Nonresponse in Web Surveys in Official Statistics

Nataša Bučar, Eva Belak, Marta Arnež, Katja Rutar

Statistical Office of the Republic of Slovenia

natasa.bucar2@gov.si

Abstract

Since 2016 at the Statistical Office of the Republic of Slovenia several household surveys have been transferred to the WEB data collection mode: Consumer Survey, Adult Education Survey, Survey on ICT Usage in Households and by Individuals (pilot and regular survey), pilot Household Energy, and Consumption Survey. These surveys, except pilots, are carried out in the mixed mode design.
The main source for the sampling frame for the household surveys is the demographic database built and updated by the Central Population Register.

For the purpose of profiling a typical non-respondent in our surveys, samples of these surveys were matched by some other administrative registers: the Real Estate Register, the Income Tax Records Register, and the Register of Households.
The goals of profiling a typical non-respondent are to be able to make some decisions regarding data collection and to take some actions to improve the response in population groups that are less likely to respond in the surveys.
To shed some light on the nonresponse bias, distributions of respondents and non-respondents will be analysed according to some register variable(s) by some socio-demographic characteristics, such as sex, age, education, activity status, household type, household income, type of housing, ownership of the dwelling, etc.

Keywords: mixed mode design, web surveys, nonresponse, nonresponse bias

1   Introduction

In recent years several surveys carried out by the Statistical Office of the Republic of Slovenia (SURS) have been transferred to WEB data collection mode (see Table 1 in the Appendix). The transition is not straightforward since the introduction of the WEB can have a considerable impact on the published estimates (non-response bias, mode effect). At the same time data collection processes, processes of the estimation as well as questionnaires need to be revised in the surveys of households and individuals. At SURS we have analysed non-response in all our WEB surveys and compared the results with surveys where face-to-face interview mode is used (f2f). The goal of non-response analysis was to learn more about the non-response patterns in WEB surveys and whether non-response is consistent across different surveys or is it survey specific.

We used the regular ICT survey and the WEB pilot ICT survey from 2017 for more detailed analysis. First non-response in both surveys by different socio-demographic characteristics and then also estimates for the key statistics for regular internet users in both surveys will be compared. In the conclusion and discussion we outlined some of the main issues that concern both data collection and estimation.

2 Non-response analysis

2.1 Sources of data and the variables

Besides the survey data, we also used the data from the following administrative and statistical registers in the analysis: Central Population Register and Register of Households (maintained by the Ministry of the Interior), Demographic Database (maintained by SURS), Real Estate Register (maintained by the Surveying and Mapping Authority of the Republic of Slovenia), telephone directory of fixed and mobile subscribers, and income tax records (maintained by the Financial Administration of the Republic of Slovenia) for the reference years 2014, 2015 and 2016.

The following variables that are available for respondents and non-respondents were included in the analysis: sex, age, education, activity status, type of household, number of household members, number of rooms, tenure status, type of housing, degree of urbanisation, assigned telephone number (telephone number for the selected person is available in the telephone directory), statistical region, type of settlement and net household income. It should be noted that net household income was matched to administrative households and that household size reported by the respondent does not always correspond to the administrative size. In the analysis we assumed that the distribution of "erroneous" households is the same for respondents and for non-respondents.

2.2 General non-response patterns in WEB vs. f2f

The completion rates of WEB surveys (AES, ICT, CS, HECS), which reached from 19% to 38%, were compared to the completion rate of the EU-SILC survey, 1st wave (57%), which is conducted every year on the field by CAPI method. EU-SILC is used as the benchmark for the non-response distribution for a typical f2f survey conducted by SURS. In the comparisons (Chart 1 - Chart 3) only persons aged 16 to 64 years were taken into account.

Chart 1 Response distribution for selected surveys

Note: The above-mentioned surveys have different target populations. Thus only persons aged 18-64 are considered in the response analysis.
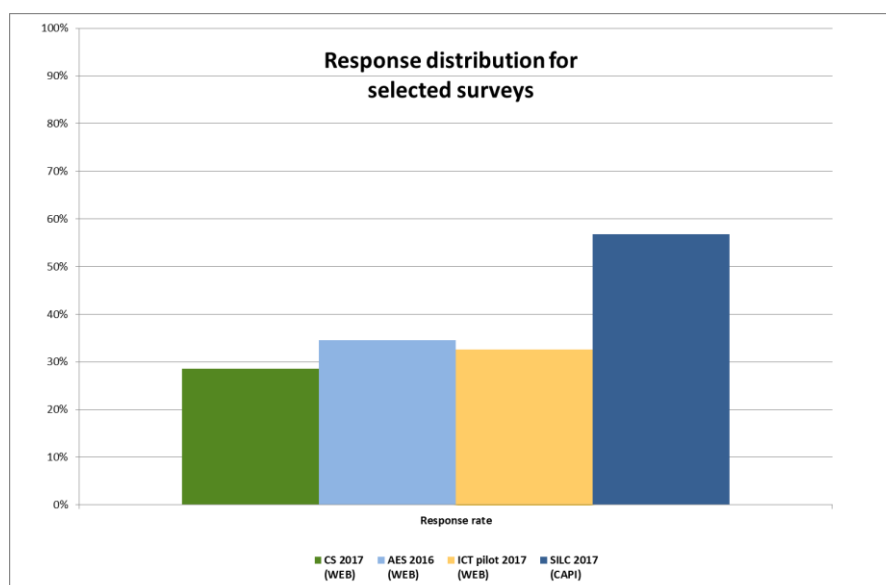
Source: SURS

In WEB and f2f surveys women respond better than men, persons who belong to a larger household are better respondents than single-member households. As regards the ownership, persons (or their household members) who own the dwelling in which they live respond better than non-owners. Similarly, persons living in larger dwellings respond more likely than persons in smaller ones. In WEB surveys younger persons cooperate significantly better (up to 61% in ICT) than other age groups. Cooperation drops with age of the respondent. If we consider education, the response rate is higher in higher educated groups in WEB surveys (up to 53% in AES and ICT), while in f2f surveys we cannot see the differences in response. But age in combination with education, we see that older respondents with higher education respond significantly better than younger ones with lower or medium education. In f2f data collection the highest response rate is in the oldest age group (18–29 years), the lowest is in the middle age group (30–49 years), while the youngest group (50–64 years) is somewhere in between. With all WEB surveys the average age of the respondents is lower than the average age of non-respondents, whereas in f2f surveys the opposite is true. Regarding the activity status, we see that in WEB students and pupils are our "best" respondents, while in f2f our "best" respondents are retired persons, but also students and pupils are responding above the average. It is also very obvious that response in WEB is increasing with income. In f2f surveys this is not so obvious, at least not in EU-SILC.

Chart 2   Response distribution for selected surveys by household income (quintile groups)
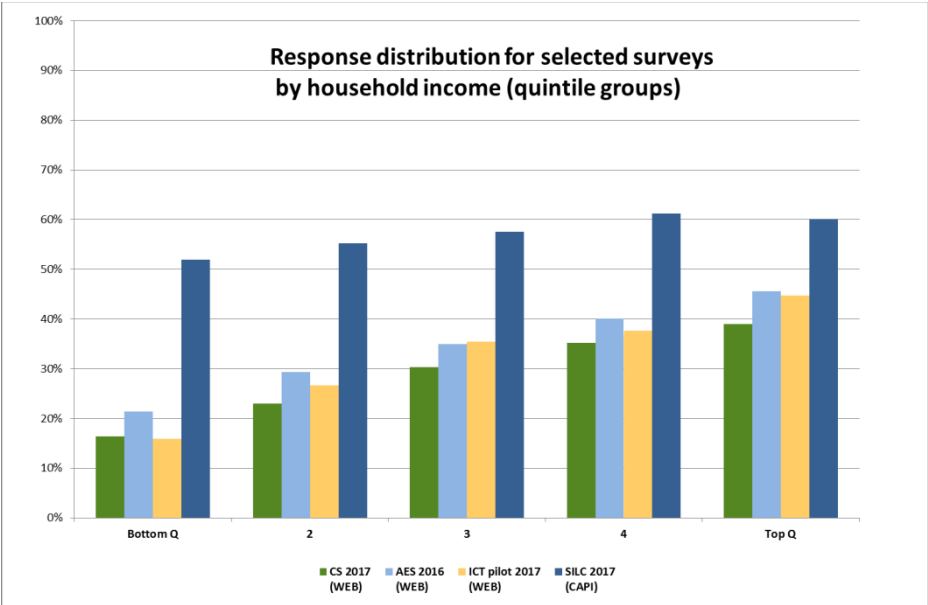
Source: SURS

Regarding the type of households, we see that respondents from households with children respond better than respondents from households without children. In f2f surveys couples without children have a higher response rate than couples with children. Type of building where the respondents resides is another distinguishing variable: in f2f the response rate in multi-dwelling housing is much lower (47%) than in individual houses (63%); in WEB there is no such difference.

Chart 3   Response distribution for selected surveys by type of housing

Response distribution for selected surveys by type of housing
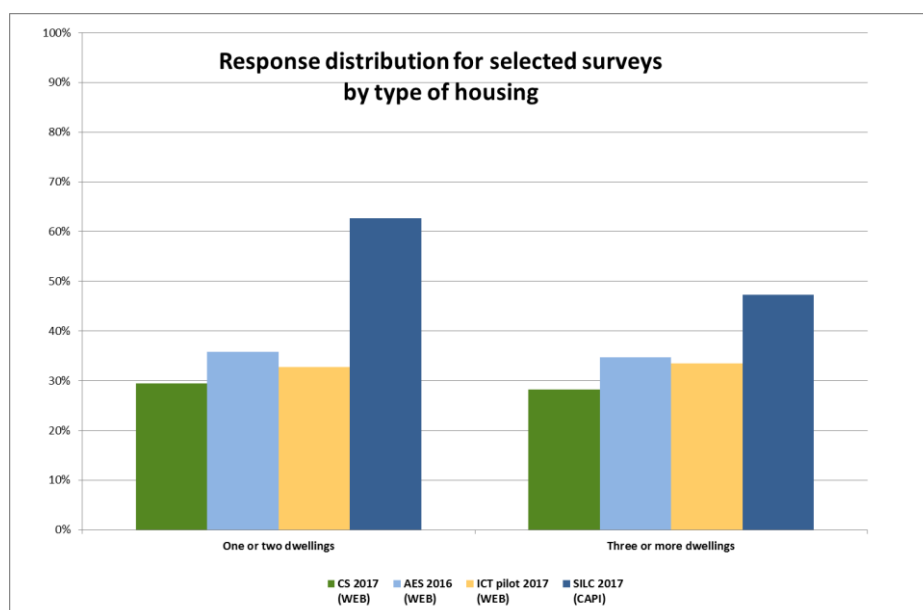
Source: SURS

In larger towns the response rate is lower (45%) than in smaller settlements (71%); in WEB there is no significant difference regarding the type of settlement.

2.3   Comparison of regular ICT 2017 and WEB pilot ICT 2017

The response distribution in the regular survey and in the WEB pilot, conducted one month after the regular survey is compared. We would like to find out whether certain socio-demographic groups are not responding to the WEB. The question is also whether the estimates for the same key statistics obtained in both surveys are significantly different and finally whether WEB mode can be used as the only data collection mode.

We have calculated weights for both surveys using the same variables for calibration (age-sex structure, activity status, statistical region, type of settlement and education structure). For the key statistics we compared weighted results of the WEB pilot survey and the regular survey.

2.3.1   Sample design and data collection

The source for the sampling frame for both ICT surveys was the Central Population Register and sample design for both was two-stage sampling. At the first stage 313 sampling units were selected and at the second stage 8 persons were selected, which is in total 2,504 persons aged between 16 and 74 years. The response rate of the regular ICT 2017 survey which was carried out in a concurrent mixed mode design (CATI and CAPI) was 68%. The data were collected from 9 March until 31 May 2017. CATI non-respondents were followed up with CAPI mode. Most responses (84%) were obtained by CAPI mode. The WEB pilot was conducted after the regular survey. The data were collected from 5 June until 26 June 2017. The questionnaires used in the pilot and for the regular survey were not

4

completely the same, because we additionally tested the ad-hoc module for 2018 in the WEB pilot survey. An introductory letter and two reminders were sent in the WEB pilot. In the regular survey separate introductory letters were sent for CATI and CAPI mode and also for persons aged less than 18. The WEB pilot ICT survey was on the other hand carried out only by WEB. The completion rate in the WEB pilot was 33%.

2.3.2   Non-response in regular ICT vs. WEB pilot ICT 2017

Both, the regular ICT survey and the WEB pilot ICT were conducted in 2017, so we had the opportunity to observe data collection mode impact on response distribution by various variables. We compared response distributions by socio-demographic variables in the WEB pilot and the regular survey in the analysis (see charts in the Appendix). We found out that the response distributions from the regular survey and the WEB pilot survey are similar for the following variables: sex (Chart 7), number of household members (Chart 8), number of rooms in the dwelling and tenure status (Chart 9). Response distribution by age (Chart 7) shows that older persons (65–74 years) cooperate significantly less in the WEB than other age groups, while in the regular survey the proportion of the responses in the oldest age group (65–74 years) is the highest. This can be explained by lack of necessary skills and usage of internet in the case of WEB and by availability of time and the need for social interaction in the case of CAPI interview mode. Similarly, response distribution by activity status (Chart 10) shows significant difference in the group of retired persons, where the response rate is lower than in other groups in the WEB survey, while in the regular survey (CATI, f2f) this group has the highest response rate. The response rate in the WEB survey grows with education, while in the regular survey the response rate by education is more evenly distributed (Chart 11). Response distribution by type of household (Chart 12) showed the biggest difference in the group "couple without children". In WEB data collection it has one of the lowest responses, while in the regular survey they have the highest response. Data collection mode has an impact also on response distribution by type of housing (number of dwellings in the building) and degree of urbanisation (Chart 13). The response rate in the regular ICT survey is the lowest in cities, a bit higher in towns and suburbs, and the highest in rural areas, while in the WEB pilot ICT survey the response rate is approximately the same in all three groups. In the WEB survey the response rate according to the number of dwellings in the building where selected persons live is approximately the same, while in the regular survey response proportion from persons living in buildings with one or two dwellings is much higher than from persons living in buildings with three or more dwellings. This may be attributed to the problems of interviewers in identifying and accessing households in multi-dwelling buildings.

In the WEB pilot the percentage of responses is increasing with household income, while in the regular survey the percentage of responses is also increasing by household income but the differences are not as big as in the WEB survey (Chart 15). Response distribution by type of housing in combination with household income (Chart 4) shows that in all quintile groups the response rate in the WEB pilot ICT is higher in buildings with three or more dwellings than in buildings with one or two dwellings, while in the regular ICT survey the opposite is true.

Chart 4   Response distribution by household income (quintile groups) and type of housing
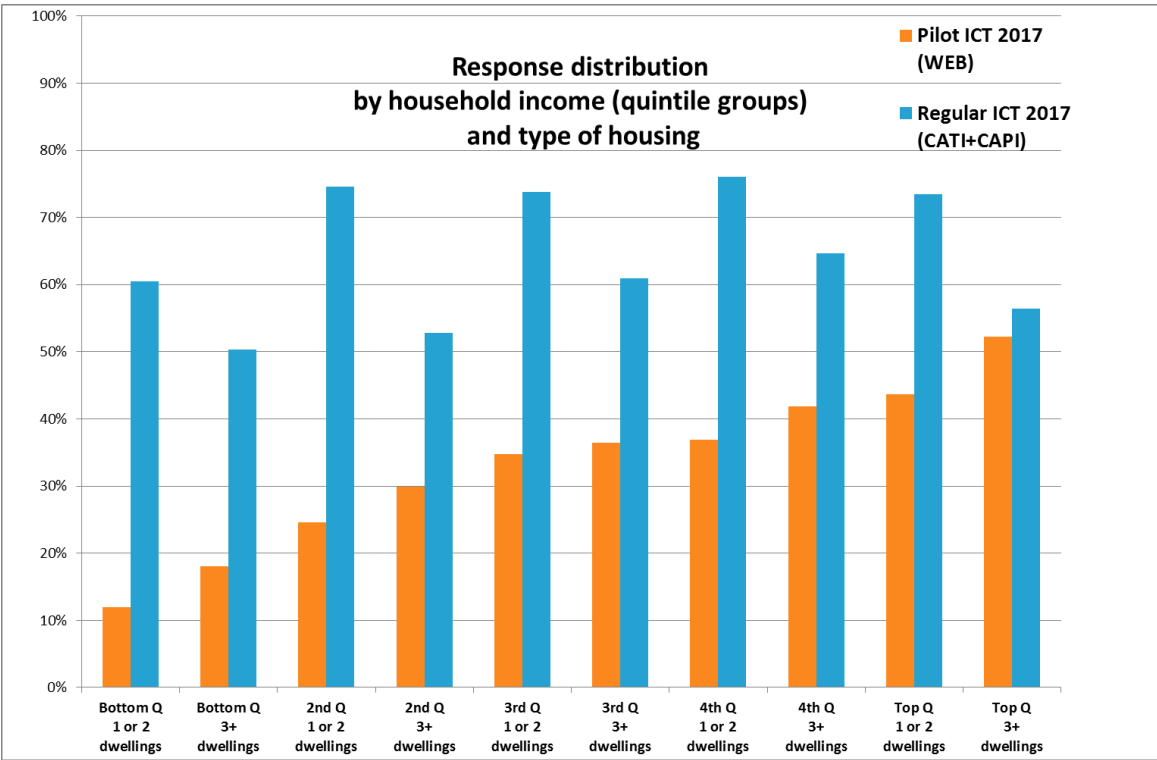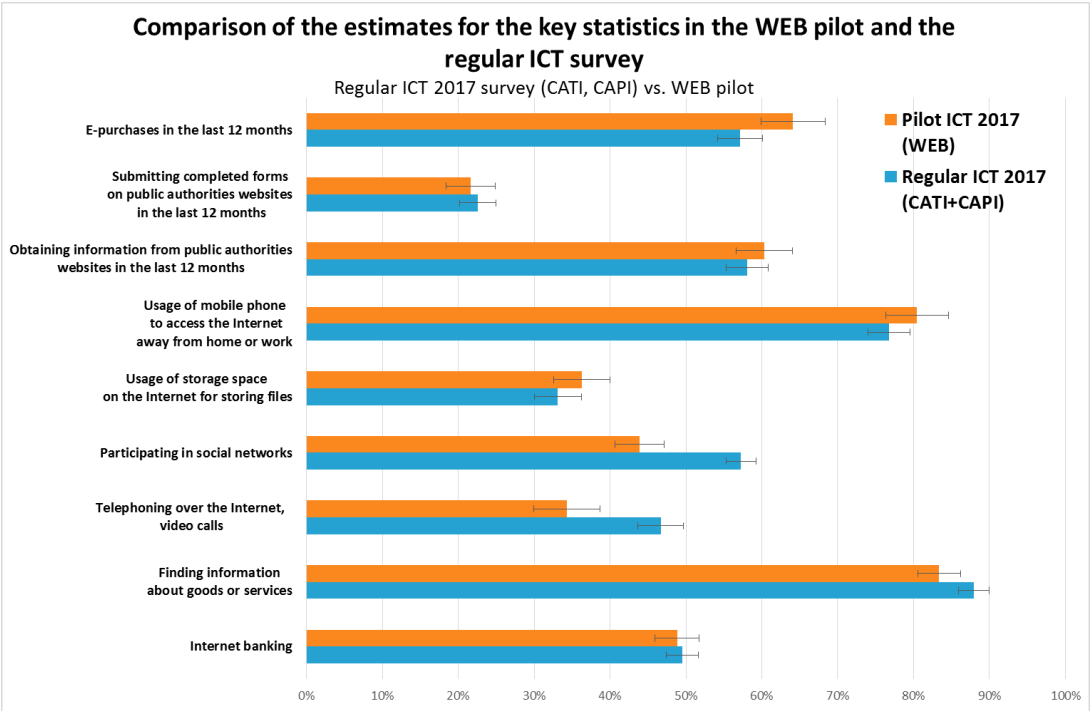

Source: SURS

From the charts (Chart 16 - Chart 20) where we analyse response by household income quintile groups, we can observe that the proportion of responses by education in all household income quintile groups is roughly evenly distributed, while in the WEB pilot ICT survey the proportion of responses generally increases with education. Response distribution by degree of urbanisation in combination with type of housing (Chart 21) shows that in the regular ICT survey response is lower in buildings with two or more dwellings, while in the WEB pilot ICT responses are more evenly distributed by type of housing.

2.3.3   Comparison of the estimates for the key statistics

Results of the estimates of the two independent samples (regular survey and WEB pilot survey) are compared. Only regular internet users, i.e. persons who used the internet in the last 3 months before the interviewing, are considered in the analysis. The results of the key statistics are presented in Chart 5. We can see that the differences are less than 5% for 6 out of 9 statistics.

The largest differences are in the statistics "E-purchase in the last 12 months", "Participating in social networks" and "Telephoning over the Internet, video calls". For the last two statistics we presume that there was a measurement error and that the questions were not correctly understood in the WEB survey, while in f2f there was an interviewer who could correctly interpret the questions.

Chart 5   Comparison of the estimates for the key statistics in the WEB pilot and the regular ICT

Source: SURS

The analysis of the WEB pilot ICT survey data from 2017 shows that using only WEB mode would give us somewhat biased estimates for some of the key statistics of the survey assuming that the regular ICT survey 2017 (conducted by CATI and CAPI) is the true value.

Chart 6   Percentage of e-purchases for internet users in the last 12 months
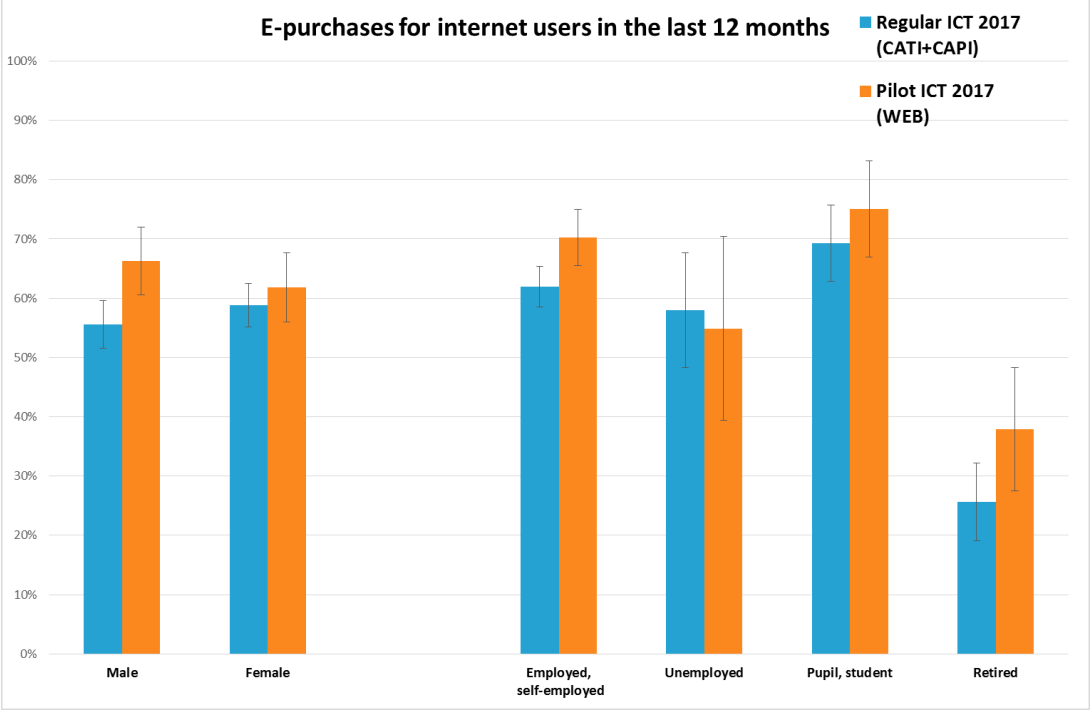
Source: SURS

From the above chart we can observe that in all sex and activity status categories except unemployed persons in the WEB pilot we obtain higher percentage of e-purchases in the last 12 months. The largest difference is in the group "retired persons". This finding is somewhat surprising since the structure of "how often they use the internet" is very similar in both groups. If we compare activity status for both groups by income, we can observe that the average income is higher for the respondents from the WEB pilot compared to the regular survey.

3   Conclusions and discussion

It is important that we introduce WEB mode in the surveys since the society is undergoing the process of digitalisation and people demand WEB questionnaires and even apps for passive data collection. In this way they can complete a questionnaire when they have time. The other reason is also that some respondents are more difficult to reach in other modes, for example persons in towns and persons living in multi-unit dwellings respond to WEB better than f2f.

However, response analysis shows that in all WEB surveys higher educated persons and persons with higher household income respond more. If we don't want to lose persons with lower income and less educated persons, we have to combine WEB with other data collection modes. Significantly higher e-purchasing rate in the WEB only pilot ICT survey can be a consequence of that bias.

General questions and points for the discussion

General non-response patterns in WEB surveys

- In multi-dwelling houses response in WEB is higher than in individual houses (all SURS's WEB surveys confirm this finding). Higher educated persons and persons that have higher income respond significantly better than persons with lower education and lower household income. We calibrate results by education structure but not by income structure.
- Using (administrative) household income as a calibration variable (experiences of other NSIs).

Specific questions to the ICT survey

- Measuring the percentage of internet users in the general population is not possible in the WEB survey. The question is whether "internet use" could be added to some other f2f survey (EU-SILC or LFS) – and in the web ICT survey only questions for internet users are asked.
- Experiences of other NSIs regarding the above question.

General points for discussion

- Transition to WEB is not straightforward: How do other NSIs introduce WEB data collection in their statistical production?
- How to decide whether WEB survey mode is enough and when it is necessary to combine it with other survey modes? Do other NSIs have any "rule of thumb" regarding this question?

References

Belak E., Arnež M., Nemec M., Zgonec M. and Steenvoorden T.: Transition to WEB data collection in household surveys at SURS, Available at: https://www.q2018.pl/wp-content/uploads/2018/06/Q2018_abstract_book.pdf.

Callegaro M., Lozar Manfreda K. and Vehovar V. (2015), Web Survey Methodology. London: Sage.
ESSnet DCSS (2014a) Query on Data Collection for Social Surveys. Available at: https://ec.europa.eu/eurostat/cros/content/d22-query-report_en (Accessed: 10 March 2018).

ESSnet DCSS (2014b) Report on the Definition, Identification and Analysis of Mode Effects. Available at: https://ec.europa.eu/eurostat/cros/content/d35-mode-effects_en (Accessed: 10 March 2018).

Thomsen, I.O., Kleven, Ø., Wang, J.E. and Zhang, L.C. (2006) Coping with Decreasing Nonresponse Rates in Statistics Norway. Recommended Practice for Reducing the Effect of Nonresponse, Reports 2006/29, Statistics Norway. Available at:
https://www.ssb.no/a/publikasjoner/pdf/rapp_200629/rapp_200629.pdf (Accessed: 9 March 2018).

Vehovar V. and Zaletel M. (1998), Non-response Patterns from the Matched Data, Advances in methodology and statistics (Metodološki zvezki), 14, pp. 115-127 Available at: http://dk.fdv.uni-lj.si/metodoloskizvezki/Pdfs/Mz14VehovarZaletel.pdf (Accessed: 9 March 2018).

Appendix

Table 1   List of mixed mode surveys at SURS using WEB as a survey mode

| Name	Main statistics	Popul-ation	Sample size and sample design	Survey length; # reminders WEB, (other mode)	Survey modes WEB completion rate ONLY(%)	Periodicity; WEB data collection period
Consumer Survey – CS	Consumer confidence indicator (CCI)	16-84	**3,000** Stratified sample	**6 min** (5 min); 1 reminder	Consecutive WEB -> CATI **29.0%**	Monthly
Pilot Adult Education Survey – AES	Participation in formal, non-formal, informal (%)	18-69	**2,075** Two stage sample design	1 reminder	WEB **28.5%**	Pilot
Adult Education Survey – AES		18-69	**8,504** Two stage sample design	**6 min** (7 min) 2 reminders	Consecutive WEB -> CATI, CAPI **34.6%**	Every 5 years
Pilot ICT in Households and by Individuals – ICT	Internet usage for different purposes (%)	16-74	**2,504** Two stage sample design	**13 min** 2 reminders	WEB **32.5%**	Pilot; June 2017
ICT in Households and by Individuals – ICT		16-74	**2,504** Two stage sample design	**13 min** 2 reminders	Consecutive WEB -> CAPI **37.5%**	Yearly, March 2018
Pilot Household Energy Consumption Survey – HECS	The total amount of consumed energy and fuels (in energy units) by energy sources and by end-use (space heating, water heating, cooking, other)	House-holds via select-ed person (18+)	**2,407** Stratified SRS: 2007 + Two-stage: 400 =2,407	**17 min** 2 reminders	WEB **19.3%**	Pilot; Sept.-Oct. 2017 (regular survey every 4 years)

Source: SURS

Chart 7   Response distribution by sex and age

Response distribution by sex and age
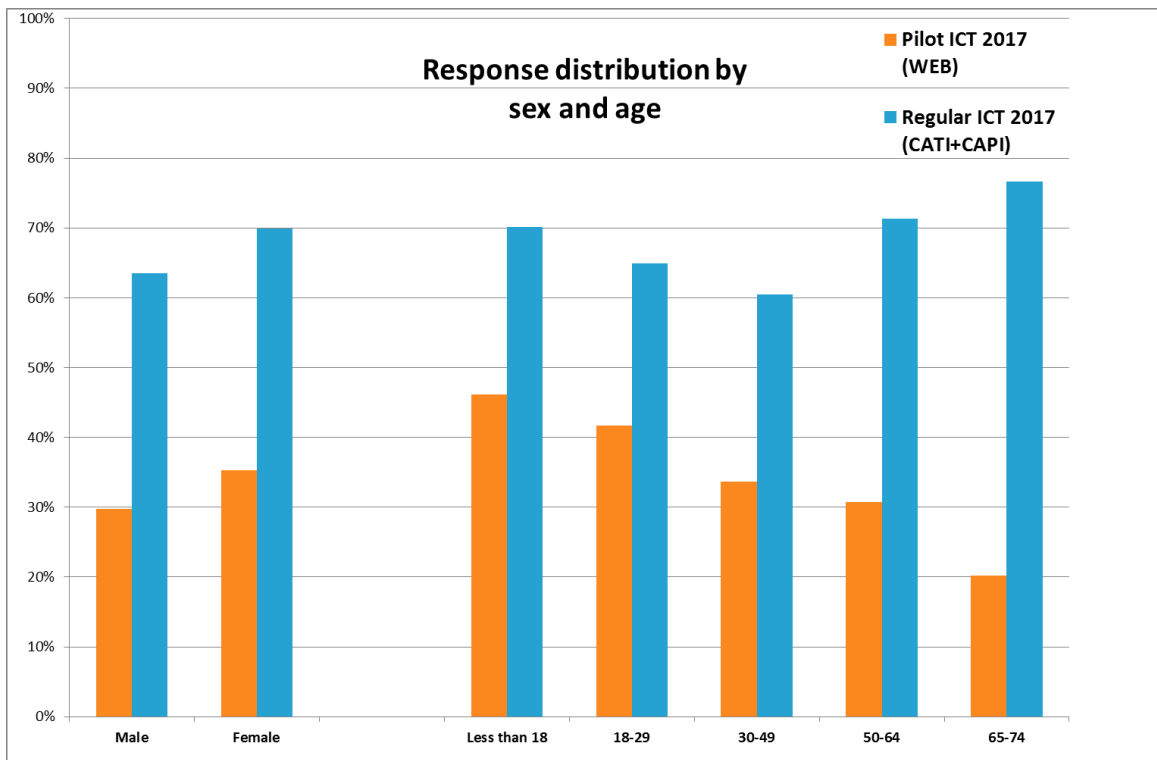
Source: SURS
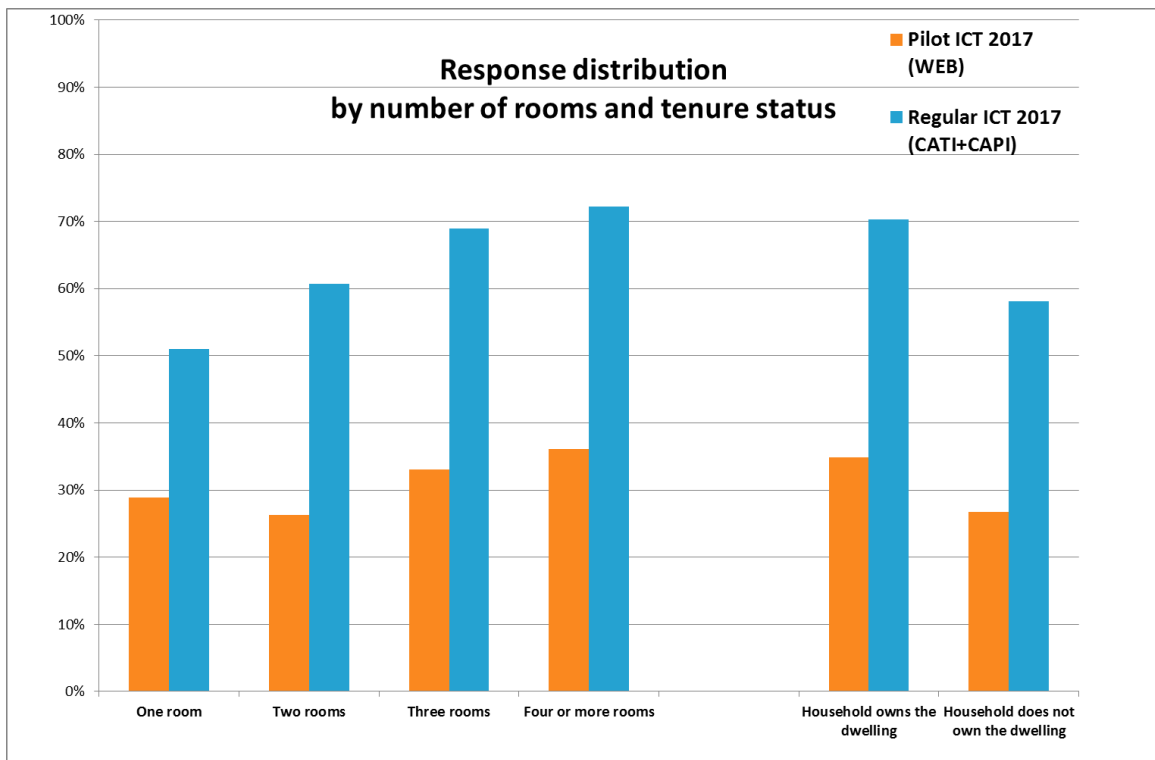

Chart 8   Response distribution by number of household members

Response distribution by number of household members
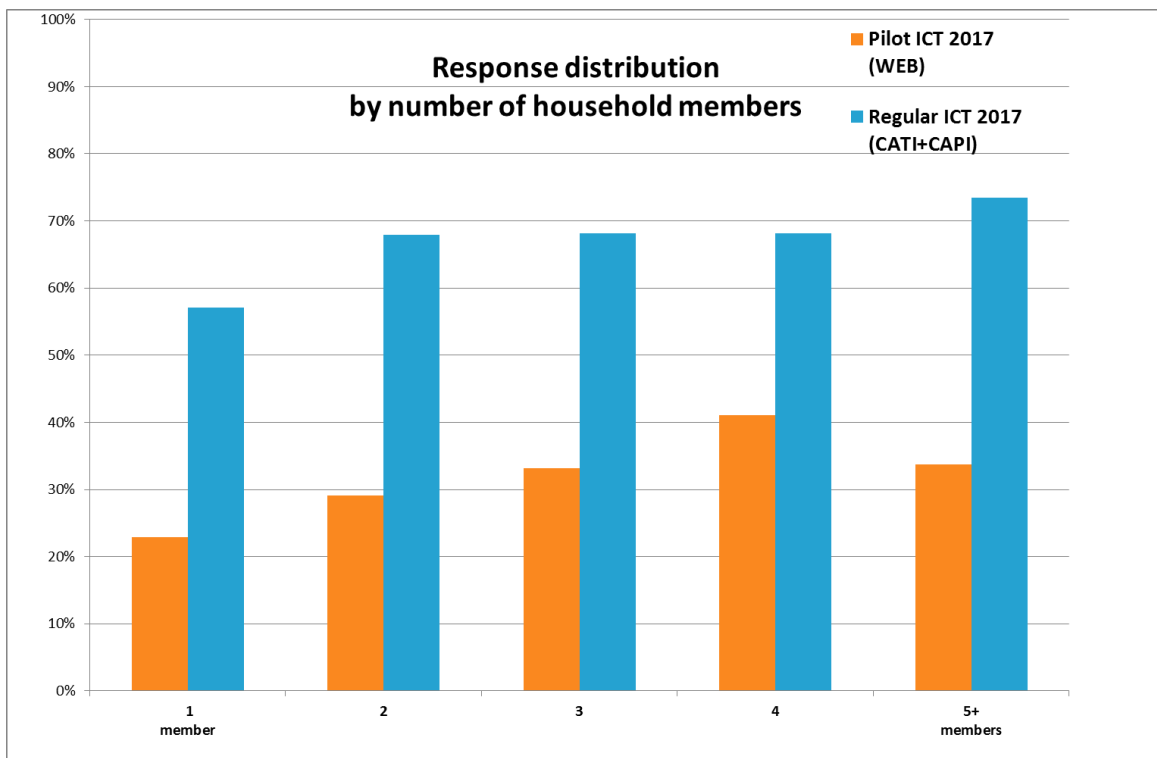
Source: SURS

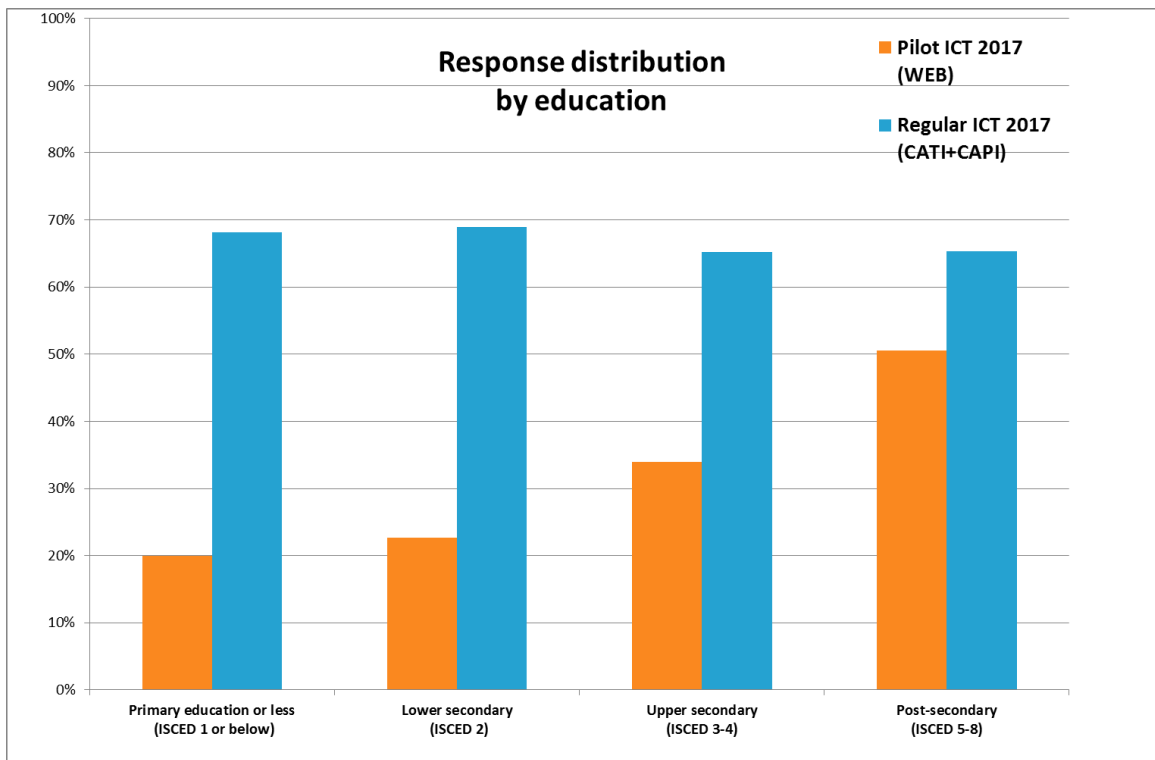Chart 9   Response distribution by number of rooms and tenure status

**Response distribution
by number of rooms and tenure status**

Legend:
■ Pilot ICT 2017 (WEB)
■ Regular ICT 2017 (CATI+CAPI)

Categories: One room, Two rooms, Three rooms, Four or more rooms, Household owns the dwelling, Household does not own the dwelling

Source: SURS

Chart 10   Response distribution by activity status

**Response distribution
by activity status**

Legend:
■ Pilot ICT 2017 (WEB)
■ Regular ICT 2017 (CATI+CAPI)

Categories: Employed, self-employed; Unemployed; Pupil, student; Retired

Source: SURS

12
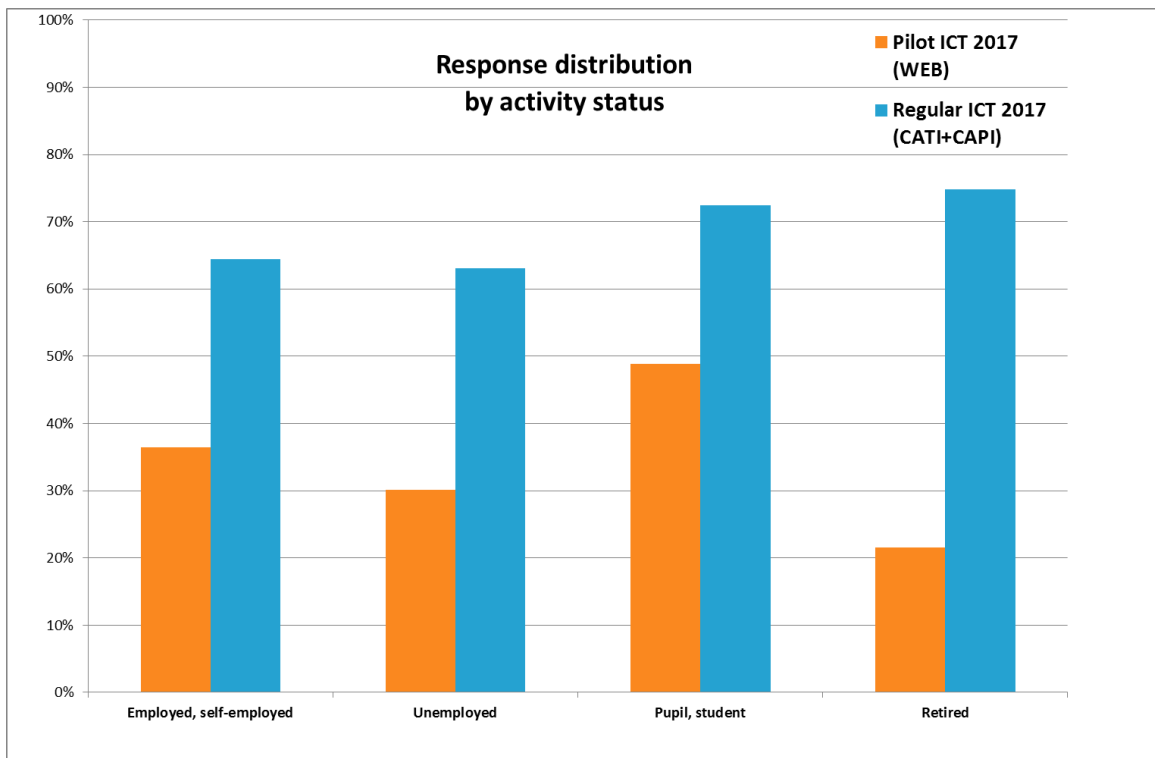
Chart 11   Response distribution by level of education

Response distribution
by education

Source: SURS

Chart 12   Response distribution by type of household
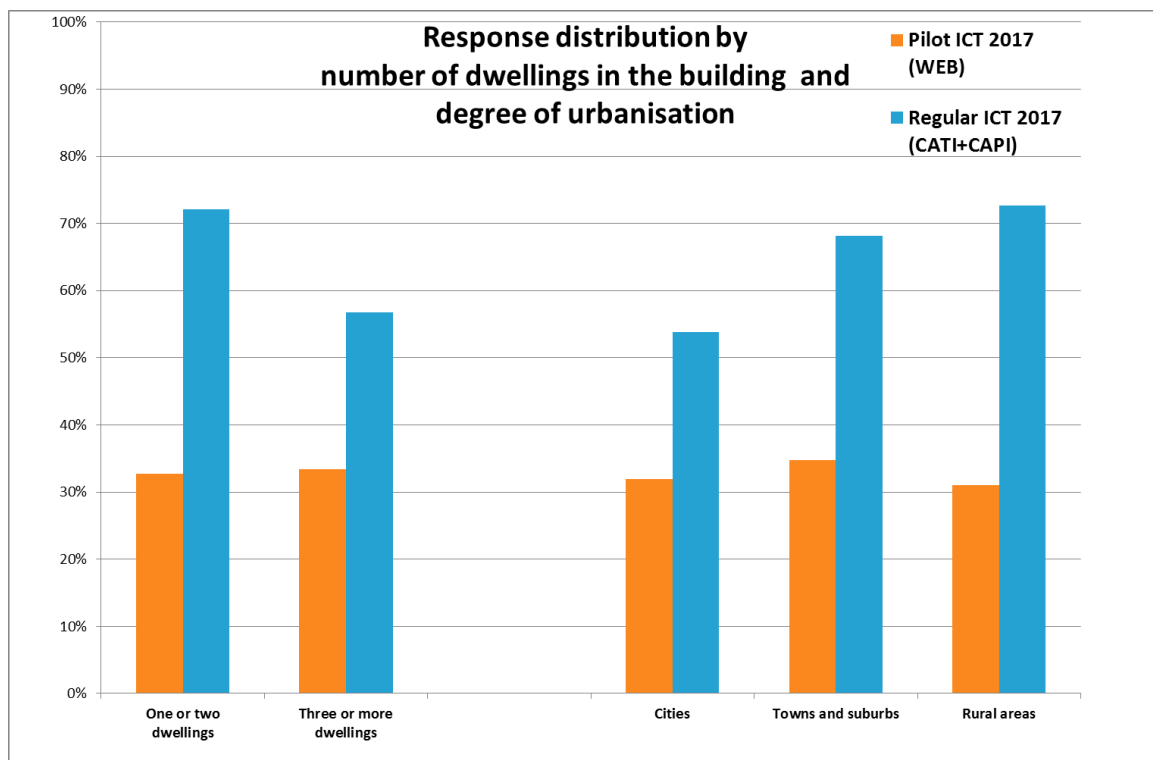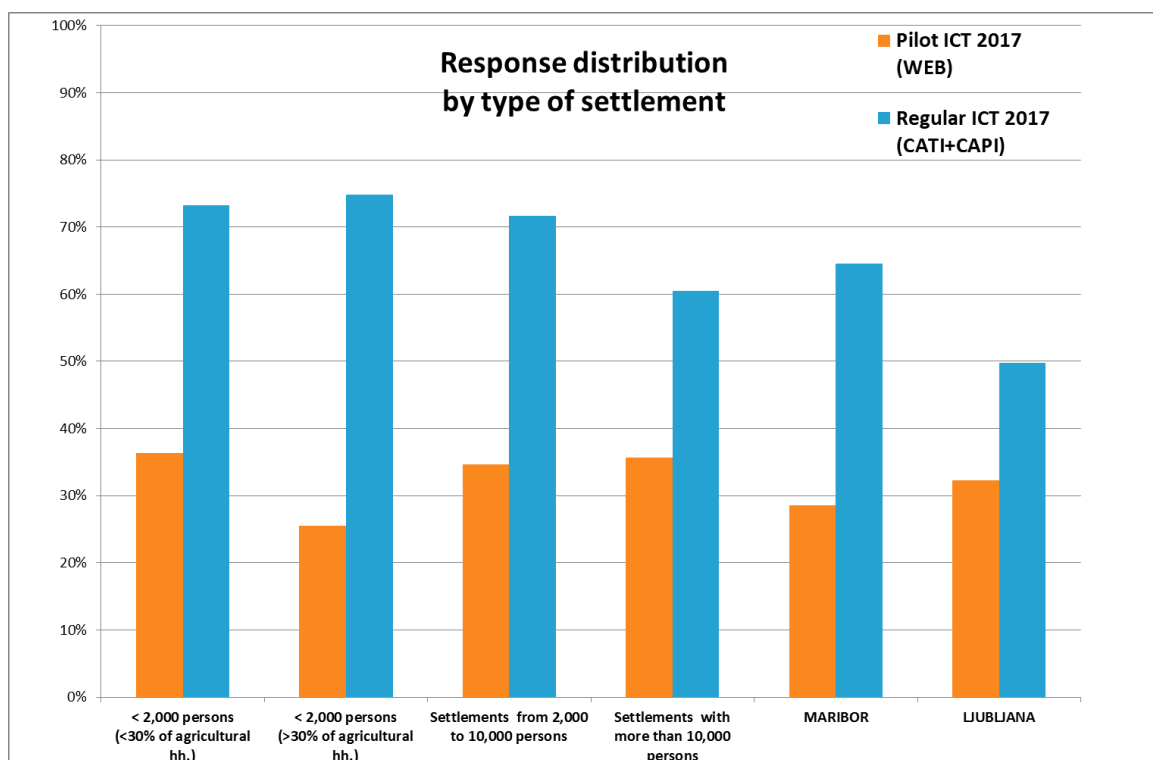

Response distribution
by type of household
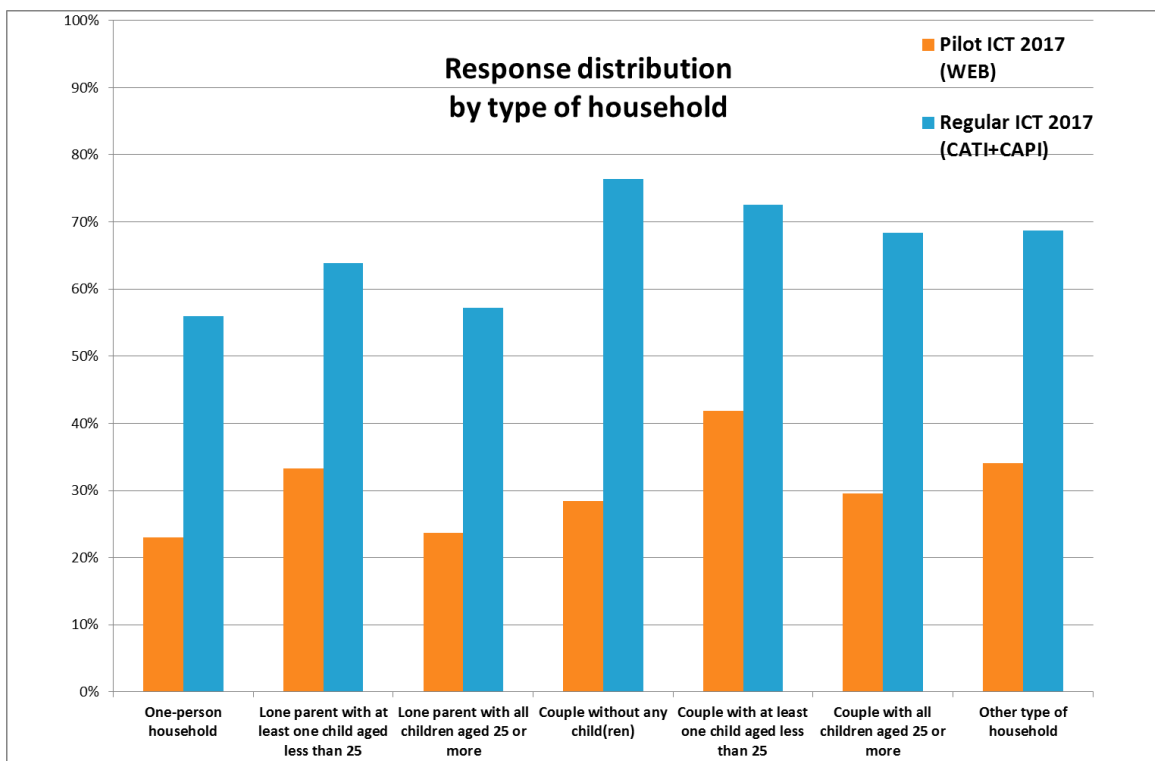
Source: SURS

Chart 13   Response distribution by number of dwellings in the building and degree of urbanisation

Response distribution by number of dwellings in the building and degree of urbanisation

Source: SURS

Chart 14   Response distribution by type of settlement

Response distribution by type of settlement

Source: SURS

Chart 15   Response distribution by household income

**Response distribution
by household income (quintile groups)**
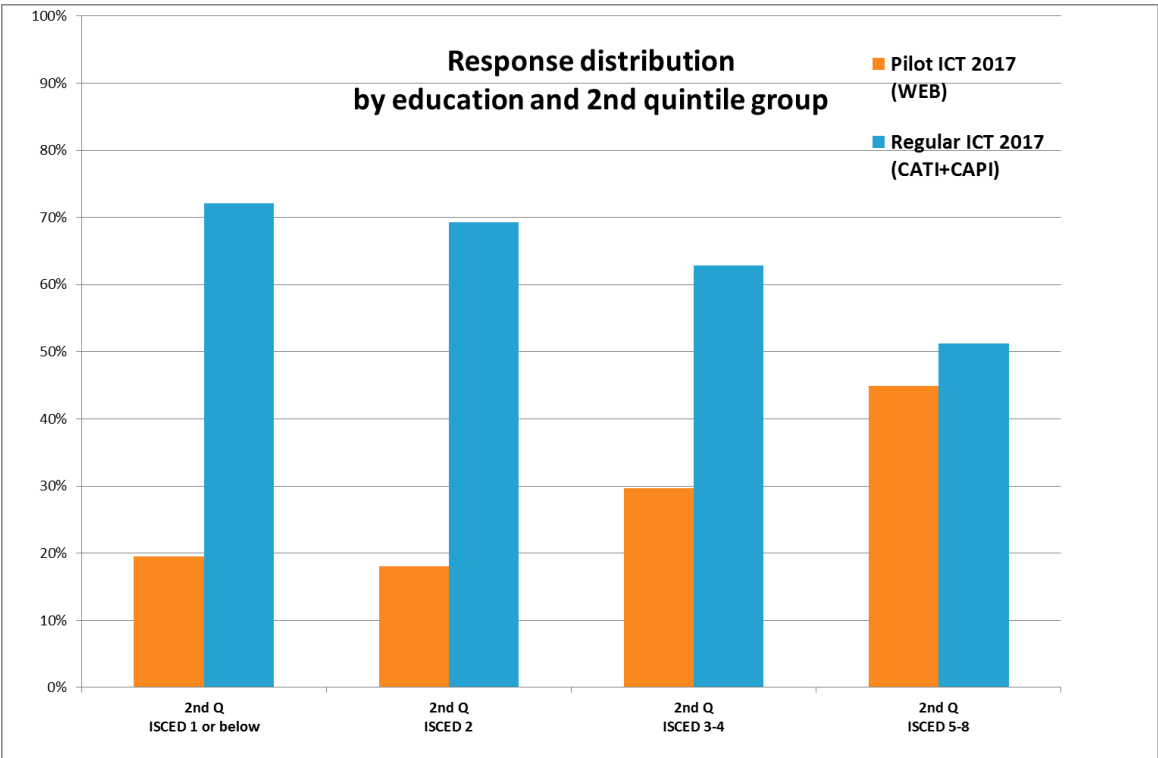
Legend:
- Pilot ICT 2017 (WEB)
- Regular ICT 2017 (CATI+CAPI)
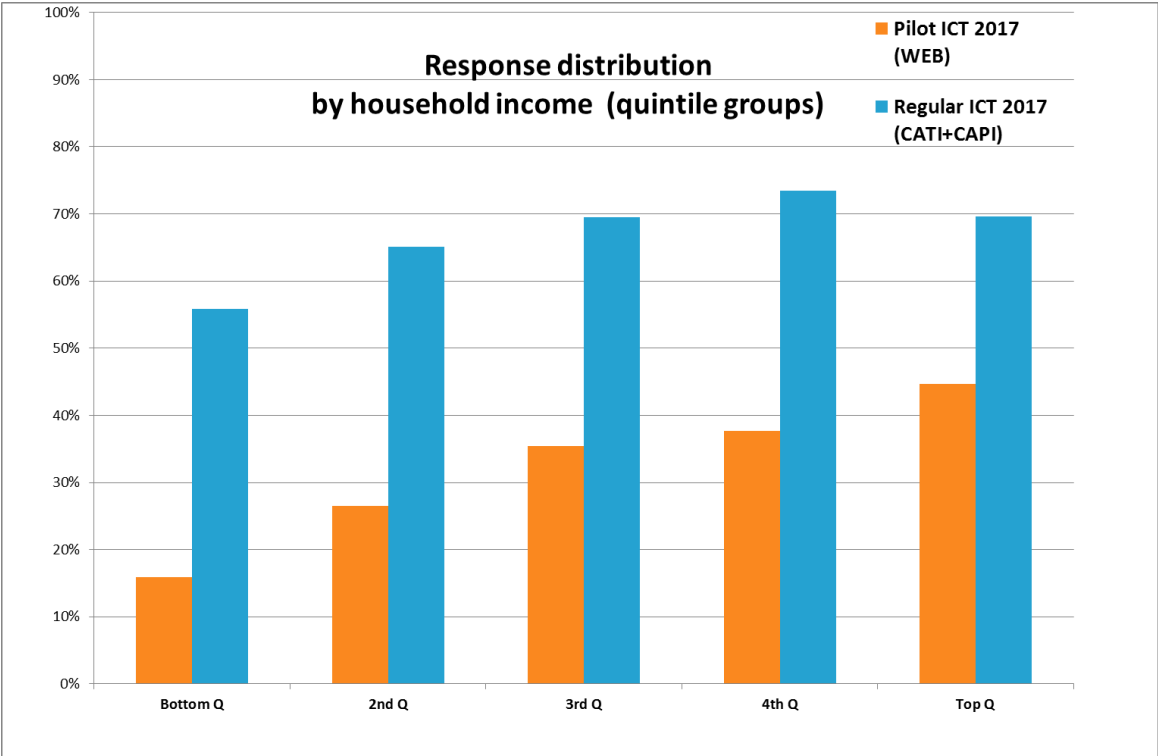
Source: SURS

Chart 16   Response distribution by education and bottom quintile group

**Response distribution
by education and bottom quintile group**

Legend:
- Pilot ICT 2017 (WEB)
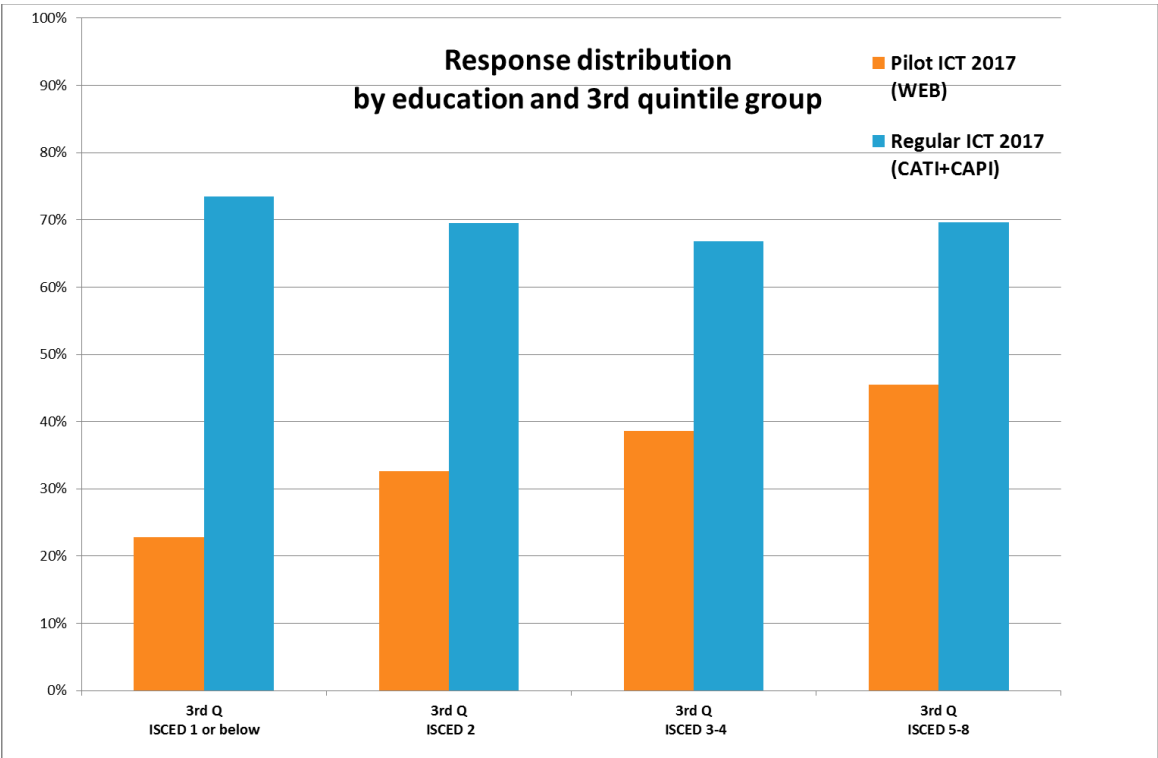- Regular ICT 2017 (CATI+CAPI)
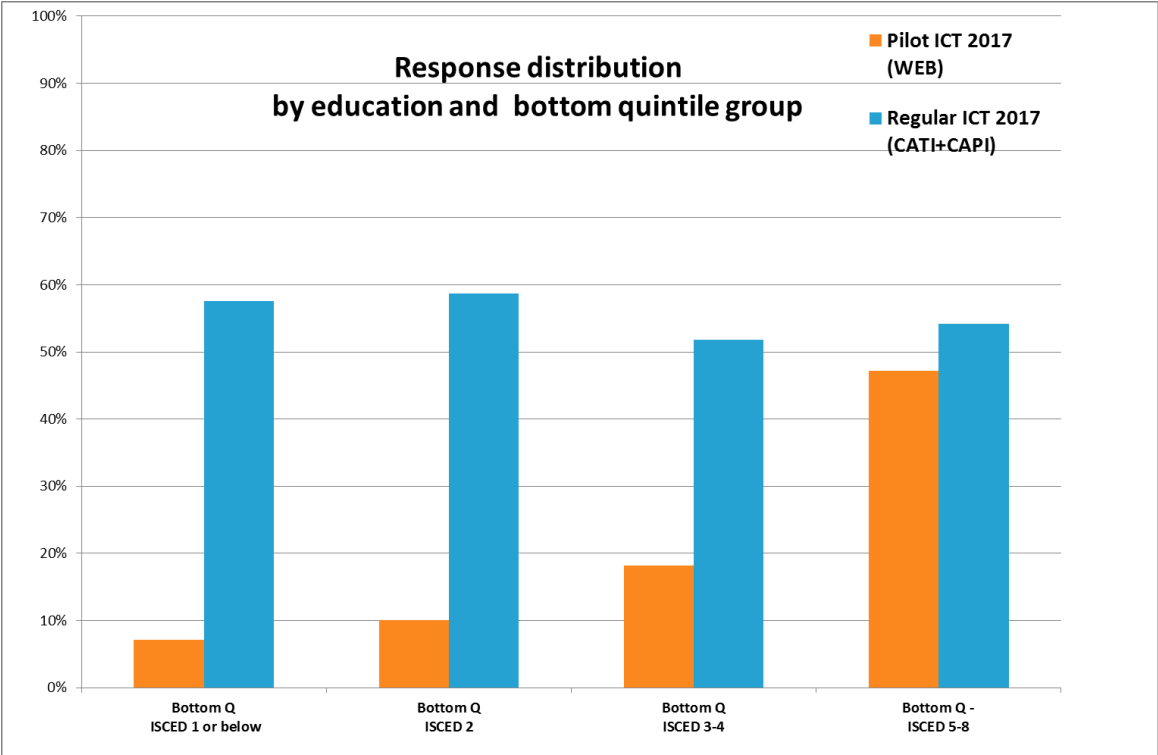
Source: SURS

Chart 17   Response distribution by education and second quintile group
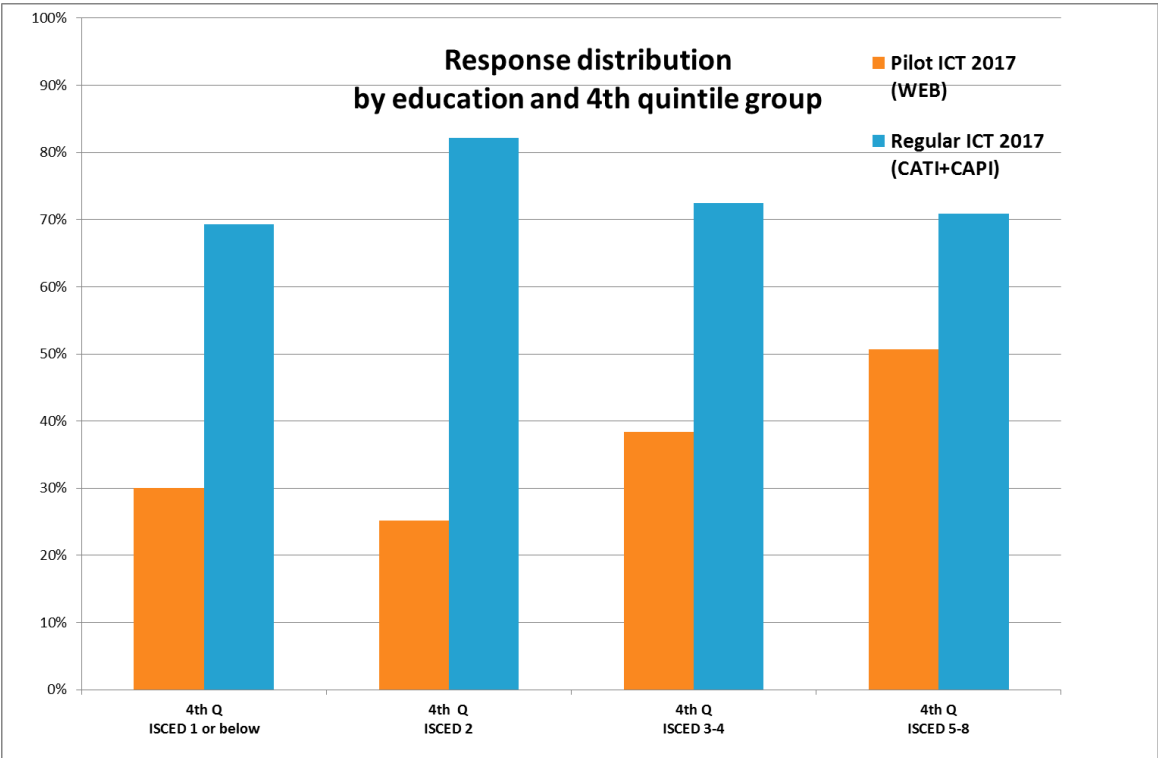

Source: SURS

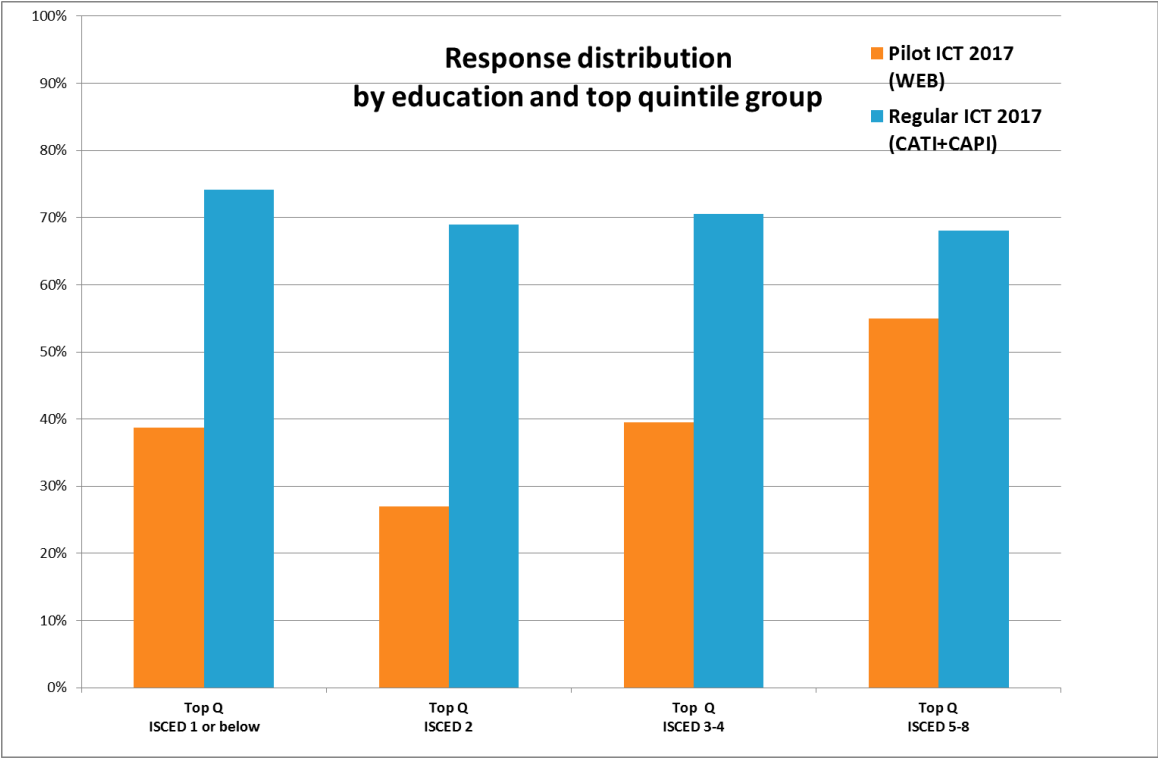Chart 18   Response distribution by education and third quintile group

Source: SURS

Chart 19   Response distribution by education and fourth quintile group

**Response distribution
by education and 4th quintile group**

- Pilot ICT 2017 (WEB)
- Regular ICT 2017 (CATI+CAPI)

Source: SURS

Chart 20   Response distribution by education and top quintile group

**Response distribution
by education and top quintile group**

- Pilot ICT 2017 (WEB)
- Regular ICT 2017 (CATI+CAPI)

Source: SURS

Chart 21   Response distribution by degree of urbanisation and type of housing

**Response distribution
by degree of urbanisation and type of housing**

- Pilot ICT 2017 (WEB)
- Regular ICT 2017 (CATI+CAPI)
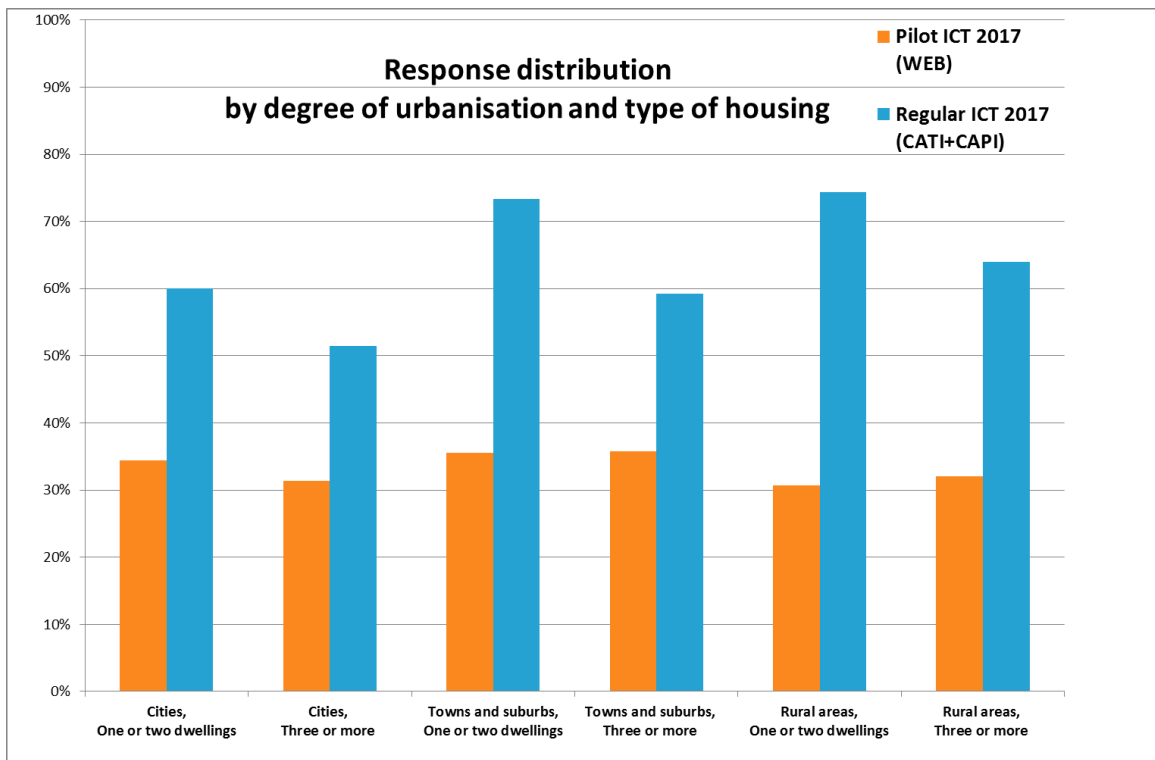
Source: SURS

Chart 22   Finding information about goods or services for regular internet users

**Finding information about goods or services
for regular internet users**

- Regular ICT 2017 (CATI+CAPI)
- Pilot ICT 2017 (WEB)
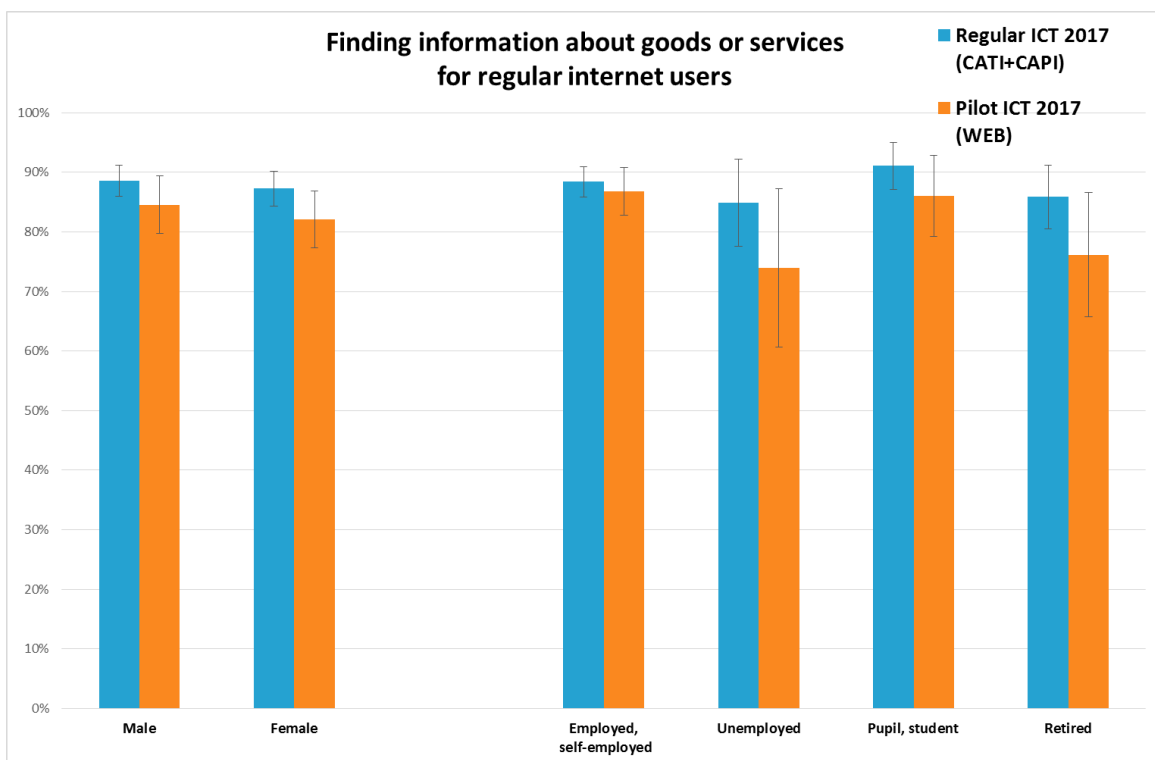
Source: SURS

Chart 23   Internet banking for regular internet users

Internet banking for regular internet users

Source: SURS