

The use of CAWI in the collection of household data in the Danish LFS

Background Statistics Denmark will begin to submit household data from the 1st quarter of 2010, and will thereby comply with Council Regulation (EC) No. 577/98. In this connection, a pilot survey was conducted by Statistics Denmark in the 4th quarter of 2009 in order to examine the possibility of collecting household data by means of a web-based solution (CAWI) and to study the effect of combining this solution with a telephone-based reminder procedure and with CATI interviews as a supplementary data-collection method.

Purpose The purpose of this evaluation is to study the effect of using CAWI and how the reminder procedure with a supplementary data-collection method (CATI) will affect the response-rate. Unfortunately, it has not been possible to analyse more closely the quality of the data collected. This must be postponed until data from the 1st quarter of 2010 are available.

Sample The pilot comprised all households with one person in the core-LFS, who was included in the 4th panel of the reference weeks 40 and 41 of 2009. The sample consisted of 1,424 household members aged 15-74.

Definition of household Two elements are central in Eurostat's definition of a household. Firstly, there must be a joint address for the household members and secondly, there must be, in addition to the rent (housing), considerable economic relations between the household members. The latter part of the definition implies that it is not possible to select the sample exclusively on the basis of the population register, while at the same time complying fully with Eurostat's definition of a household.

Validation test As part of the pilot the possibility of creating a sample by means of the validation, which fully complies with Eurostat's concept of a household, was examined. The validation must be performed one quarter prior to the reference quarter to make sure that the validated sample can be applied.

During the first step of the validation process, information on persons registered at the respondent's address was linked to the Central Population Register (CPR). The outcome of the linkage with the CPR register constitutes a kind of gross sample, which corresponds to a sample exclusively selected on the basis of registers. The gross sample fulfils the requirement that the persons must live at the same address, but does not necessarily fulfil the requirement that there must be considerable financial relations between the household members.

A validation process was tested in order to create a sample that fully complied with Eurostat's household definition. The core-respondent got the opportunity, in this process, to validate the register-based composition of the household by deleting or adding household members, depending on if there were considerable financial relations between the household members or not.

The difference between the validated sample and the register-based sample was about 6 pct. However, this difference can be reduced to less than 3 pct. by means of a combination, which involves that the register-based sample is selected at a

later stage of the production process and that a semi-validation of the sample is performed by adding directly a check question in CAWI. This implies, together with the heavy workload involved in the production process of the validation that the benefit of a complete validation must be considered to be comparatively small.

Pilot design The sample of the 1,424 household members was divided into two groups, each comprising 712 household members. A letter was sent to the two groups stating that they were to participate on the first Monday following the reference week. Furthermore, username and password were sent to the household members in order to enable them to fill out the web-based questionnaire.

Subsequently, there was no further follow-up on the first group, called the ALDI-group. Consequently, the ALDI-group is acting as a control group and represents a situation, where very few resources are used for the collection of household data.

For the second group, called the IRMA-group, the household members who had not yet filled in the CAWI questionnaire after 4 days, were included in the first round of telephone calls. (See figure 1.). The respondents included in the first telephone round were called and reminded. It was also possible for this group to conduct CATI interviews in those cases, where it was considered by the interviewer that this was the only possibility of conducting the interview. The round of calls took place over three days.

Following the three days, the respondents had another two days in which participation in CAWI was possible. The household members in the IRMA-group, who had not yet participated in the survey after this period, were included in a second round of calls. The second round of calls took place over two days.

Figure 1.

IRMA-group	Week 41							Week 42							Week 43							
	Ma	Tu	We	Th	Fr	Sa	Su	Ma	Tu	We	Th	Fr	Sa	Su	Ma	Tu	We	Th	Fr	Sa	Su	
Ref 40	letters				1:st reminder					2:nd reminder					stop							
Ref 41								letters				1:st reminder					2:nd reminder				stop	

Support In cooperation with Capacent, Statistics Denmark made it possible for respondents to obtain support from Statistics Denmark. In the letters that were sent to the respondents, it was stated that Statistics Denmark could be contacted between 9.00 a.m. and 3.00 p.m. if there were any questions or problems. It was, in reality, also possible to obtain assistance outside office hours, where telephone calls were passed on to Capacent. However, this possibility was not stated in the letters.

A relatively low number of persons made use of this support, and no telephone calls were passed on to Capacent. A possible reason for this could be due to the time interval stated in the letter, which did not cover evenings or weekends. In order to examine whether there is any need to provide a longer period of time, this will be tested in the 1st quarter of 2010.

General response-rate The general response-rate achieved in the two reference weeks was 45.9 pct. for the IRMA-group and 20.6 pct. for the ALDI-group. A relatively large share of the interviews for the IRMA-group was conducted by telephone. 118 of the 327 interviews conducted in the IRMA-group were conducted via CATI. One single

person was interviewed by telephone in the ALDI-group. This concerned one telephone call to the support line.

Table 1.

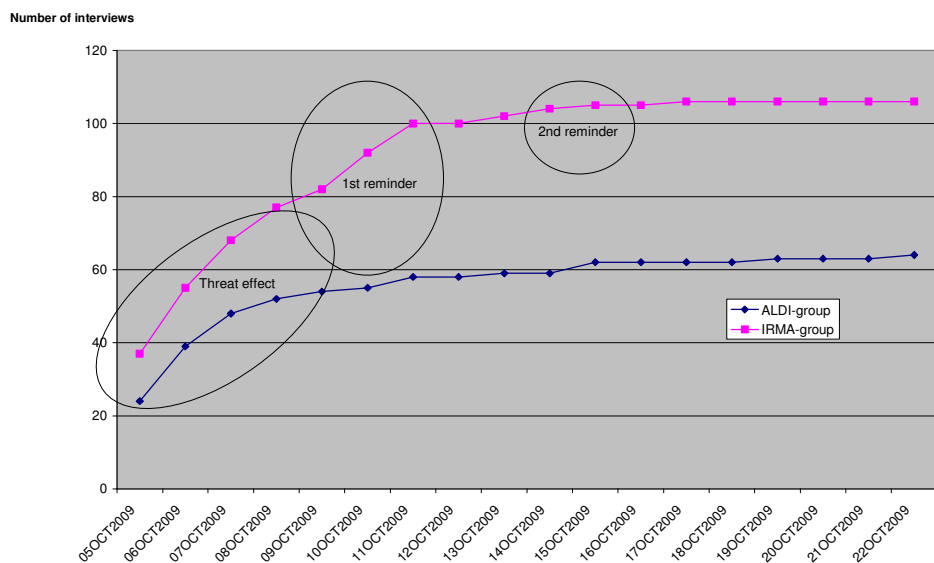
	IRMA	ALDI
Sample survey	712	712
Web-based	209	146
Telephone	118	1
Total response-rate	327	147
Rate of response as a percentage	45.9	20.6

The fact that the IRMA-group had to be reminded twice, combined with the active possibility of conducting the interview through CATI, implied that the response-rate in the IRMA-group exceeded that of the control group ALDI by 25.3 pct. points, i.e. more than twice as high.

Web-based A considerable part of the pilot survey was to test how the response-rate in the web-based part was affected by “the threat effect”. That is, the effect derived from stating in the letters to the IRMA-group that participants would receive a telephone call from Statistics Denmark if they had not filled out the questionnaire within the first four days. Furthermore, it was also important to test the effect of a reminder procedure. This part was tested by means of the two rounds of telephone calls, which were performed for the IRMA-group.

Threat effect reference week 40 The threat effect clearly appears in reference week 40 (see figure 2). The letter they have received is the only difference between the two groups in the first four days. The letter itself resulted in an effect, which implied that the household members in the IRMA-group decided, to a greater extent, to participate in CAWI. For the IRMA-group, the average response rate in CAWI was 2.6 pct. points higher than for the ALDI-group in the first 4 days.

Figure 2. Number of web-based int. accumulated, reference week 40



Reminder procedure reference week 40 Figure 2 also clearly shows the effect of the first round of telephone calls, which was undertaken in the period 9 Oct. to 11 Oct. During these three days (Friday, Saturday and Sunday), the respondents who had not participated in CAWI

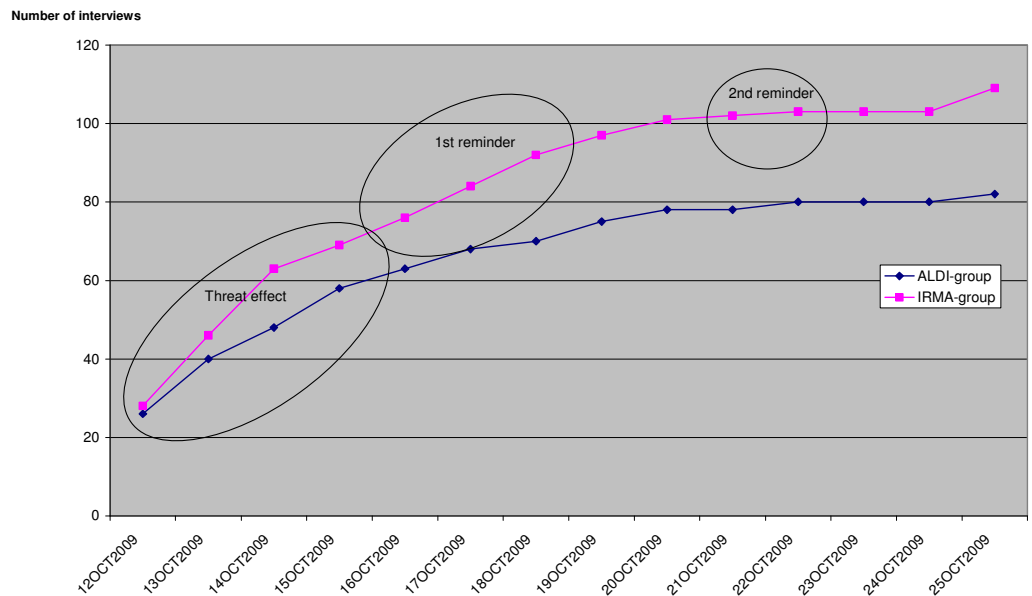
interviews were called and reminded. For the IRMA-group, the number of CAWI conducted increased by 3.2 pct. points in this period. In the same period, the corresponding figure for the ALDI-group was 0.8 pct. point. In other words, the number of household members participating was 4 times higher in the IRMA-group, compared to the ALDI-group, which participated in CAWI interviews during the period when telephone calls were made.

The other round of calls, which took place in the period 14 and 15 October, did not have any noticeable effect. The reason for this could be that the remaining respondents, at this time, are not affected by reminder calls and/or that the reminders of the second round of calls were undertaken on weekdays.

Threat effect reference week 41

The threat effect can also be seen for reference week 41, but the effect here is not as high as that of reference week 40. The difference in the response-rate for the IRMA- and ALDI-group was an average of 1.2 pct. point for the first four days. However, the comparison between the two reference weeks is, to some degree, influenced by the autumn holiday in the period 12 Oct. to 16 Oct.

Figure 3. Number of web-based int. accumulated, reference week 41



Reminder procedure reference week 41

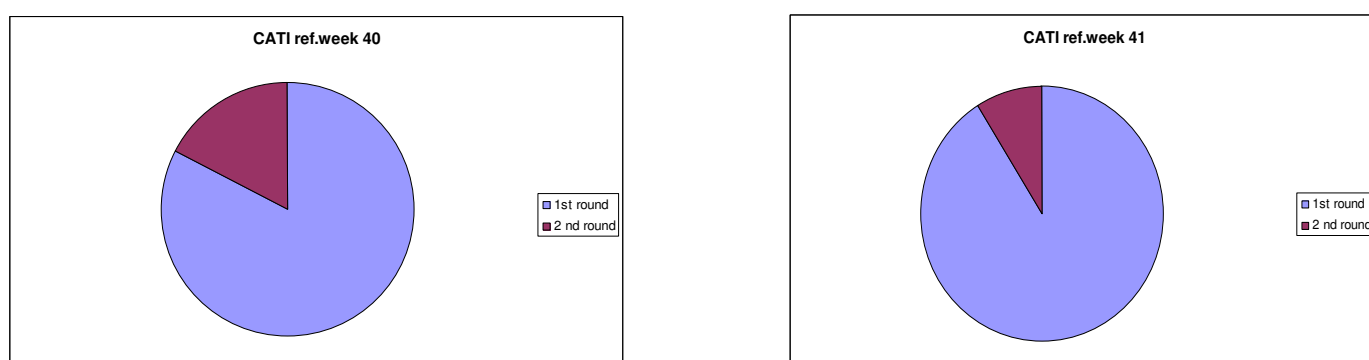
The first round of telephone calls resulted in a relative increase in the number of conducted CAWI interviews. There was an increase in the response-rate of 3.2 pct. points in the IRMA-group after expiry of the period of reminders. By way of comparison, the response-rate in the ALDI-group was increased by 1.7 pct. points in the same period. In the second round of calls, there was no noticeable effect on the response-rate for the IRMA-group, which was exactly as in reference week 40.

The first round of telephone calls had an important influence on the response-rate achieved in connection with CAWI interviews. Partly due to the threat that the respondent would receive a telephone call, if he or she did not participate in CAWI, which in itself leads to a greater increase in the response-rate achieved, and partly because the reminder in itself implies that the respondents, to a higher degree, are inclined to participate. It also seems to be an advantage that the reminder calls take place during a weekend.

The second round of calls did not have any noticeable influence on the participation in CAWI.

CATI CATI interviews were applied in the pilot survey as a supplementary data-collection method for the IRMA-group. CATI interviews were conducted in those cases, where it was considered by the interviewer that this was the only possibility of conducting the interview. A relatively large number of CATI interviews were conducted. All together, 63 and 55 telephone-based interviews, respectively, were conducted for reference weeks 40 and 41. The contribution of the total number of telephone-based interviews to the total response-rate reached 16.6. pct. points for the IRMA-group.

Figure 4. Distribution of CATI interviews by round of telephone calls and reference week



Applying CATI as a supplementary data-collection method has resulted in a considerably higher response-rate. However, the greatest effect was seen in the first week of calls. The second round of calls has only marginally contributed to the total response-rate, as most telephone-based interviews were conducted in the first round of calls. While the second round of calls resulted in a total of 16 CATI interviews for both reference weeks, 103 CATI interviews were conducted in the first rounds of telephone calls.

Effects due to age differences

The number of CATI interviews was relatively high in the pilot. However, this should be viewed against the background that elderly people were overrepresented in the sample of the pilot, which was due to an error involved in selecting the sample.

Figure 5 shows that the share using CATI increases concurrently with age. The group of persons aged 65-74 accounts for the highest share of CATI interviews, while the lowest share is seen among persons aged 15-29. This implies that a falling number of CATI interviews can be expected due to the circumstance that the persons in the sample will have a lower average age in the long term.

A sample with a lower average age will probably also result in a lower response-rate. The highest total response-rate in the pilot was seen for persons aged 55-64. In the IRMA-group, 61 pct. in this age group participated either through CAWI or CATI, while the corresponding figure among persons aged 15-29 was 26 pct.

Figure 5. Response-rate by age and data-collection method

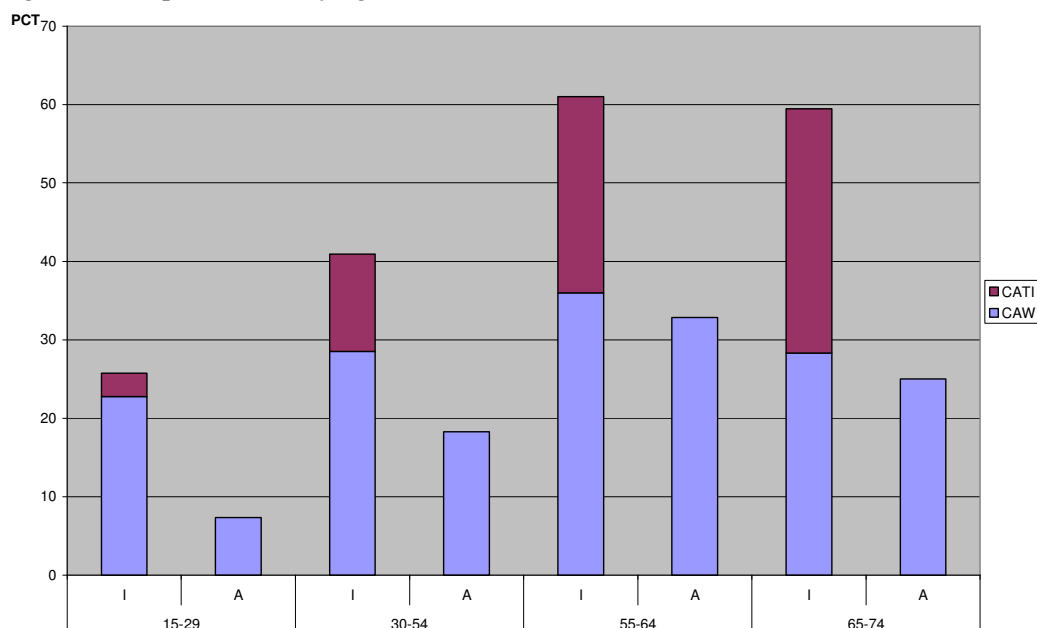


Figure 5 also shows how participation in CAWI differs among the IRMA-group and ALDI-group depending on age, as the relative effect of the reminder procedure decreases concurrently with increasing age. Among persons aged 15-29, 23 pct. of the household members in the IRMA-group participated in CAWI, while the figure for the same age group in the ALDI-group was only 7 pct. For persons aged 55-64 and persons aged 65-74, the difference between the groups was only 3 pct. points.

Consequently, the combination of reminders and CATI as a supplementary data-collection method results in different effects depending on age. The supplementary data-collection method is of lesser importance among young people, while it is of great importance among elderly people. While the reminder procedure is of importance in order to increase the response-rate among young people, the effect of the reminder procedure with respect to elderly people is smaller.

Effects due to sex

More men (54 pct.) than women (46 pct.) participated in the pilot. This can be compared with the core-LFS, where typically more women than men participate. An example is the 2nd quarter of 2009, where the non-weighted percentage of men participating in the survey was 47.

However, the distribution between the sexes differs depending on the data-collection method. 76.5 pct. of the men, who participated in the pilot, did so through CAWI interviews. The corresponding percentage for women was 73. This implied that 55 pct. of the persons who participated in CAWI interviews were men, while the distribution among the sexes in CATI was 50.4 pct. for men and 49.6 pct. for women.

Future setup

The results from the evaluation of the pilot seem to indicate that the combination of reminders and CATI as a secondary data-collection method is of great importance for maximising the response-rate. Simultaneously, the benefit derived from the 2nd round of telephone calls is highly limited. Consequently, a good setup in an operating situation is to undertake only one single round of telephone calls and to supplement the data collection with CATI interviews. Furthermore, the following setup results in a simplification of the data collection,

as it is not necessary to undertake reminders/to conduct CATI interviews over several reference weeks at the same time.

Future resources needed

In connection with conducting the pilot, approximately 30 hours for interviewing were used per reference week in the first round of telephone calls. In the second round of calls, 8 and 11 hours, respectively, were used for reference weeks 40 and 41.

In the 1st quarter of 2010, the number of household members in the sample per reference week will be twice as high, as in the pilot. However, the Section for Interview Services holds the view that this will entail economies of scale, and that 46 hours for interviewing will be used per reference week.

The number of hours for interviewing will be reduced when the average age of the sample is changed. A fall in age will imply that fewer CATI interviews have to be conducted, but it will probably be necessary to make more reminder calls.

In addition to the hours for interviewing, about 63 hours per quarter for other work, which will be carried out by the Section for Interview Services, will be necessary, as well as about 72 hours per quarter to be used internally by Statistics Denmark's Section for Labour Force Survey. This adds up to a total consumption of 707 hours corresponding to 1.56 full time equivalents.

Number of hours, FTE's per quarter. (One round of calls per reference week)

	Hours	FTE's (of 1,850 hours)
Interviewers	598	0.32
Interview Services, other	63	0.03
LFS	72	0.04
Total	733	0.39

Quality issues

One of the main problems when removing the interviewer from the questionnaire is undoubtedly the correct coding of NACE and ISCO. It is known that this may be a problem even when an interviewer is present – how then to produce credible results without this assistance?

In the Danish web questionnaire, this was solved by creating a search box.

The top one is for ISCO – the bottom one for NACE. The respondent is invited to write what he or she thinks is the most appropriate description of their job/business sector, and then press the search button to find the corresponding code, e.g. a search for mailman will give the results seen on the right. The respondent can then click the description that fits best, and then the code will be added to the data.

Afterwards the NACE code can be added in a similar way – on the left a search for post office is shown. Again clicking the Sector description that seems most appropriate will lead to a code being written in the background. Finally the respondent will see a screen

Skriv din stilling i feltet
Tryk herefter på 'Søg' for at vælge den nøjagtige betegnelse.

Skriv din branche i feltet
Tryk herefter på 'Søg' for at vælge den nøjagtige betegnelse.
Som udgangspunkt er vi interesseret i hvilken slags virksomhed det er.
Prøv at søge på virksomhedens type, fx. postvirksomhed, supermarked, folkeskole.

posthus

Resultater 1 - 10 af 12:

Resultatside: 1 2 [Næste](#)

brancheteket
postbetjener
postvirksomhed
udlejning, postboks
postlæge, fremstilling
postordreforretninger, detalhandel
kompostering
posten, engros-handel
leverposten, engros-handel
leverposten, fremstilling
posten, papir, fremstilling

Vælg den ønskede og tryk OK

postbud

Resultater 1 - 10 af 14:

Resultatside: 1 2 [Næste](#)

txt
postbud
landpostbud
overpostbud
pakkepostbud
ekstrapostbud
postbudafeser
postbudformand
reserverpostbud
postbudaspirant
postbetjent

Vælg den ønskede og tryk OK

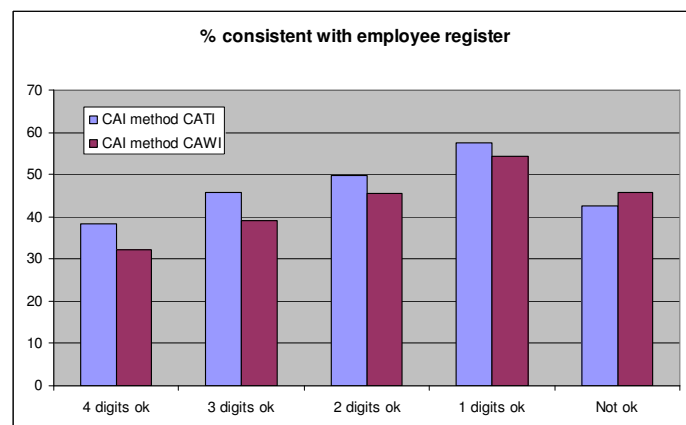
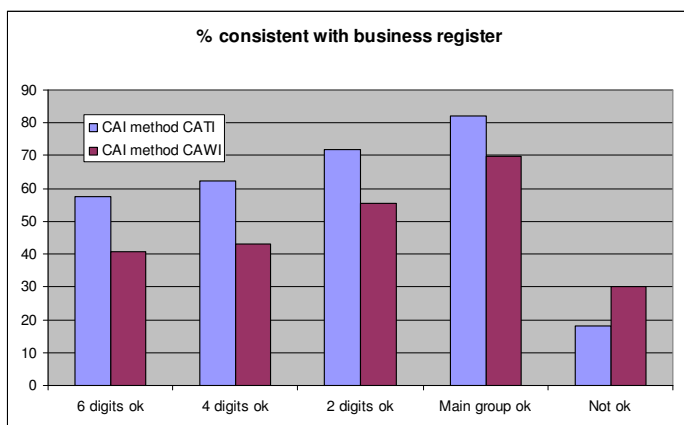
looking like this.

The codes are not shown, as they are deemed not to contain any relevant information to the respondent.

Results When measuring the quality of the codings – two figures come to mind. First of all the amount of respondents with unknown NACE or ISCO. In the ordinary Danish LFS, where interviews are conducted by CATI unknowns are negligible – the rate is zero or very close to zero. This is because of the manual coding, which is added to the automated one, and is a very labour-intensive method. For the household survey the coding is only automatic and the respondent has to find a code that he or she finds appropriate.

Unfortunately the sample size in the pilot study was too small as to make it possible to analyse the data quality in detail. Instead we analysed the responses from the first quarter of 2010. In total there were 1555 employed persons among the CAWI-respondents. Of these 48 or 3.1 pct. had unknown ISCO, while 110 or 7.1 pct. had unknown NACE. This is in line with the general idea that it is normally easier to describe ones own work than the companies' in general.

Secondly, we have in Denmark the possibility to compare the answers to the corresponding value in the register. Because of the long time it takes to update registers, only respondents employed at the same place since 2007 are considered for this comparison. In the first quarter of 2010 there were 1320¹ reference persons, who had been employed since 2007– interviewed via CATI in the standard LFS – while there were 1259 household members, who had been employed that long – interviewed via CAWI. The quality level is compared below.



It is of course by no means certain that the register is correct – but it is a good proxy.

In the ordinary LFS, with manual validation of codes and also the possibility to copy the register information if this seems valid, almost 60 pct. of respondents have consistent codings on the 6 digit level. More than 80 pct. are consistent within the ten main sector groupings used at Statistics Denmark (we very rarely publish on a more detailed level).

In the CAWI-household-LFS the level is significantly lower, as there is no manual validation, and therefore no possibility to apply the register value directly. But even then 40 pct. are consistent on the most detailed level and 70 pct. (12 percentage-points less than in the ordinary LFS) are within the same main sector. When looking at the ISCO codes correspondence is lower, and the interviewer effect seems to be smaller. Differences between register and LFS for ISCO-coding

¹ Only panel 4 of the core-LFS is relevant in this aspect.

are well known at Statistics Denmark. These occur mostly because the responsible recording ISCO at a firm is often far away from the person that is recorded. Therefore they tend to have very different impressions of the kind of work that is done. At the most detailed level (4 digits) the consistency is less than 40 pct. in the ordinary LFS and just over a third in the CAWI-based interviews. This increases significantly when looking at less detailed levels, and at 1-digit level, which is the most detailed we normally publish at Statistics Denmark, the consistency is 58 pct. in the ordinary LFS and 54 pct. in CAWI. This means the quality drop is much lower than for the NACE codes.

- Conclusions*
- Eurostat’s household definition is not fully compatible with a sample selected on the basis of the Central Population Register. However, the benefit derived from a complete validation must be considered relatively small. The difference between a sample complying fully with Eurostat’s definition and a register-based sample is expected to be less than 3 pct. This type of sample can potentially contain a too low number of household members, as it is only possible for the household to delete and not to add other household members
 - The first round of calls has had an important influence on achieving a higher response-rate due to:
 - ”the threat”, stated in the letter sent to the respondents, that the respondents will receive a telephone call from Statistics Denmark has resulted in a higher response-rate in CAWI
 - the reminder in itself increases the participation in CAWI
 - most CATI interviews are conducted in the first round of calls
 - The second round of calls did not have any noticeable influence on the response-rate achieved via CAWI. Furthermore, the second round only resulted in a very small number of CATI interviews. Consequently, only one round of telephone calls is necessary.
 - The results from the pilot seem to indicate that it is an advantage to undertake the reminder calls during a weekend.
 - Different effects, depending on age, are achieved by means of the combination of reminders and application of CATI as a supplementary data-collection method. The supplementary data-collection method is of minor importance in the case of young people, but for this age group the reminder procedure has a relatively great influence on the response-rate achieved. A reverse situation applied for elderly people, where the combination of reminders with a supplementary data-collection method is, consequently, of great importance in order to maximise the response-rate for different age groups.
 - The benefit derived from the second round of telephone calls is highly limited. Consequently, an expedient design in connection with an operating situation is to undertake only one single round of telephone calls and to supplement the data collection with CATI interviews.
 - The quality of the NACE and ISCO codes in the CAWI setup is lower than in the ordinary CATI setup. For NACE 12 pct-points fewer codes match the register, while for ISCO the difference is four pct-points. Taking in to consideration how much less labour is invested in the CAWI setup, the discussion must be whether this is acceptable.