act = 686 Com = 794	Act = 248 Com = 283

As a last step, the types of purchases are considered in order to evaluate diversity in products reported. Table 11 shows the average household purchase amount in 11 categories. The 11 categories are based on the shop names and shop types reported by households. The shop names/types are used instead of the COICOP classification, because it was infeasible to fully annotate and evaluate the products on all approximately scanned receipts within the given time span. While for manual data entry COICOP categories are available, for scan data entry they need to be based on machine learning models and/or receipt text string matching. As COICOP classification performance still needs to be evaluated for all three countries, it would be too risky to base conclusions on the predicted categories. However, within the time span since data collection was completed, all receipts were manually checked on amounts, numbers of products and on shop names. Using the existing shop lists, all

reported potentially adjusted/supplemented names were recoded to 1) petrol station, 2) superstores/department stores/stores with a broad product range, 3) supermarkets, 4) specialised food and drink stores (such as bakeries, butchers, poultries, liquor store, etc), 5) clothing and shoe stores, 6) drug store/perfumery, 7) home interior stores, 8) garden centers and do-it-yourself stores, 9) cafes, restaurants and canteens, 10) media stores and book stores, and 11) a remainder category for all other stores. This classification was chosen based on the regular Dutch HBS 2020 in which store names/types were asked as well.

Table 11 shows that for LU and NL most of the expenditures fall in the remainder category, followed by supermarkets. In ES, supermarkets are the largest category. However, clearly on average responding households report purchases in all categories and doing so follow instructions given to them in the invitations. This is true for both active households and complete households. Complete households show larger average amounts, as expected, but the diversity across categories is very similar. This is a very good result as households that stop prematurely were conjectured to show also less variation in types of purchases.

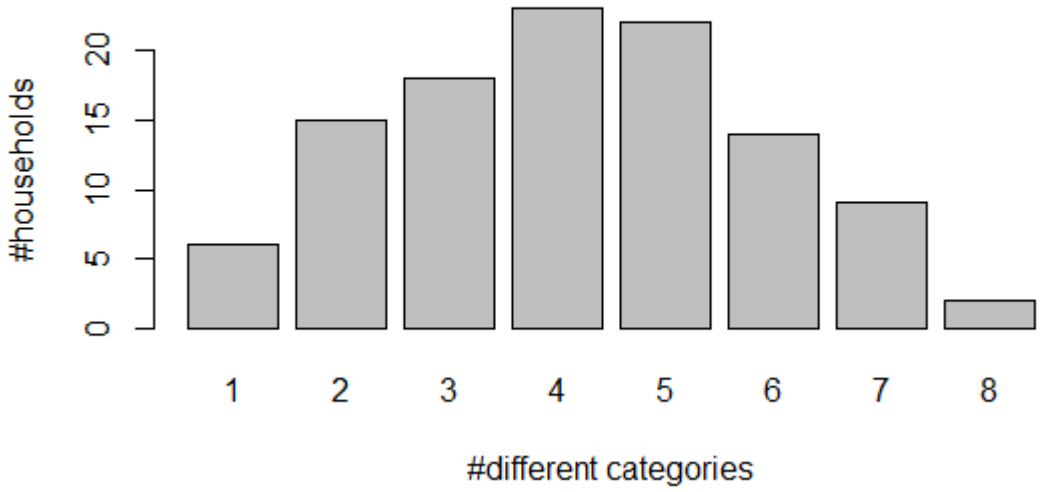

Figure 1: Barplot of the number of different shop categories for the active households in ES

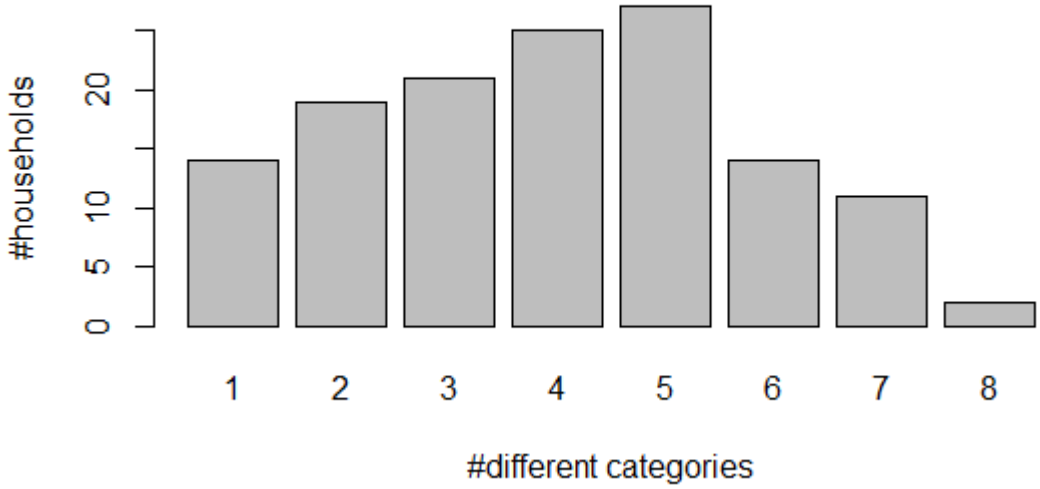

Figure 2: Barplot of the number of different shop categories for the complete households in ES.

Figures 1 to 6 depict country barplots of the number of different shop categories, as defined above, in which, respectively, active and complete households reported purchases. Tha maximum number os 11, which only occurred in NL. The modus for active households in both LU and NL around five to six different categories, and around two third of the active households reported at least five different categories. The modus for complete households shifts slightly to the right and now around three fourth of the households has five or more different categories. For ES, where the reporting period was one week instead of two weeks, the modus is four to five different categories. Differences between complete an active households, naturally, are also smaller, since there was less time to break off.

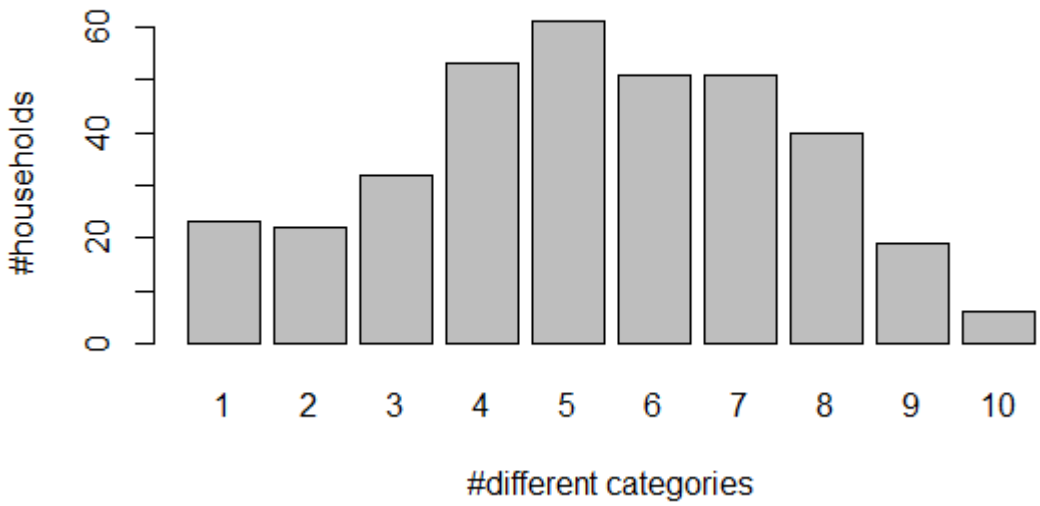

Figure 3: Barplot of the number of different shop categories for the active households in LU.

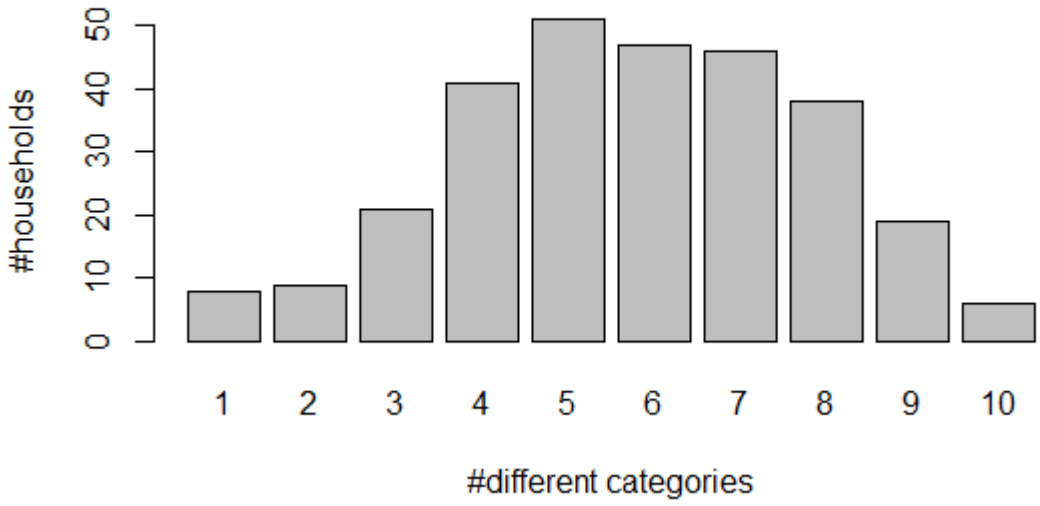

Figure 4: Barplot of the number of different shop categories for the complete households in LU.

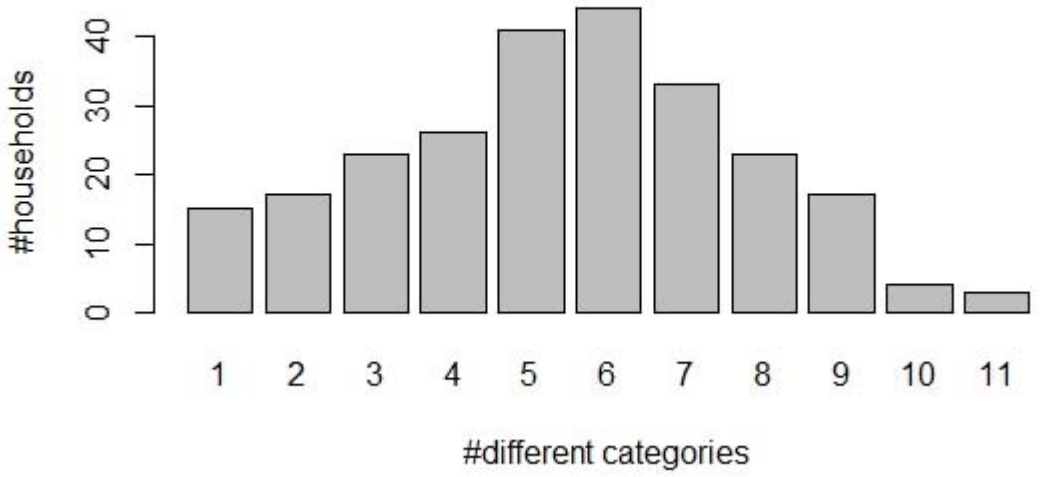

Figure 5: Barplot of the number of different shop categories for the active households in NL.

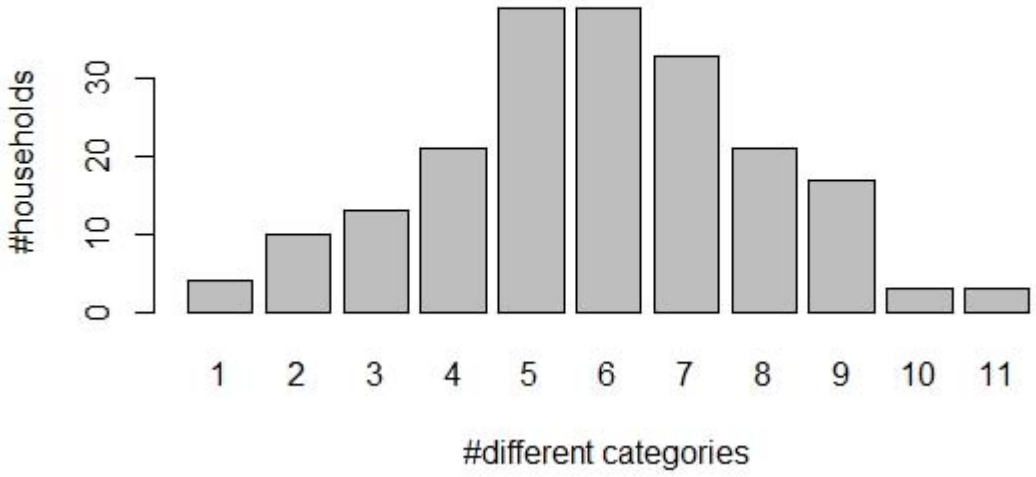

Figure 6: Barplot of the number of different shop categories for the complete households in NL.

All in all, the results point at diversity in the type of purchases reported by households. In subsection 5.4, these descriptives are compared to the regular HBS.

5.3 Data quality and data diversity as a function of time

At the start of the phase 3 fieldwork it was conjectured that data quality and data diversity are a function of time. More specifically, it was expected that in the second reporting week for LU and NL fewer purchases would be reported and variation in amounts and numbers of products would decrease. This decrease has been the main reason for a decision in NL ten years ago to restrict reporting of all purchases to a week only.
In this subsection, the reporting period is split in weeks and in half weeks and amounts and numbers of products are deduced per household for each sub-period. Results are shown for LU and NL only as ES had a week reporting period.

In order to get a sense of respondent behaviour, paradata on in-app navigation has been evaluated extensively. These evaluations are not in scope for ESSnet Smart Surveys and the phase 3 HBS field test, but will be reported in Eurostat-funded project @HBS2. Nonetheless, to get some intuition on how much time households spent in the app, figures 7 to 9 display the average amount of seconds active households spent on each reporting day in the app.
Paradata records show that on the first day much more time is spent since the intro questionnaire needs to be filled in and households tend to make short tours through the app main screens. When a scan is made for the first time, households also get to see a short movie that shows them how to scan receipts and crop the resulting images. From figures 7 to 9 it becomes clear that on the first day households spent between 10 and 15 minutes on average in the app. The intro questionnaire took around five minutes on average.

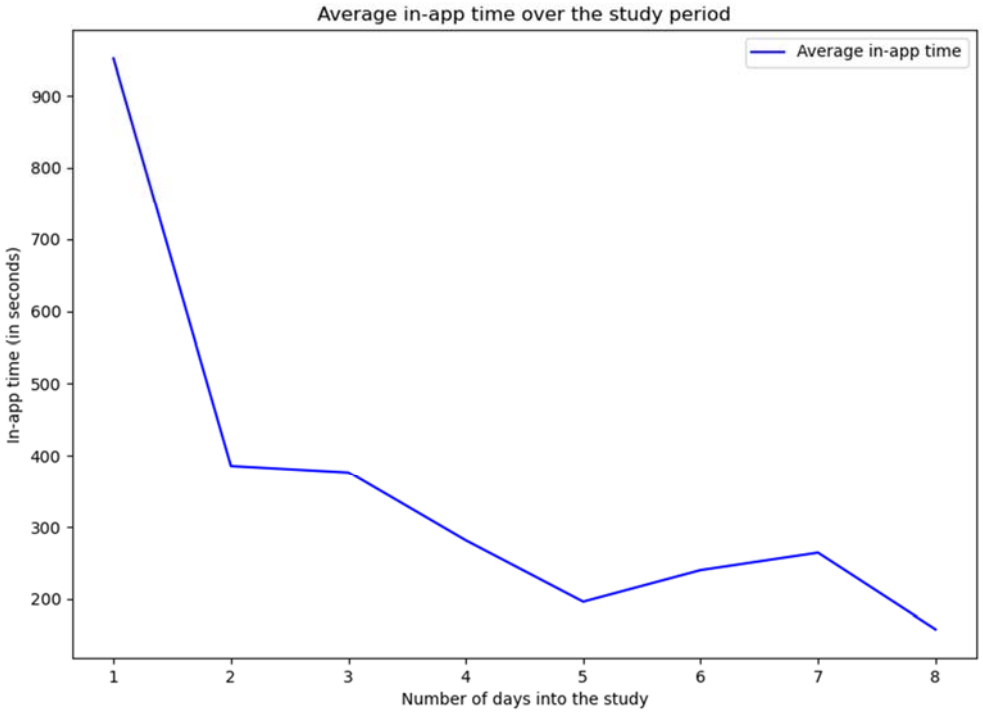

Figure 7: Average time spent in the app in seconds for ES as function of time

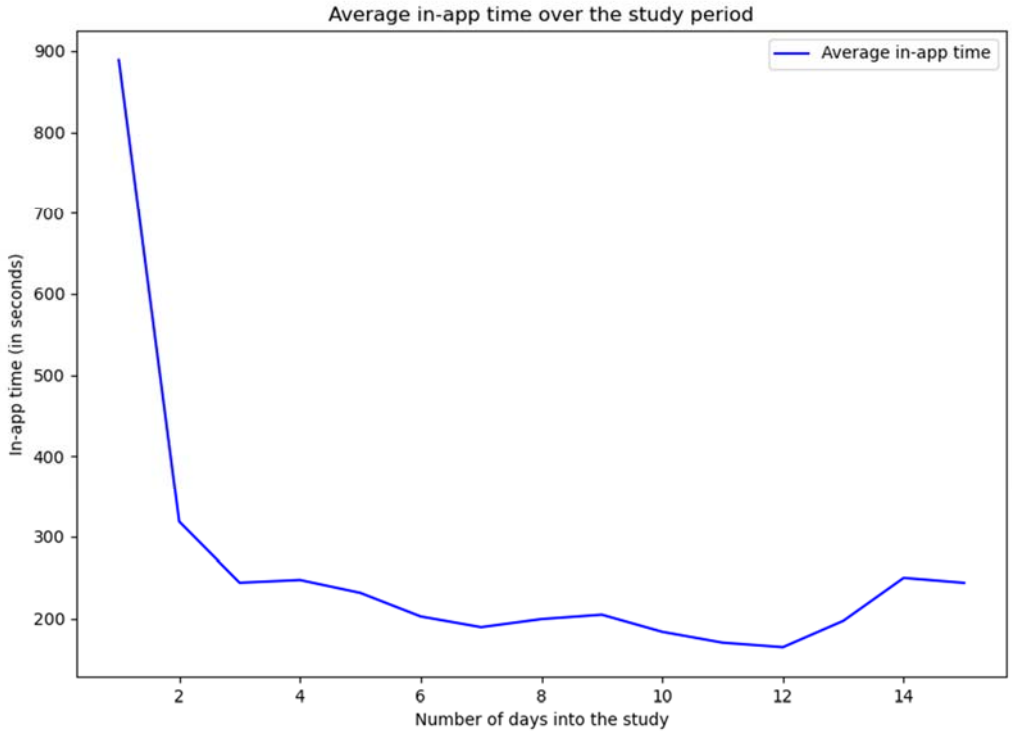

Figure 8: Average time spent in the app in seconds for LU as function of time

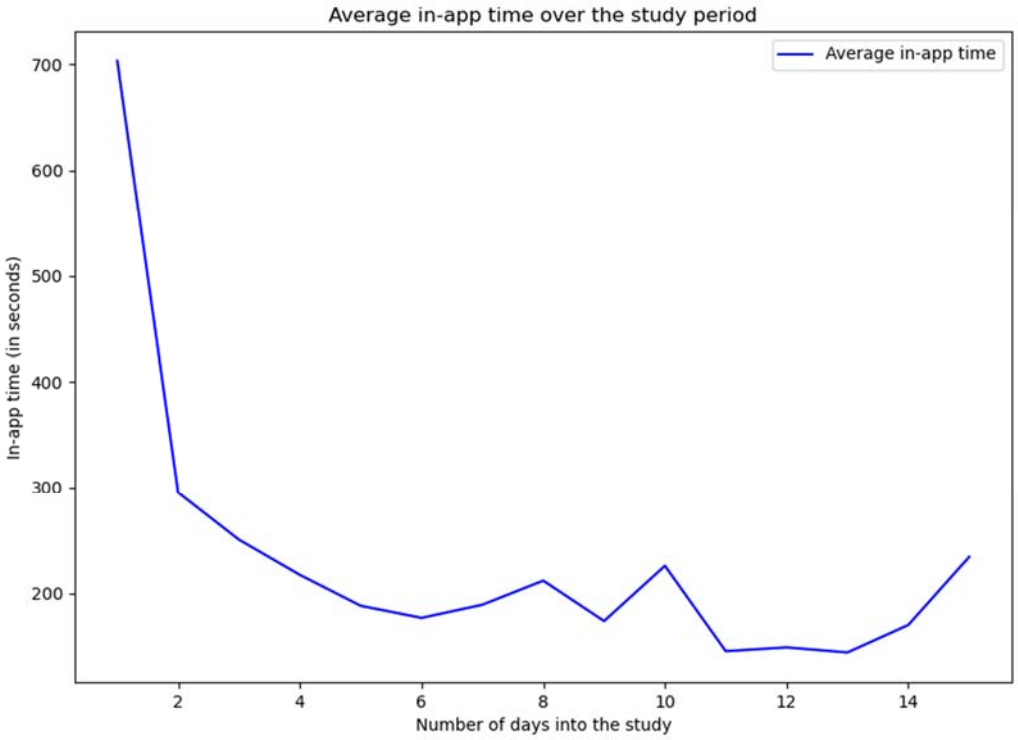

Figure 9: Average time spent in the app in seconds for NL as function of time

A sharp decrease in in-app time can be seen after the first day. On the second day around five to six minutes are spent in the app, after which the in-app time hoovers around three to four minutes on average per day. This drop seems to be the result of a learning curve and not so much the result of households spending less and less time in the app; already after three days the in-app time is quite stable.

Tables 12 and 13 present a range of statistics on purchases as a function of time for NL and LU, respectively. A distinction is in nine reporting periods: 1) the first full week (w1), 2) the second full week (w2), 3) the full second week extended with three extra days (w2+), 4) days 0 to 3, 5) days 4 to 7, 6) days 8 to 11, 7) days 12 to 14, 8) days 15 to 18, and 9) days 19 and further. Part of the households reported purchases after the first full two weeks. The HBS backend database only allows to see the exact date on which a purchase was made when respondents explicitly change the data in the entry screen. For this reason, it is very likely that most purchases shortly after the first two full weeks still belong to the reporting period. After day 18 a strong decline is observed in the number of purchases that are reported.

The first set of rows in tables 12 and 13 give statistics for the amounts and the second set of rows for the numbers of products. One observation stands out immediately and that is the much larger range of amounts reported by households in LU. While medians between LU and NL have a comparable magnitude, it must be concluded that LU households reported also large expenditures in the app. Detailed exploration of the types of purchases still need to be made in order to deduce whether this was the result of recurrent expenditures also being reported by LU households, LU households simply

spent more money, and/or NL households not reporting larger expenditures. Another observation that stands out is the cyclical nature of amounts and numbers of products in LU. This is the result of weekends and week days, where more purchases are reported in weekends. For NL, this cyclical pattern is not visible, because interviewers visited and recruited households at different days of the week.

Table 12: Descriptives of household purchase amounts and numbers of products as function of time period for NL active households.

Indicator	Year	Min	1st q	Median	Mean	3rd q	Max	SD
Total amount per household	2021 W1	0.0	113	240	304	411	1232	267
	2021 W2	0.0	62	178	273	380	3270	355
	2021 W2+	0.0	80	208	303	424	3700	376
	2021 0-3	-133	44	119	161	224	1136	172
	2021 4-7	0.0	24	88	144	188	1217	186
	2021 8-11	-20	9	98	159	228	1091	192
	2021 12-14	0.0	0.0	50	114	127	3154	267
	2021 15-18	0.0	0.0	0.0	31	2.2	956	98
	2021 19+	0.0	0.0	0.0	46	0.0	5195	350
Number of products per household	2021 W1	0	14	36	47	68	196	42
	2021 W2	0	6	26	38	52	214	43
	2021 W2+	0	7	28	43	57	341	50
	2021 0-3	0	5	18	27	41	153	28
	2021 4-7	0	3	11	20	28	141	25
	2021 8-11	0	1	10	23	32	167	31
	2021 12-14	0	0	5	15	22	214	25
	2021 15-18	0	0	0	5	1	127	16
	2021 19+	0	0	0	2	0	71	11

Table 13: Descriptives of household purchase amounts and numbers of products as function of time period for LU active households.

Indicator	Year	Min	1st q	Median	Mean	3rd q	Max	SD
Total amount per household	2021 W1	0.0	187	406	784	837	36164	2223
	2021 W2	0.0	88	313	678	705	36541	2082
	2021 W2+	0.0	129	404	763	847	36541	2102
	2021 0-3	-3	68	207	465	420	36156	2151
	2021 4-7	0.0	15	152	319	384	4736	512
	2021 8-11	0.0	0.0	108	409	346	34079	1897
	2021 12-14	0.0	0.0	120	269	299	4220	490
	2021 15-18	0.0	0.0	0.0	85	16	4849	324
	2021 19+	0.0	0.0	0.0	54	0.0	3580	281
Number of products per household	2021 W1	0	10	23	35	49	214	35
	2021 W2	0	4	13	25	34	244	32
	2021 W2+	0	5	17	27	36	300	35
	2021 0-3	0	5	12	20	26	144	22
	2021 4-7	0	1	8	15	19	129	21
	2021 8-11	0	0	5	15	19	178	24
	2021 12-14	0	0	4	10	12	123	10
	2021 15-18	0	0	0	3	1	144	10
	2021 19+	0	0	0	2	0	89	9

From the statistics, it can be concluded that week 1 and the extended week 2 show very similar amounts and numbers. This is a very positive finding as there is no clear sign of respondent fatigue or respondent satisficing behaviour.

Perhaps the most interesting statistic of all is the standard deviation of purchase amounts and purchase numbers of products. This statistic essentially reflects diversity in purchases. Again there is no sign that diversity in purchases deteriorates in the second week. In fact, for NL the standard deviation of amounts is even somewhat higher.

From tables 12 and 13, it also becomes very clear that LU households reported larger expenditures than NL households. This most likely has to do with the instruction of interviewers and the content of invitation letters. In NL, the focus was on the non-recurrent costs and not the fixed costs such as housing. In LU, the HBS app was presented as an all-in-one solution. It is expected that when recurrent expenditure questionnaire are added, this difference will disappear.

Table 14: Descriptives of within household range of purchase amounts and purchase numbers of products for active NL households.

Reporting period	Median	Mean
Amounts week 1	71	110
Amounts week 2	62	110
Amounts week 2 extended	68	116
#products week 1	13	17.1
#products week 2	10	14.7
#products week 2 extended	11	15.2

For NL, another sub-metric is evaluated in order to determine diversity, namely the range of within household purchase amounts and numbers of products. In order to compute this metric, per household the minimum purchase amount is subtracted from the maximum purchase amount and the minimum number of purchase products is subtracted from the maximum. Table 14 shows the results for the various reporting periods for NL. It is concluded that the range is relatively stable across reporting periods. There is a small decrease in the range for the number of products, but for the range of amounts a small increase.

The conclusion from these analyses is that there is no strong indication that active households decreased the quality of their reported purchases.

5.4 Comparison to the regular HBS

As a final evaluation, a comparison is made between the regular HBS and the app-assisted HBS in ES and NL. A main motivation to go smart is to improve data richness and diversity.

From January to December 2020, Statistics Netherlands performed its regular HBS. A sample of households was invited through mailed letters to participate in the HBS. Non-responding households with a registered phone number (around 60% of the household population) were called by a telephone interviewer. Hence, no face-to-face interviewers were employed in the regular Dutch HBS. As a result, the 2020 HBS only mildly suffered from COVID-19 regulations and restrictions. The recruitment rate for the 2020 HBS was around 15%, i.e. comparable to the non-interviewer sample in the ESSnet phase 3 field test. Since the ESSnet field test in NL was conducted from September 1st to November 30th,

there is an average time lag of around nine months between the two data collections. Here, it is assumed that there is no strong difference in expenditure patterns between the two data collections. This assumption cannot be checked, however, and demands for some caution in drawing strong conclusions. Another reason for some caution is the confounding of selection and measurement. The regular HBS and the app-assisted HBS may differ because responding provide different answers, but also because different types of households respond when they are invited to participate through an app.

In ES, the regular HBS is an annual survey being conducted by face-to-face interviewers. Pre-notification letters are sent. Interviewers conduct a starting interview, hand over paper diaries and return to the address to pick up diaries including receipts that the households have collected. In 2021, fieldwork had to be reorganized and part of the sample was re-allocated to telephone. Telephone interviewers did starting interviews but materials were sent and returned by households. The recruitment rate in 2021 was around 61%.

In LU, the regular HBS is an annual survey that is carried out by face-to-face interviews using freelance interviewers. The mode of data collection is PAPI. However, in 2020, due to the outbreak of the Covid crisis, face-to-face interviews had been stopped between March and September 2020. In order to keep the survey running, survey materials were sent directly to the households. Those were asked to fill in the diaries and the questionnaires themselves (interviewers could still be reached by telephone if needed) and send them back to the office using the prepaid envelope that had been provided.
See Household budget survey - Statistics Portal - Luxembourg (public.lu).

Two comparisons are made, one by considering purchase amounts and product numbers and one by looking at the type of purchases.

Table 15: Comparison between descriptives of the regular HBS 2020 and the app-assisted HBS phase 3 field test in 2021 for active households in NL.

Indicator	Year	Min	1st q	Median	Mean	3rd q	Max	SD
Total amount per household	2020	0.68	142	245	359	429	30003	544
	2021 all	2.39	262	488	654	874	5784	616
	2021 W1	0.00	113	240	304	411	1232	267
	2021 W2	0.00	62	178	273	380	3270	355
	2021 W2+	0.00	80	208	303	424	3700	376
Average amount per purchase per household	2020	0.14	18	26	36	39	15001	154
	2021 all	2.39	21	28	33	38	304	27
Number of products per household	2020	1.00	33	53	58	78	371	35
	2021 all	1.00	33	70	92	123	415	78
	2021 W1	0.00	14	36	47	68	196	42
	2021 W2	0.00	6	26	38	52	214	43
	2021 W2+	0.00	7	28	43	57	341	50
Average number of products per purchase per household	2020	1.00	4	5.3	6.2	7.6	79	4.4
	2021 all	1.00	2	4.7	5.6	7.3	43	5.1
Number of entries per household	2020	1.00	6	10	10.8	14	91	6.2
	2021 all	1.00	11	18	19.8	26	77	12.7
Number of categories per household	2020	1.00	3	5	4.7	6	11	1.9
	2021 all	1.00	4	6	5.5	7	11	2.3

Table 15 presents the statistics for household purchases in the regular HBS in NL, indicated as '2020', and the smart HBS, indicated as '2021'. For some of the statistics again a further distinction is made between different reporting weeks. The regular Dutch HBS asks households to report all expenditures for a week, but also uses separate questionnaires for recurrent expenditures such as housing, healthcare, insurance, etc.
A look at total amounts, average amounts, total number of products and average numbers of products reveals that means and medians for the regular HBS and for the smart HBS are similar in size. Since the regular HBS is much larger in size, it is less interesting to look at minima, maxima and standard deviations, as these are taken across all participating households. As a consequence, more exceptional purchases are observed. Drawing repeated samples of the same size as the smart HBS (results not shown) show that the smart HBS actually has a broader range and more extreme quantiles.
The mean and median number of different types of purchases is higher for the smart HBS. The mean increases from 4.7 to 5.4, and the median from 5 to 6. This indicates more diversity in the range of products reported. For this reason a closer look at the types of purchases may give the final answer.

Table 16 provides a slightly different comparison to the regular HBS 2021 in ES. The regular HBS data does not carry information if and how products are clustered within single purchases. For this reason it is not possible to compare purchase statistics and the comparison is restricted to amounts and numbers of products reported. Two distinctions are made, namely 1) the regular HBS through the whole of 2021 and the regular HBS only during the app field test period, and 2) only reference persons and all household members. The comparison where the regular HBS is restricted to the field test period but expanded to all household members should be closest to the app-assisted HBS. Amounts and are higher in the app-assisted HBS while numbers of products are comparable.

Table 16: Comparison between descriptives of the regular HBS 2021 and the app-assisted HBS phase 3 field test in 2021 for active households in ES. Within the regular HBS 2021 a distinction is made based on time period (the whole of 2021 or the same period as the HBS app field test) and household (only the reference person or all household members).

Indicator	Type	HH	Period	Min	1st q	Median	Mean	3rd q	Max	SD
Total amount per household	regular	no	all	1,00	121	213	293	360	22071	405
	regular	no	test	1,44	95	168	238	284	15587	461
	regular	yes	all	1,00	133	240	330	409	22111	434
	regular	yes	test	1,44	106	195	276	333	15587	477
	app	yes	test	0,90	101	221	327	383	2755	386
Number of products per household	regular	no	all	1,00	28	49	56	76	344	38
	regular	no	test	1,00	21	37	41	56	203	25
	regular	yes	all	1,00	30	51	60	81	372	40
	regular	yes	test	1,00	23	40	44	60	205	27
	app	yes	test	1,00	18	40	44	61	193	31

Next, reporting diversity in purchase types is considered for NL, ES and LU.

Table 17 presents the amounts reported in the 11 categories for NL. In both the regular HBS and the smart HBS supermarkets correspond to about a third of purchase amounts. The second largest category is the remainder types of purchases. The third and fourth largest are specialised food and drink stores and lunching and dining. The difference in the remainder category is remarkable. In the app, relatively more purchases are reported in a diverse set of shops/service providers. From this, it is concluded that the app HBS data are at least as diverse as regular HBS data.

Table 17: Comparison between types of purchases for the regular HBS 2020 and the app-assisted HBS phase 3 field test in 2021 for NL for active households.

Category	Type	Country	Median	Mean	SD	Total	%
Petrol station	Regular	NL	0.00	0.48	0.77	4970	4.4
	App	ES	0.00	0.59	0.83	79	5.0
	App	NL	0.00	0.89	1.44	218	4.5
Superstores	Regular	NL	0.00	0.52	0.87	5401	4.8
	App	ES	0.00	0.22	0.67	30	1.9
	App	NL	0.00	0.76	1.25	187	3.8
Supermarket	Regular	NL	3.00	3.77	2.62	39297	35.0
	App	ES	3.00	3.33	3.16	446	28.3
	App	NL	6.00	6.60	4.83	1624	33.3
Specialised food-drink stores	Regular	NL	1.00	1.38	1.79	14394	12.8
	App	ES	1.00	2.33	2.90	312	19.8
	App	NL	1.00	1.55	2.45	382	7.8
Clothing and shoes	Regular	NL	0.00	0.36	0.77	3801	3.4
	App	ES	0.00	0.49	0.96	65	4.1
	App	NL	0.00	0.59	1.06	145	3.0
Drug store/perfumery	Regular	NL	0.00	0.51	0.78	5348	4.8
	App	ES	0.00	0.08	0.35	11	0.7
	App	NL	0.00	0.78	1.14	192	3.9
Home interior	Regular	NL	0.00	0.09	0.35	982	0.9
	App	ES	0.00	0.04	0.23	5	0.3
	App	NL	0.00	0.12	0.43	30	0.6
Garden and house	Regular	NL	0.00	0.47	0.85	4928	4.4
	App	ES	0.00	0.12	0.39	16	1.0
	App	NL	0.00	0.56	1.07	137	2.8
Lunching and dining	Regular	NL	0.00	1.09	1.62	11303	10.0
	App	ES	1.00	2.14	3.11	287	18.2
	App	NL	1.00	2.59	3.45	636	13.0
Media and books	Regular	NL	0.00	0.30	0.62	3082	2.7
	App	ES	0.00	0.28	0.67	38	2.4
	App	NL	0.00	0.42	0.79	104	2.1
All other types	Regular	NL	1.00	1.81	1.90	18855	16.8
	App	ES	2.00	2.16	2.43	289	18.3
	App	NL	4.00	4.98	4.90	1226	25.1

Table 18 displays the diversity in products for the app-assisted HBS and the regular HBS in ES. Both concern 2021. Since store types are not collected in the regular HBS in ES, now the comparison is made based on first digit COICOP. In order to make the comparison all receipts have been manually classified. Again a distinction is made in whole year versus app field test period and in reference person only versus whole household. It can be observed that throughout the first digit COICOP there is close similarity in reported types of products. Relatively, the proportion of products falling under food and

non-alcoholic beverages, that contains most of the supermarket purchases, is slightly lower in the app, but overall diversity is very similar.

Table 18: Comparison between types of purchases for the regular HBS 2021 and the app-assisted HBS phase 3 field test in 2021 for ES for active households. Within the regular HBS 2021 a distinction is made based on time period (the whole of 2021 or the same period as the HBS app field test) and household (only the reference person or all household members).

COICOP	Type	HH	Period	Min	1st q	Median	Mean	3rd q	Max	SD	Total	%
Food and non-alcoholic beverages	regular	no	all	0,0	18,0	36,0	41,4	58,0	255,0	30,7	784.915	73,6
	regular	no	test	0,0	13,0	27,0	29,2	41,0	153,0	20,3	84.557	71,7
	regular	yes	all	0,0	19,0	36,0	42,1	59,0	260,0	31,1	797.215	70,7
	regular	yes	test	0,0	13,0	27,0	29,7	42,0	153,0	20,5	86.070	67,8
	app	yes	test	0,0	11,0	25,0	28,3	43,0	98,0	21,2	3.796	65,0
Alcoholic beverages and tobacco	regular	no	all	0,0	0,0	1,0	1,6	2,0	36,0	2,5	29.462	2,8
	regular	no	test	0,0	0,0	0,0	1,0	1,0	15,0	1,6	3.015	2,6
	regular	yes	all	0,0	0,0	1,0	1,8	2,0	44,0	2,9	33.731	3,0
	regular	yes	test	0,0	0,0	1,0	1,2	2,0	19,0	1,9	3.558	2,8
	app	yes	test	0,0	0,0	0,0	0,9	1,0	7,0	1,4	120	2,1
Clothing and footwear	regular	no	all	0,0	0,0	0,0	0,8	1,0	34,0	2,0	16.050	1,5
	regular	no	test	0,0	0,0	0,0	0,8	1,0	23,0	1,8	2.259	1,9
	regular	yes	all	0,0	0,0	0,0	0,9	1,0	34,0	2,1	17.694	1,6
	regular	yes	test	0,0	0,0	0,0	0,9	1,0	23,0	1,9	2.541	2,0
	app	yes	test	0,0	0,0	0,0	1,2	1,0	17,0	2,5	163	2,8
Housing, water, electricity, gas and other fuels	regular	no	all	0,0	0,0	0,0	0,1	0,0	27,0	0,6	2.545	0,2
	regular	no	test	0,0	0,0	0,0	0,1	0,0	7,0	0,5	285	0,2
	regular	yes	all	0,0	0,0	0,0	0,1	0,0	27,0	0,6	2.765	0,3
	regular	yes	test	0,0	0,0	0,0	0,1	0,0	7,0	0,5	329	0,3
	app	yes	test	0,0	0,0	0,0	0,3	0,0	24,0	2,1	37	0,6
Furnishings, household equipment and routine household maintenance	regular	no	all	0,0	1,0	3,0	4,6	6,0	72,0	4,8	87.494	8,2
	regular	no	test	0,0	1,0	2,0	3,4	5,0	56,0	3,6	9.710	8,2
	regular	yes	all	0,0	1,0	3,0	4,7	7,0	72,0	4,8	88.717	7,9
	regular	yes	test	0,0	1,0	2,0	3,4	5,0	56,0	3,6	9.909	7,8
	app	yes	test	0,0	0,0	2,0	3,1	4,0	19,0	3,7	416	7,1
Health	regular	no	all	0,0	0,0	0,0	0,7	1,0	33,0	1,3	13.410	1,3
	regular	no	test	0,0	0,0	0,0	0,5	1,0	11,0	1,0	1.515	1,3
	regular	yes	all	0,0	0,0	0,0	0,8	1,0	33,0	1,4	14.571	1,3
	regular	yes	test	0,0	0,0	0,0	0,6	1,0	11,0	1,1	1.692	1,3
	app	yes	test	0,0	0,0	0,0	0,9	1,0	7,0	1,3	115	2,0
Transport	regular	no	all	0,0	0,0	0,0	0,7	1,0	34,0	1,3	14.033	1,3
	regular	no	test	0,0	0,0	0,0	0,6	1,0	14,0	1,1	1.861	1,6
	regular	yes	all	0,0	0,0	1,0	1,0	1,0	34,0	1,6	19.229	1,7
	regular	yes	test	0,0	0,0	0,0	1,0	1,0	14,0	1,5	2.759	2,2
	app	yes	test	0,0	0,0	0,0	1,5	2,0	54,0	5,2	201	3,4
Communication	regular	no	all	0,0	0,0	0,0	0,0	0,0	7,0	0,3	874	0,1
	regular	no	test	0,0	0,0	0,0	0,0	0,0	3,0	0,2	88	0,1
	regular	yes	all	0,0	0,0	0,0	0,1	0,0	7,0	0,3	1.051	0,1
	regular	yes	test	0,0	0,0	0,0	0,0	0,0	3,0	0,2	113	0,1
	app	yes	test	0,0	0,0	0,0	0,1	0,0	1,0	0,2	8	0,1
Recreation and culture	regular	no	all	0,0	0,0	0,0	1,3	2,0	41,0	2,6	24.537	2,3
	regular	no	test	0,0	0,0	0,0	1,1	1,0	23,0	2,0	3.190	2,7
	regular	yes	all	0,0	0,0	0,0	1,6	2,0	41,0	2,9	29.672	2,6

	regular	yes	test	0,0	0,0	0,0	1,4	2,0	23,0	2,3	4.006	3,2
	app	yes	test	0,0	0,0	1,0	1,6	3,0	9,0	2,1	214	3,7
Education	regular	no	all	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-	0,0
	regular	no	test	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-	0,0
	regular	yes	all	0,0	0,0	0,0	0,0	0,0	1,0	0,0	1	0,0
	regular	yes	test	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-	0,0
	app	yes	test	0,0	0,0	0,0	0,0	0,0	1,0	0,1	1	0,0
Restaurants and hotels	regular	no	all	0,0	0,0	0,0	2,3	3,0	55,0	4,4	43.961	4,1
	regular	no	test	0,0	0,0	0,0	2,0	3,0	30,0	3,3	5.830	4,9
	regular	yes	all	0,0	0,0	1,0	3,8	5,0	106,0	6,2	71.841	6,4
	regular	yes	test	0,0	0,0	1,0	3,5	5,0	49,0	5,0	10.001	7,9
	app	yes	test	0,0	0,0	1,0	3,8	5,0	33,0	6,1	503	8,6
Miscellaneous goods and services	regular	no	all	0,0	0,0	2,0	2,6	4,0	49,0	3,0	49.553	4,6
	regular	no	test	0,0	0,0	1,0	1,9	3,0	25,0	2,4	5.639	4,8
	regular	yes	all	0,0	1,0	2,0	2,7	4,0	49,0	3,1	51.642	4,6
	regular	yes	test	0,0	0,0	1,0	2,1	3,0	25,0	2,5	5.970	4,7
	app	yes	test	0,0	0,0	1,0	2,0	3,0	21,0	2,6	267	4,6
TOTAL	regular	no	all	1,0	28,0	49,0	56,3	76,0	344,0	38,1	1.066.834	100
	regular	no	test	1,0	21,0	37,0	40,7	56,0	203,0	25,2	117.949	100
	regular	yes	all	1,0	30,0	51,0	59,5	81,0	372,0	40,1	1.128.129	100
	regular	yes	test	1,0	23,0	40,0	43,8	60,0	205,0	27,3	126.948	100
	app	yes	test	1,0	18,0	40,0	43,6	61,0	193,0	31,0	5.841	100

Table 19: Amounts in Euros per COICOP first digit for the regular HBS 2020 and app-assisted HBS in 2021 for LU. The regular HBS data is weighted using the standard LU weighting model. The HBS app data are both unweighted and weighted for comparison.

COICOP 1st digit	App		Regular
	unweighted	weighted	
Food and non-alcoholic beverages	6723	6446	6451
Alcoholic beverages and tobacco	709	713	994
Clothing and footwear	1625	1815	2968
Housing, water, electricity, gas and other fuels	424	546	22515
Furnishings, HH equipment and maintenance	3355	3115	4193
Health	1096	964	1597
Transport	2050	1997	10549
Communication	177	187	1720
Recreation and culture	1553	1751	3774
Education	88	23	260
Restaurants and hotels	2098	1797	4772
Miscellaneous goods and services	19055	18003	6299
TOTAL	38953	37356	66091

Finally, table 19 compares reported amounts of the regular HBS 2020 to the app-assisted HBS in 2021 for LU. The regular HBS 2020 amounts are weighted. The app-assisted HBS amounts are both weighted and unweighted. It must be stressed that in the 2021 app field tests, the focus was on the non-recurrent costs. This means that categories such as housing costs, education and transport were more or less out of scope, although respondent could report them. Recurrent expenditure questionnaires

are the next step for the app-assisted HBS. The amounts for the typical non-recurrent costs categories Food and non-alcoholic beverages and Alcoholic beverages and tobacco are quite similar for the regular and app-assisted HBS. For Clothing and footwear, Recreation and culture, Restaurants and hotels the amounts reported in the app are lower. Remarkably, the category Miscellaneous goods and services is almost three times as large in the app as in the regular HBS. Further, in-depth, evaluations are needed to see what causes this large difference.

In sum, it is concluded that the smart HBS shows similarity to the regular HBS and there is evidence that there is larger diversity both in amounts and types reported. The inclusion of recurrent expenditure questionnaires in the near future will allow for a more accurate comparison.

6. DISCUSSION AND NEXT STEPS

This deliverable set out to evaluate the data quality and data richness of the ESSnet smart household budget survey. It explored selection biases, it derived various statistics for purchase amounts, numbers of products and types, and it compared HBS statistics to the regular HBS.
The smart HBS achieves registration rates and completion rates that are relatively similar across the three participating countries depending on whether interviewers are employed or not. An in-depth analysis of selection biases was not in scope and is part of the upcoming deliverable 2.4. However, biases across countries display similar patterns and are mostly observed on age and educational level. For household income and household assets there was no clear pattern and further analyses are needed.
Statistics on amounts and numbers of products show that households report a wide range of purchases in the HBS app, from small single product purchases up to large expenditures. Especially in LU the range of purchase amounts was very large. This difference between countries is likely due to the presentation of the HBS app as an all-in-one solution in LU, whereas in other countries it was introduced for the non-recurrent expenditures.
Statistics on types of purchases reveal a strong diversity in types of products reported; purchases are found for all main types of shops/service providers and the majority of active households reported five or more different categories.
When comparing the smart HBS to the regular HBS, similar statistics are found, but the observed diversity in types of products seems to be somewhat larger. Some caution is needed because of the average time lag of nine months between the two data collections for LU and NL, COVID-19 measures being different, and the confounding of household selection and household reports.
Summarizing, the analysis of the smart HBS indicates a plausible set of HBS statistics and a strong diversity in the types of purchases being reported across all three countries that participated.

There are some limitations to this study. The first limitation is that in the smart HBS, recurrent expenditures have been ignored. Respondents were made aware of this, except in LU, and did not report housings costs, healthcare costs, insurance costs, and other recurrent costs. Even so, the smart HBS does not yet provide a complete tool to report all household expenditures; separate questionnaire for recurrent expenditures are under development and not in scope for ESSnet Smart Surveys. The second limitation is that, due to COVID-19, it was not possible to run the smart HBS in parallel to the regular HBS in NL. A direct comparison is, therefore, hampered and leads to some extra caution in drawing conclusions. Finally, there are still some open questions about subtle differences

found between countries that require further evaluations. These will be part of deliverables 2.3 on respondent feedback and 2.4 on interviewer assistance.

Although, not in scope of this deliverable, it is noted here that respondents in majority followed instructions in invitation letters, brochures and tutorials and that longer receipts posed no major issues in text recognition. These claims will be substantiated in deliverables by Eurostat-funded project @HBS2 that are due for March and April 2022.

Next steps for the HBS app are the expansion to recurrent expenditure questionnaires, the introduction of multi-device options for desk/laptops, i.e. bigger screens, and the inclusion of an upload data entry option for digital receipts. Especially, the digital receipts will further assist respondents in entering data. In the current application, respondents had to own and use two devices.

ANNEX A : COUNTRY ANSWERS TO FUNCTIONAL TEST QUESTIONNAIRE
See attachment

ANNEX B: RECRUITMENT MATERIAL FOR ES, LU AND NL
See attachment

ANNEX C: REGISTRATION, ACTIVITY AND COMPLETION FOR BACKGROUND VARIABLES

The tables in this appendix display registration, activity and completion rates in the three countries. Tables C.1 to C.4 are for ES, tables C.5 to C.9 are for LU, and tables C.10 to C.15 are for NL.

Table C.1: Registration, activity and completion rates for age in ES. Standard errors by normal approximation in brackets.

Age	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
<25 years	22 (5.5)	21 (5.3)	16 (4.8)
25-34 years	26 (3.5)	24 (3.4)	21 (3.2)
35-44 years	21 (3.2)	18 (3.0)	15 (2.8)
45-54 years	21 (3.5)	17 (3.3)	16 (3.2)
55-64 years	17 (3.4)	17 (3.4)	13 (3.0)
65-74 years	11 (3.0)	9 (2.8)	7 (2.4)
75+ years	3 (1.9)	2 (1.5)	1 (1.0)

Table C.2: Registration, activity and completion rates for household size in ES. Standard errors by normal approximation in brackets.

Household size	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
1	13 (2.6)	11 (2.4)	9 (2.2)
2	14 (2.3)	12 (2.1)	12 (2.1)
3	24 (3.1)	22 (3.0)	16 (2.7)
4+	21 (2.5)	19 (2.4)	16 (2.3)

Table C.3: Registration, activity and completion rates for educational level in ES. Standard errors by normal approximation in brackets.

Household size	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Primary/illiterate	14 (3.4)	10 (3.0)	10 (2.9)
Secondary	13 (2.1)	10 (2.0)	9 (1.9)
Lower tertiary	16 (2.3)	15 (2.2)	12 (2.0)
Higher tertiary	28 (2.8)	25 (2.8)	21 (2.6)

Table C.4: Registration, activity and completion rates for educational level in ES. Standard errors by normal approximation in brackets.

Household size	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
(Very) strong	18 (1.5)	17 (1.5)	15 (1.4)
Not - moderate	18 (2.6)	13 (2.3)	10 (2.0)

Table C.5: Registration, activity and completion rates for age in LU. Standard errors by normal approximation in brackets.

Age	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
<25 years	19 (4.4)	14 (3.9)	14 (3.8)
25-34 years	38 (4.4)	30 (3.9)	23 (2.5)
35-44 years	38 (2.8)	30 (2.7)	23 (2.0)
45-54 years	37 (2.3)	22 (2.2)	19 (2.1)
55-64 years	21 (2.3)	15 (2.0)	13 (1.9)
65-74 years	14 (2.8)	10 (2.4)	8 (2.1)
75+ years	10 (5.6)	-	3 (3.4)

Table C.6: Registration, activity and completion rates for household size in LU. Standard errors by normal approximation in brackets.

Household size	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
1	30 (2.9)	23 (2.6)	17 (2.3)
2	29 (2.0)	22 (1,8)	18 (1.7)
3	30 (2.3)	23 (2.1)	19 (2.0)
4+	28 (2.0)	22 (1.8)	17 (1.7)

Table C.7: Registration, activity and completion rates for educational level in LU. Standard errors by normal approximation in brackets.

Edu level	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Primary	25 (2.9)	17 (2.5)	14 (2.3)
Secondary	28 (2.0)	21 (1.8)	17 (1.6)
Tertiary	31 (1.5)	24 (1.4)	20 (1.3)

Table C.8: Registration, activity and completion rates for household income quintiles in LU. Standard errors by normal approximation in brackets.

Income	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
0-20%	29 (2.5)	21 (2.2)	16 (2.0)
20-40%	30 (2.5)	23 (2.3)	19 (2.2)
40-60%	33 (2.6)	23 (2.4)	20 (2.3)
60-80%	29 (2.5)	24 (2.3)	19 (2.1)
80-100%	25 (2.4)	19 (2.2)	16 (2.0)

Table C.9: Registration, activity and completion rates for nationality in LU. Standard errors by normal approximation in brackets.

Nationality	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Other	36 (2.9)	25 (2.6)	20 (2.4)
LU	26 (1.5)	20 (1.3)	16 (1.2)
DE-FR-PT	31 (2.1)	24 (2.0)	20 (1.9)

Table C.10: Registration, activity and completion rates for age in NL. Standard errors by normal approximation in brackets.

Age	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
<25 years	21 (5.1)	17 (4.6)	15 (4.5)
25-34 years	25 (3.0)	22 (2.9)	18 (2.7)
35-44 years	27 (2.9)	23 (2.8)	21 (2.6)
45-54 years	23 (2.6)	20 (2.5)	16 (2.3)
55-64 years	25 (2.6)	21 (2.5)	19 (2.3)
65-74 years	19 (2.6)	17 (2.5)	15 (2.4)
75+ years	8 (1.9)	6 (1.6)	6 (1.6)

Table C.11: Registration, activity and completion rates for urbanization in NL. Standard errors by normal approximation in brackets.

Urbanization	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Very strong	20 (2.0)	16 (1.9)	14 (1.8)
Strong	21 (1.5)	18 (1.4)	15 (1.3)
Moderate	23 (3.9)	18 (3.6)	16 (3.4)
Not/little	21 (3.2)	19 (3.1)	17 (2.9)

Table C.12: Registration, activity and completion rates for household income quintiles in NL. Standard errors by normal approximation in brackets.

Income	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Unknown	15 (5.2)	13 (5.0)	11 (4.6)
0-20%	11 (2.1)	10 (1.9)	9 (1.8)
20-40%	16 (2.2)	11 (1.9)	10 (1.7)
40-60%	19 (2.3)	17 (2.2)	13 (2.0)
60-80%	24 (2.5)	21 (2.4)	18 (2.2)
80-100%	34 (2.7)	30 (2.6)	27 (2.6)

Table C.13: Registration, activity and completion rates for household type in NL. Standard errors by normal approximation in brackets.

Household	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Single	18 (1.6)	15 (1.5)	13 (1.4)
Couple	20 (1.9)	17 (1.8)	15 (1.7)
Couple + children	27 (2.4)	23 (2.2)	19 (2.1)
Single + children	23 (4.2)	18 (3.8)	18 (3.8)

Table C.14: Registration, activity and completion rates for highest attained educational level in NL. Standard errors by normal approximation in brackets.

Edu level	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Primary	14 (3.7)	11 (3.4)	11 (3.4)
Lower secondary	15 (3.2)	12 (2.9)	11 (2.8)
Higher secondary	25 (2.2)	21 (2.1)	17 (2.0)
Lower tertiary	37 (3.6)	32 (3.4)	27 (3.3)
Higher tertiary	44 (5.0)	41 (4.9)	36 (4.8)
Unknown	13 (1.4)	10 (1.2)	9 (1.2)

Table C.15: Registration, activity and completion rates for household total assets quintiles in NL. Standard errors by normal approximation in brackets.

Assets	Registration rate in % (SE)	Activity rate in % (SE)	Completion rate in % (SE)
Unknown	15 (5.3)	13 (5.0)	11 (4.6)
0-20%	20 (2.5)	17 (2.4)	12 (2.1)
20-40%	15 (2.0)	12 (1.8)	12 (1.8)
40-60%	24 (2.6)	20 (2.5)	17 (2.3)
60-80%	25 (2.5)	22 (2.4)	20 (2.3)
80-100%	24 (2.5)	21 (2.5)	18 (2.3)

ANNEX D: SHOP LIST

petrol station
department store
store with broad product range
supermarket
bakery/confectionery
vegetable shop
butcher shop/poulterer
fish shop
dairy shop
liquor shop
other shops for food (specialised)
tobacconist
clothing shop/textile shop
shoe shop
home furnishings (furniture, carpet, curtains, etc)
lamps/lighting shop
electronics shop/kitchen appliances shop
computer shop
household items shop
construction market
flowery/garden center
bike/moped shop
toy shop
Photo/film shop
optics shop/jewelry shop
bookstore/book kiosk
sports and leasure shop
pharmacy
drug store/perfumery
souvenir shop/gift shop
other shops (non-food)
mobile phone shop
record/cd shop
car repair
secondhand from shop
secondhand from private person
leather goods/travel goods shop
antiques shop
do it yourself shop
shop for hardware/paint/wood/etc
pet shop
market/bought in street
mail order

at-the-door/drive-by shop
wholesale
farm/horticulture
company canteen
sports canteen
catering (hotel, cafe, restaurant, gelateria, etc)
all other purchases and services
sold
gift