ESSnet DCSS

Mixed mode and web data collection

Experiences in the ESS

Annemieke Luiten

Centraal Bureau voor de Statistiek

Contents

- ESSnet on web and mixed mode data collection
- CBS and other experiences with web and mixed mode data collection
  - How to mix within waves
  - How to mix between waves
  - Who uses the web
- Mode effects
- System requirements

Some words on the ESS

- The **European statistical system (ESS)**
  - partnership between Eurostat and NSIs responsible for European statistics.

- The ESS functions as a **network**,
  - Eurostat plays a leading role in harmonising statistics
  - Work is coordinated with OECD, the United Nations, the International Monetary Fund and the World Bank.

The ESSnet on mixed mode data collection

- Mixed mode data collection current practice in many member states
- Most have moved from PAPI to computer- assisted modes (CATI – CAPI)
- Some have started considering web-based data collection (CAWI)

→ unique opportunity to coordinate and harmonise the introduction of CAWI across the ESS from the onset

→ launch of dedicated ESSnet on mixed mode data collection, including web

Research aims

ESSnet DCSS

- Web data collection, challenges in developing web questionnaires
  - I**mplementation** issues (length questionnaire, unimode vs. functional equivalence
  - **Design** issues (navigation, error checks, instructions, don't know)
- Mixed mode data collection:
  - **Organisation** (fieldwork, case management, design, timeliness)
  - **Mode effects** (understanding, prevention, adjustment)

Status of web and MM data collection

- Questionnaire among ESS members
  - Status quo of mixed-mode and web data collection in social surveys, in ESS and beyond
- Response
  - 34 ESS + 3 overseas countries
  - Response rate in Europe 93%
  - March-June 2013
- This presentation: Some snapshots
  - Modes used in the LFS and in other social surveys
  - Implementation of CAWI instruments and potential for future cooperation

Design	EU-SILC	LFS1	LFS2+	ICT	Census
Single mode					
Capi	10	9	2	2	1
Cati	1	4	6	8	
Papi	6	8	5	7	5
Pap	1			1	4
register					4
Idem + register					
Capi + register	1	2			
Cati + register	3	3	3	2	
Mixed mode interviewer - interviewer					
Capi - Cati	1	2	8	4	
Cati- Capi - Papi	2	2	1	1	
Cati - Papi			3		
Capi - Papi	1	1			
Idem + register					
Cati - capi + register	4	2		1	1
Mixed mode interviewer - noninterviewer					
Capi - pap		1	1		
Cawi - cati	1			1	
Cawi - cati - pap				1	
Papi - pap				1	
Cati - capi - papi - pap				1	
Cati - papi - pap				1	
Cawi - papi - pap					1
Idem + register					
Cawi - capi - cati - register		1			
Cawi - cati - register			1		
Papi - Pap - register				1	
Cawi - capi - register					1
Cawi - papi - register					1
Cawi - Capi - Papi - register					1
Mixed mode noninterview - noninterview					
Cawi - pap				1	5
Cawi - pap - register					1
n different designs					

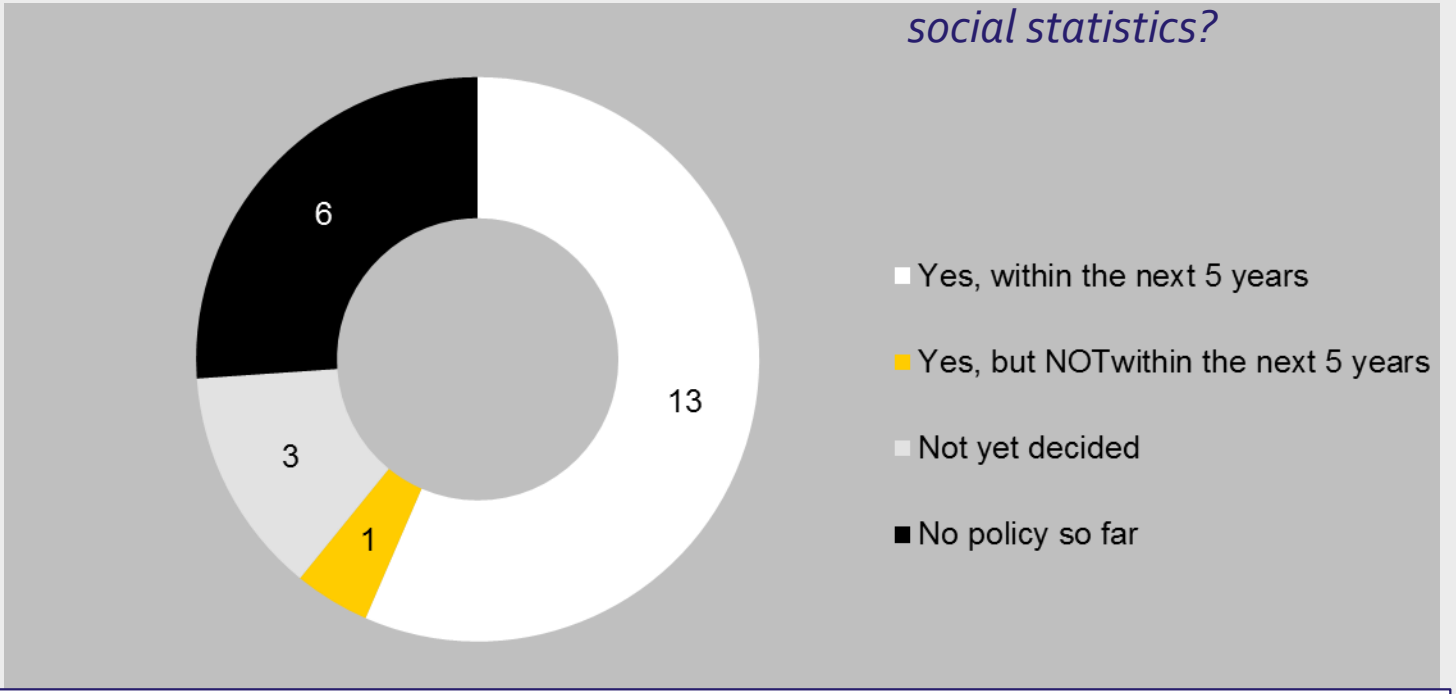
Mixing modes: How? Which ones?

Modes involved

Modes	EU-SILC	LFS1	LFS2+	ICT	Census
Cawi	1	1	1	3	11
Cati	11	11	22	17	2
Capi	19	18	15	11	7
Papi	7	9	7	9	10
Pap	1	1	1	6	9
Registries	7	4	5	2	14

Future Perspective – CAWI Plans

Does your organisation follow a policy towards the implementation of web data collection for social statistics?



→ *More than half of NSIs have concrete CAWI plans for the near future*
→ *40% of NSIs have no CAWI policy or have not decided on this matter*

Functionality and usability issues Web Q

- Don’t know option
- Hard and soft checks
- Error messages
- Navigation
- Individual vs household approach
- Functional buttons (send, save, load, print..)
- Dependent interviewing
- Using grids
- Look and feel
- Welcome page

Critial questions LFS

NL

- Looking for work
- Available
- Educational attainment

DE

- Main Activity Status
- Occupation
- Economic sector
- Second Job
- Full-time/Part-time

What seems to be the trouble?

ESSnet DCSS

- Do we merely have problems with some specific questions?
- Do problems result from the basic concepts of the LFS?
- Or do problems occur due to our choice of mode (CAWI)?
- In principle, are questions and the sequence of questions in the current format practicable?

Conclusions web Q (NL)

- Validity interacts with lay-out issues AND specific content of LFS questions.
- Critical questions need to be rephrased, to give web respondents the right context.
- In the advance letter the purpose of the LFS has to be made more explicit as an interviewer is lacking to convince respondents to participate in the LFS survey.
- Questions throughout the LFS should be personalized for the household members.

Most important reasons for MM - ESS

1. Costs (16)
2. Easier for respondents (10)
3. Response rates (6)
4. Timeliness (4)
5. Coverage (1)

Mixed mode designs used - ESS

- Web is first mode in a sequential design. Non-respondents are followed up in other modes.
  - NL (LFS and other SS), Bulgaria (census), FR (ICT), SP (EHSIS), US CB (ACS).
- Web is offered concurrently, alongside other options, as a choice for respondents.
  - DK (LFS 2), BE (SILC), IT, HU, UK (census), SW (HS), NO (RS), Switzerland (for the Structural Survey)
- Web is offered to non-respondents of earlier CATI or CAPI trials
  - FR (LFS)
- Web is offered in subsequent waves, after a first wave in a different mode.
  - DK (LFS) and NO (RS). DE is considering (LFS).

Mixed mode designs - continued

Which design is chosen, is closely linked with the **purpose** of the mixed mode design:

- If purpose = decrease coverage or non-response error:
  - Concurrent design where respondents are given a choice of modes,
  - Consecutive design where the most expensive mode is presented first
- If purpose = reduce costs:
  - Consecutive design with cheapest mode first

Measurement errors - ESS

Evidence from ESSnet that measurement errors are an important source of differences between modes in some but not all surveys

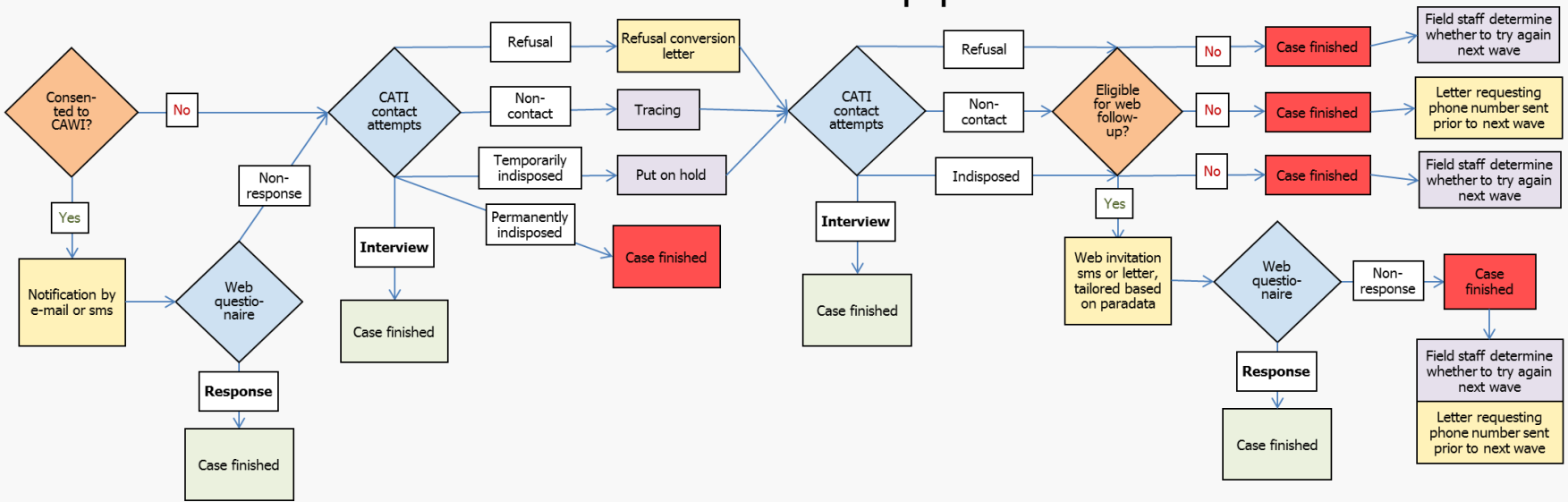
- LFS: differences between modes (mail, web, cati, capi) can be explained with common weighting variables (NL – DE)
- Safety Monitor: large mode differences (NL) → restriction of modes
- Opinions Survey: mode effects can be explained, but need more than standard weighting variables (UK)
- Consumer sentiments: large mode differences (FI) → no introduction of web

Costs - ESS

- Implication of MM for costs is a function of the **MM design chosen**, and the **design prior** to the introduction of the additional mode.
- Introduction of web will save costs, but other designs will not necessarily do so.
- While data collection costs mostly went down, other costs, like costs of sampling, data processing, case management, and adjustment may increase as a result of the greater complexity of the mixed mode process.
- A number of countries therefore specified that some costs went up, while others went down.
- Some indicate that, although other costs may have gone up, this is easily compensated by the decrease in costs of data collection.
- Most countries foresee that within the next five years mixed mode data collection will become more profitable, especially by the introduction or the increased use of web surveys.

1.1 How to mix within waves - NO

- Depends on the objective
  - Save money: cheapest mode first = web
  - Maximise respose rate: most efficient mode first, cheaper modes for follow-up
  - Maximise representativity: different modes for different respondents based on mode-specific response propensity
  - Combination of objectives: Finding the right moment for switching modes + use of CAWI in both initial and follow-up phase:



1.2 How to mix between waves - NO

- Use interviewers/CATI to establish contact and recruit into the panel, and for 'difficult' cases. Use CAWI once contact has been established for 'easy' cases
  - Proposed between-wave mixed-mode design for the LFS:

First wave	Second wave
CATI main mode for entire sample	CAWI main mode for consenting persons with standard working/working time arrangements
	CATI main mode for non-consenting persons with standard working/working time arrangements
	Main mode CATI for persons with non-standard working time arrangement and the unemployed
	Main mode CATI for 1st wave nonrespondents

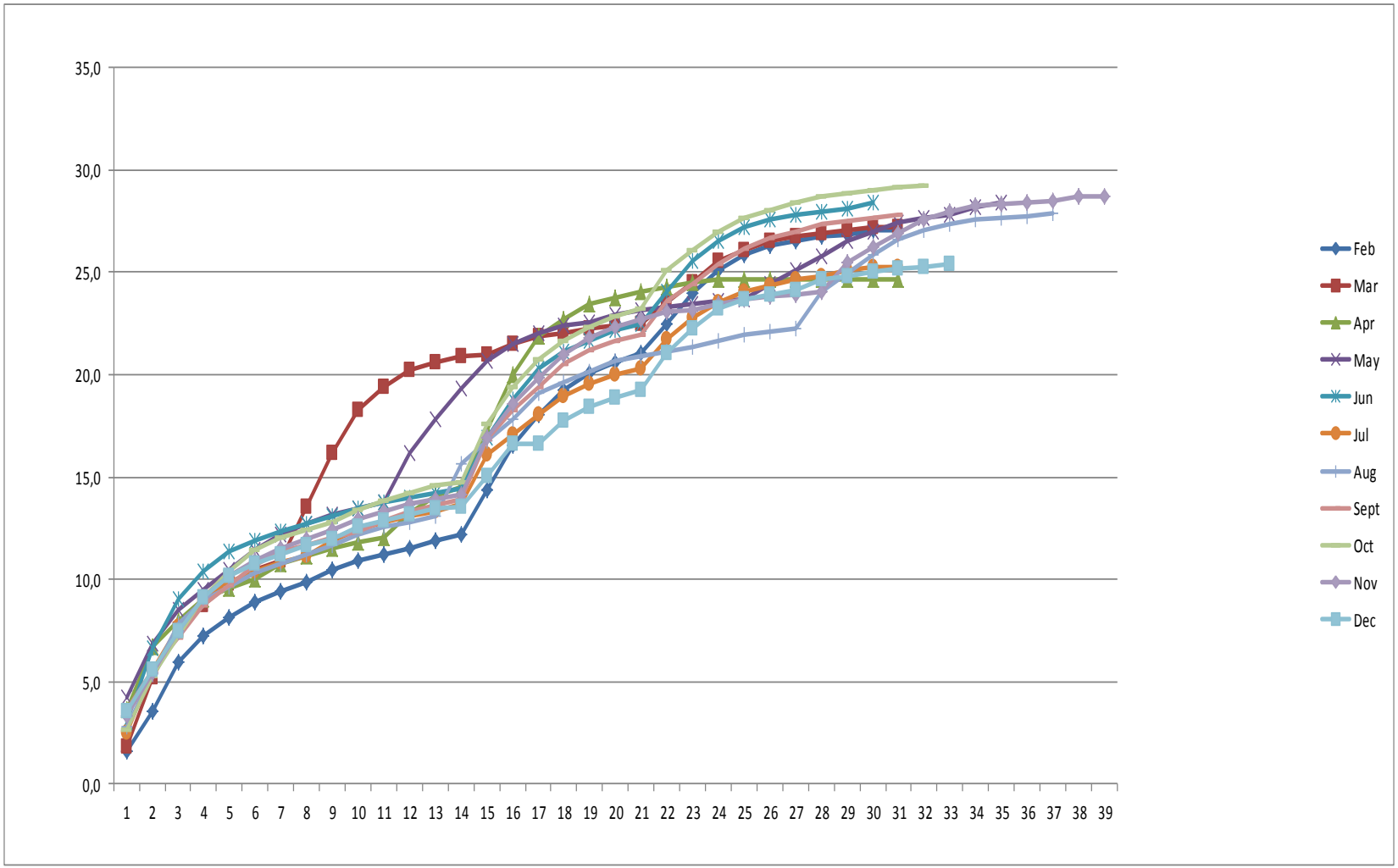
1.2 How to mix between waves - NL

% Recruitment subsequent waves	cawi	cati	capi	total
No objection	66	93	88	79
recruited and telephone provided	27	93	88	59

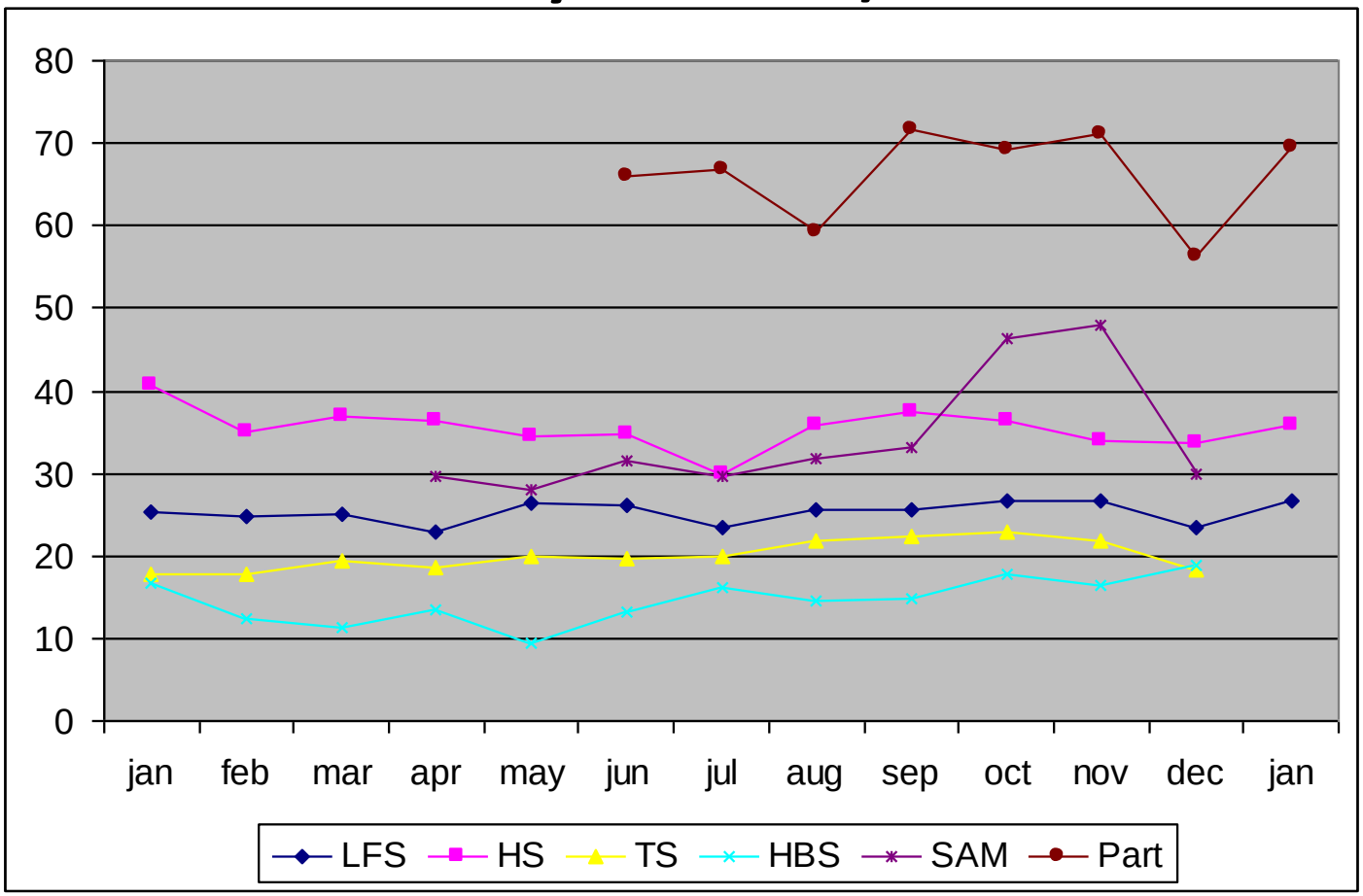
	cawi	cati	total
% Response wave 2	22	76	69
wave 1 = cawi	36	75	56
wave 1 = cati	7	79	80
wave 1 = capi	16	75	79

- Recruitment for following waves much lower in CAWI
- Effect made worse because low number of respondents providing telephone number
- Low cawi wave 2 response, especially …
- Cawi response by HH that did not response in CAWI in wave 1 extremely low

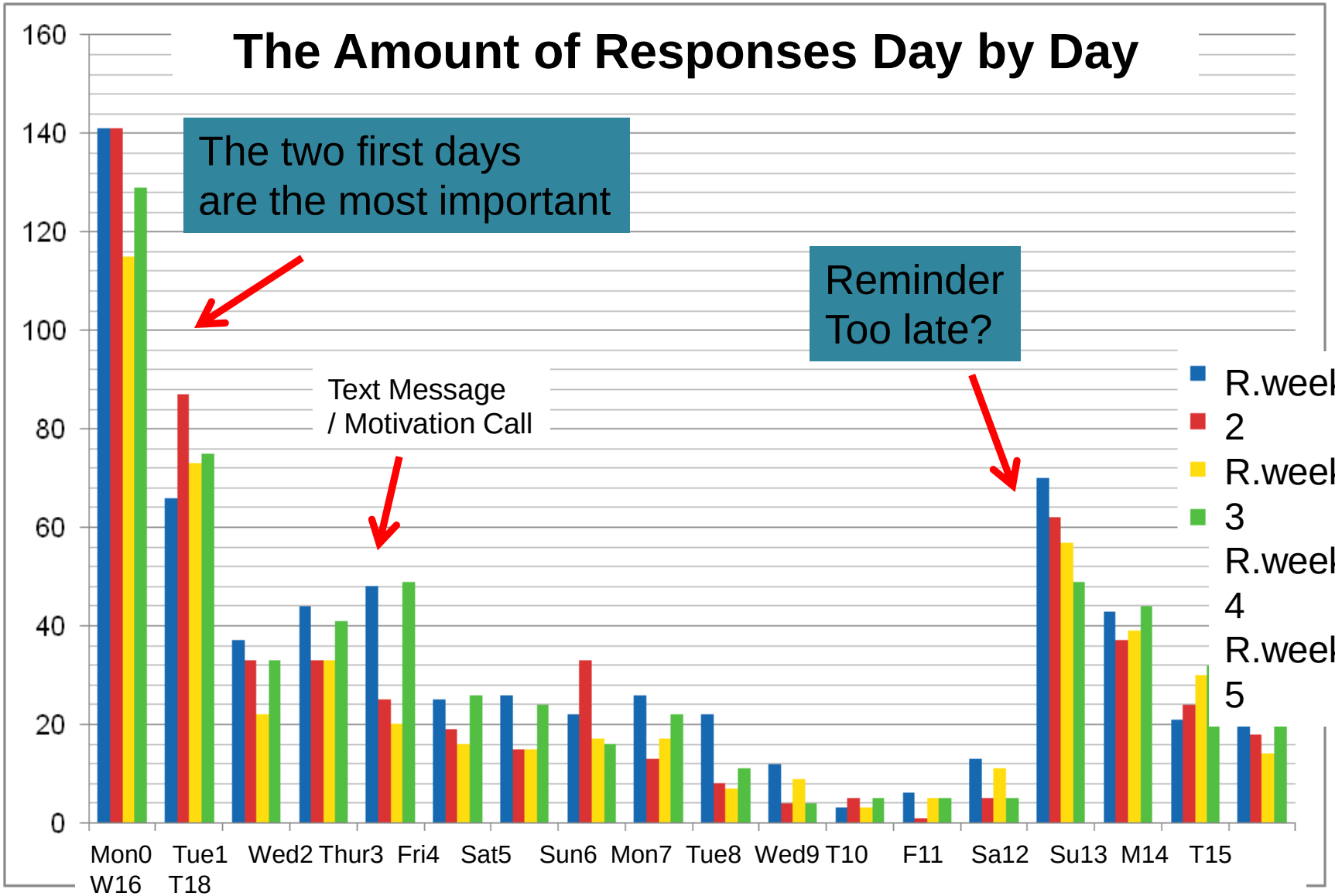
5.1 Response by fieldwork day – LFS - NL



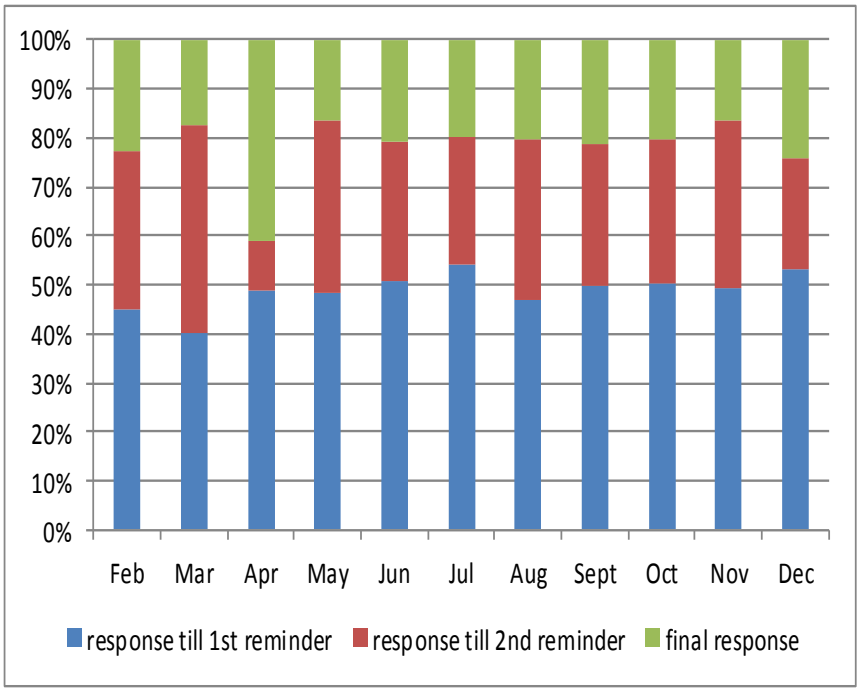
Web Response Rates (first or only wave) - NL

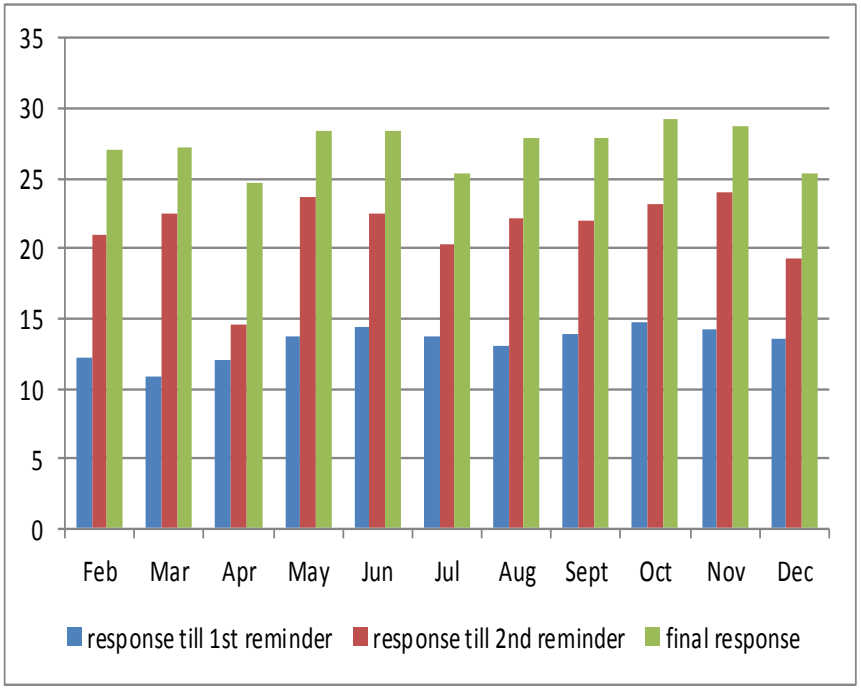


5 Timeliness



2.4 Response by reminder LFS - NL

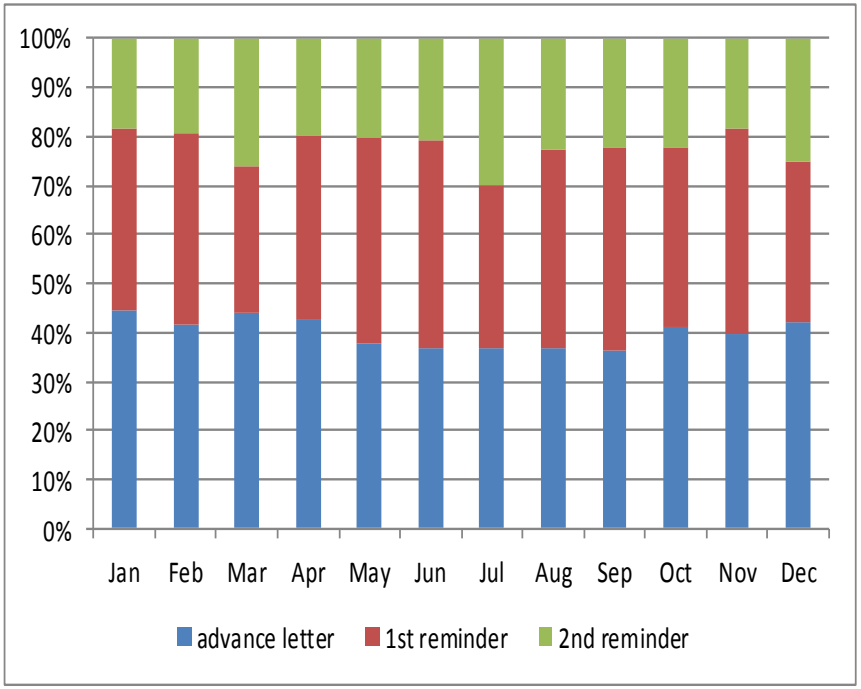


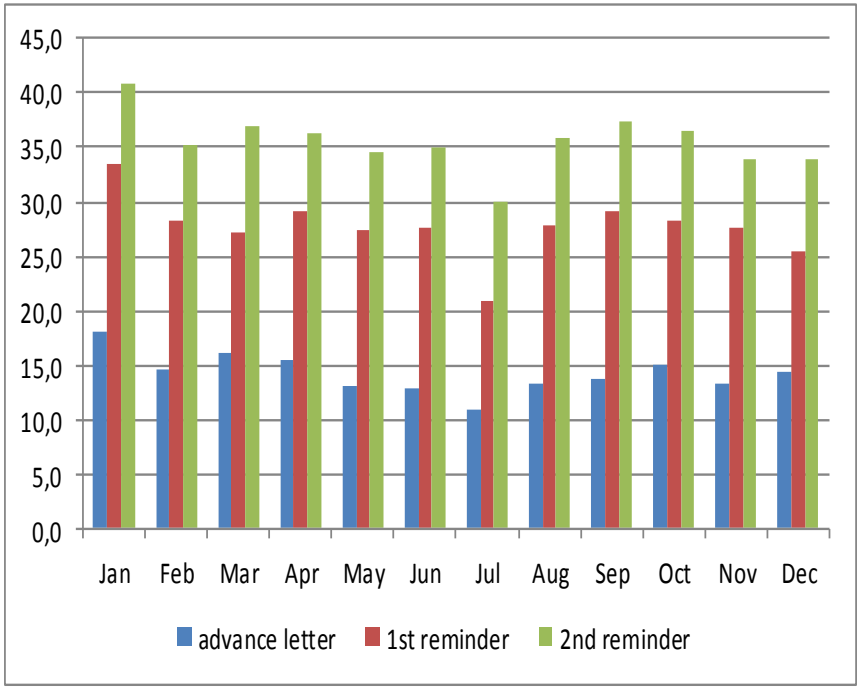


50% of response after advance letter
30% of response after 1st reminder
20% of response after 2nd reminder
Fieldwork period: 4 weeks
1st reminder after 2 weeks
2nd reminder after 3 weeks

Mean response: 27,3
Max response: 29,2
Min response : 24,7

2.4 Response by reminder HS - NL

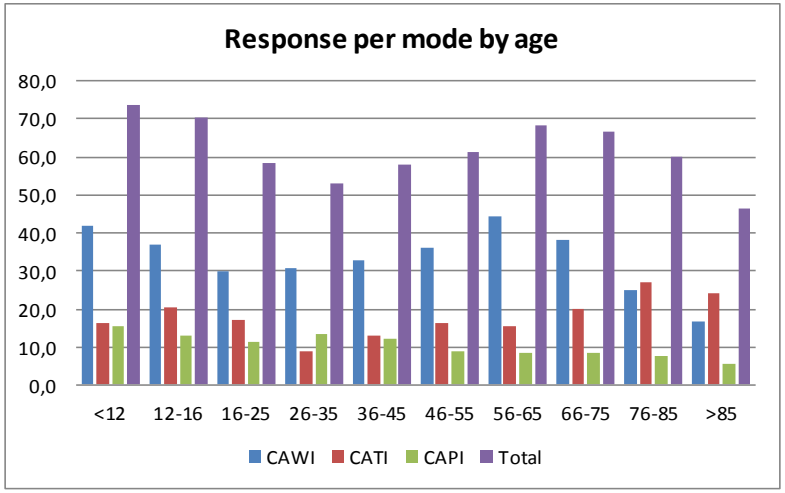


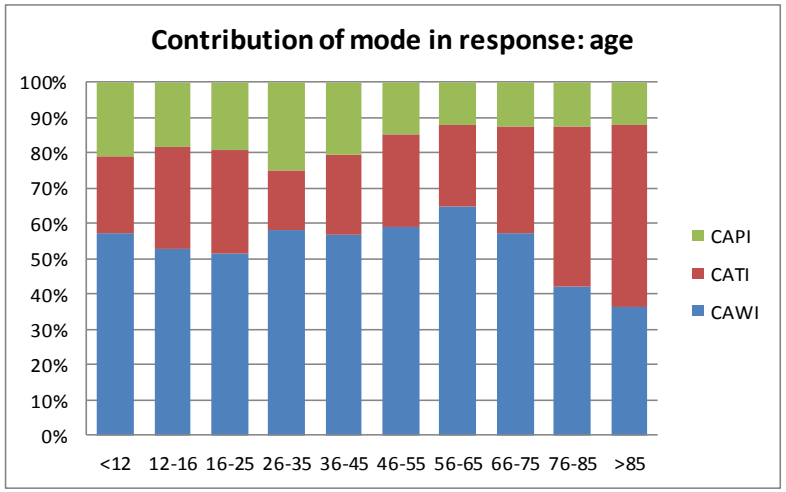


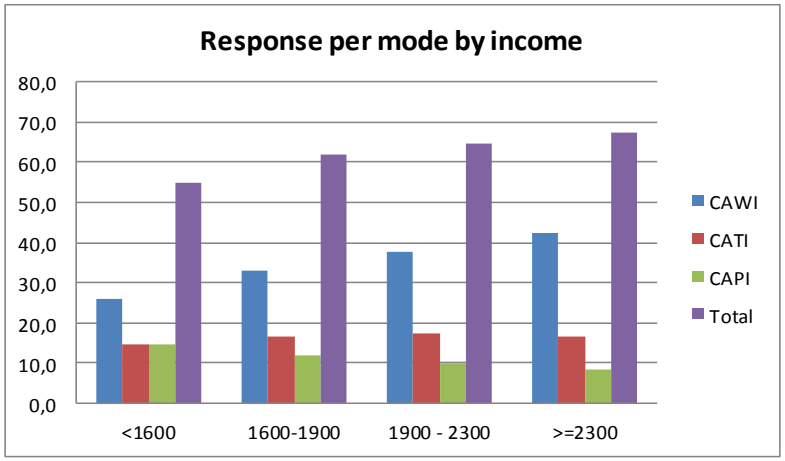
40% of response after advance letter
40% of response after 1st reminder
20% of response after 2nd reminder
Fieldwork period: about 3 weeks
1st reminder after 1 week
2nd reminder after 2 weeks.

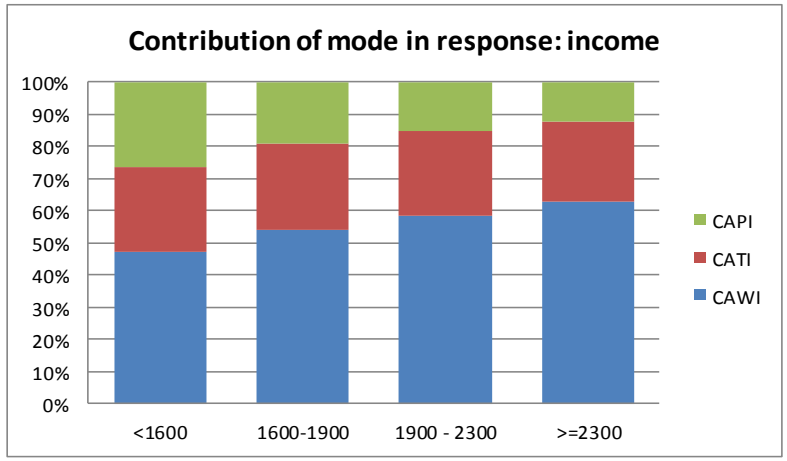
Mean response 35,5
Max response 40,9
Min response 29,9

2.5 Who uses the web* - NL



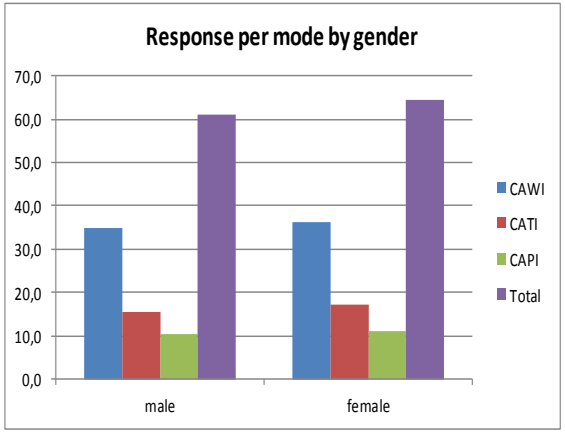


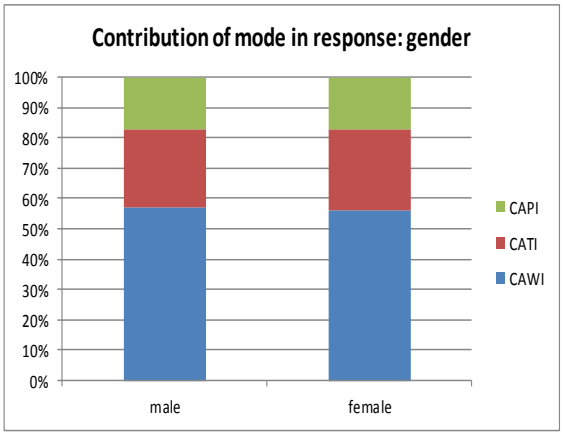




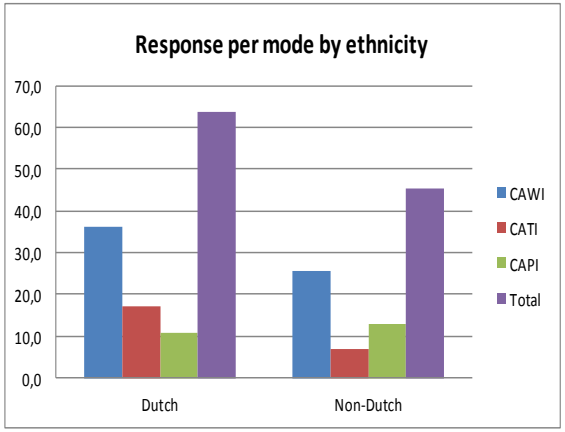
- Web is used most by the age 56-65 group (and parents of young children)
- Surprisingly high use in the age 66-75 group
- Web response lowest in 85+ age group, but still present (35% of responses in this group)!
- The higher the income, the higher the response, and the higher the web response

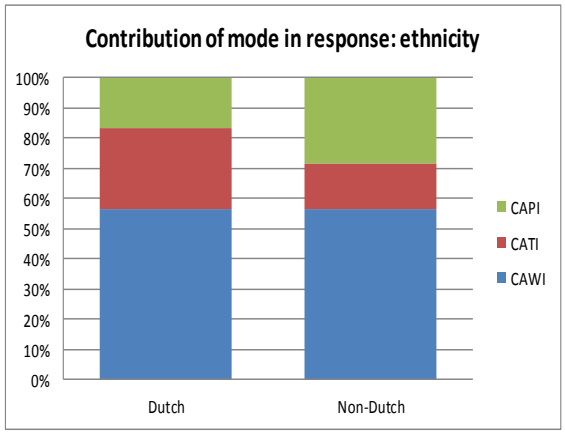
* Health Survey



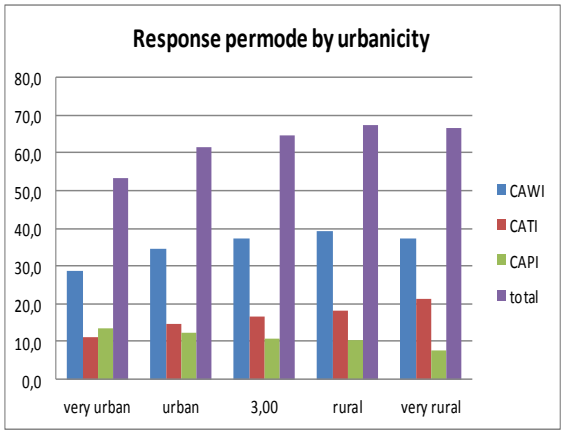


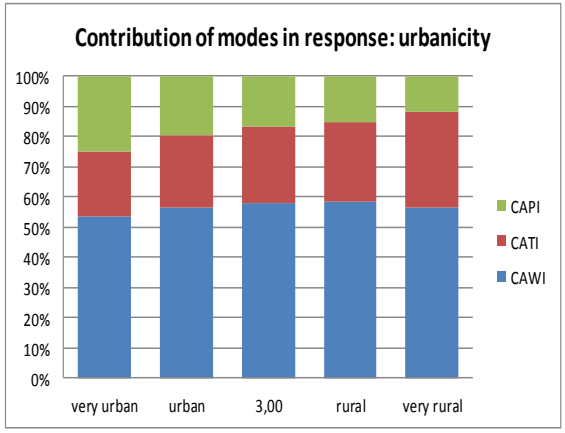
- No difference at all by gender





- Higher response for Dutch, but relative use of web is the same in both groups

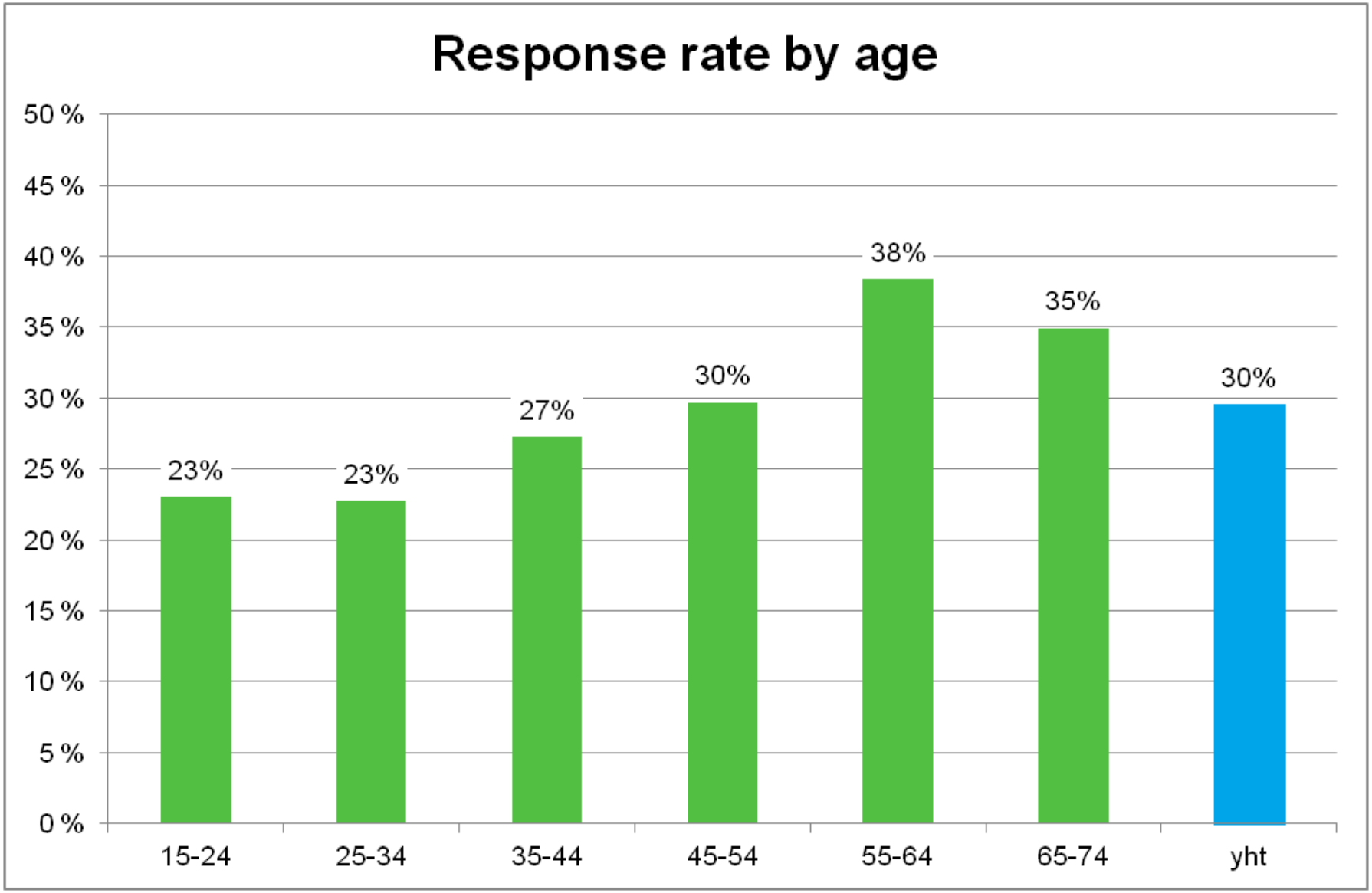




- Lower web response in urban areas, but relative contribution the same.

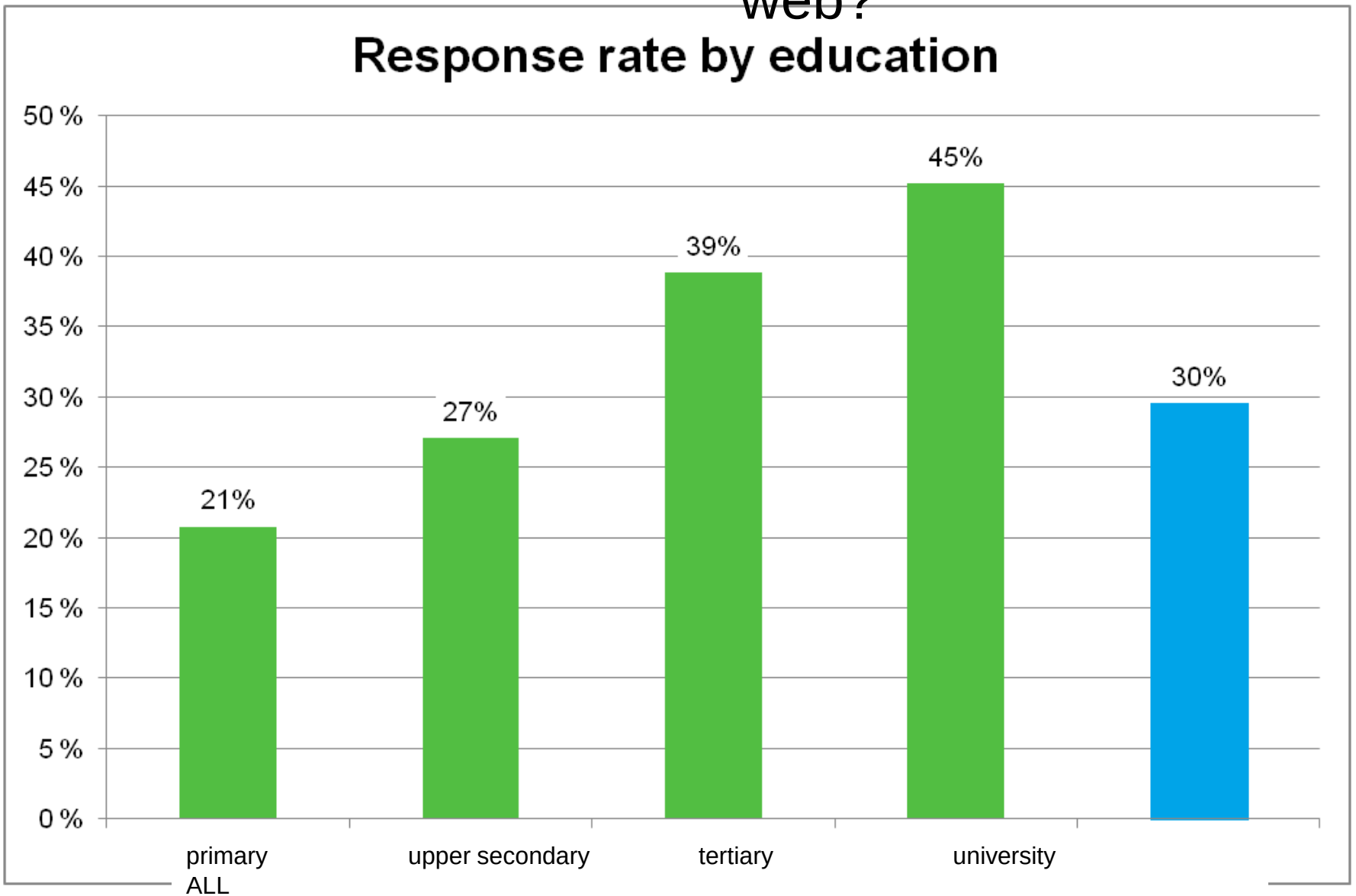
Statistics Finland

2.5 Who uses the web?



Statistics Finland

2.5 Who uses the web?



2.5 Who uses the web?

- Data from follow-up pilot
  - Representativity perspective
- High response rate among
  - Women
  - Age group 30-44
  - Highly educated
  - Previous round respondents

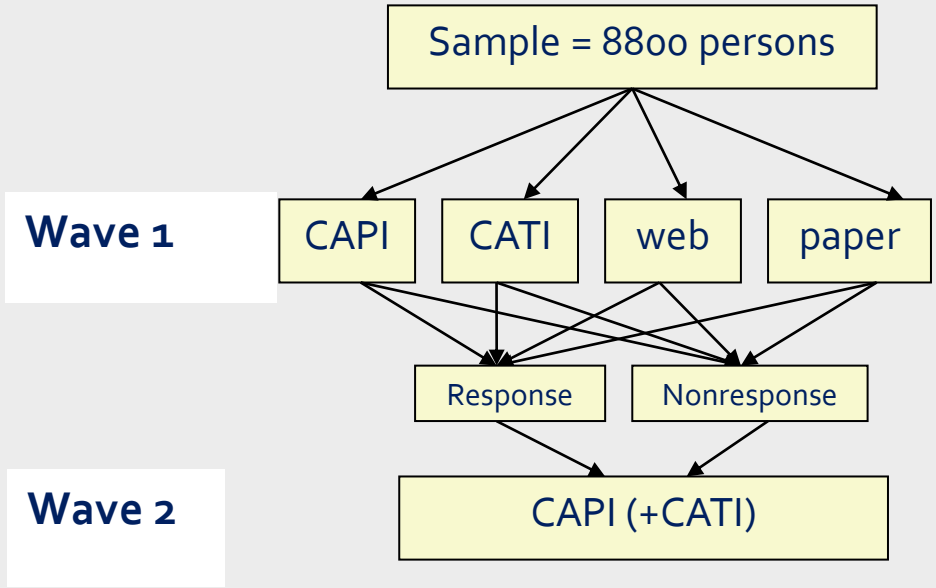
	Response rate	Gross sample
Total	17.0	599
Gender		
Men	13.6	301
Women	20.5	298
Age groups		
16-29	14.8	171
30-44	28.6	207
45 and older	20.2	155
Urbanity		
3 largest cities	16.3	135
Rest of country	17.2	464
Education		
Low	8.4	214
Middle	16.4	238
High	30.6	147
Result 2nd quarter		
Respondents (CATI)	31.1	171
Nonrespondents	11.0	428

Minimizing mode effects

Methodological options:

- Prevent by questionnaire design (WP II)
- Avoid by (adaptive) data collection design (WP III)
- Adjust by estimation design (WP III)
- Stabilize by calibration on mode (WP III)

Decomposition: Experimental design



Mode effect decomposition LFS

Employment status (CAPI is benchmark)

	Coverage bias		Nonresponse bias		Measurement bias		Total bias	
	CATI	Web	CATI	Web	CATI	Web	CATI	Web
Employed	-0,4%	4,2%	0,1%	5,5%	0,7%	-0,7%	0,4%	8,9%
Unemployed	0,0%	0,0%	-1,1%	-0,9%	-1,9%	-1,7%	-3,0%	-2,6%
Non-labour	0,4%	-4,2%	1,0%	-4,6%	1,2%	2,4%	2,6%	-6,3%

Also available:

- Decompositions for employment status subquestions (#jobs, wishing a job, searching a job, being available for a job) and for educational level
- Decompositions for employment status for age and gender strata
- Decompositions LFS key variables as a function of contact effort

Mode effect decomposition LFS

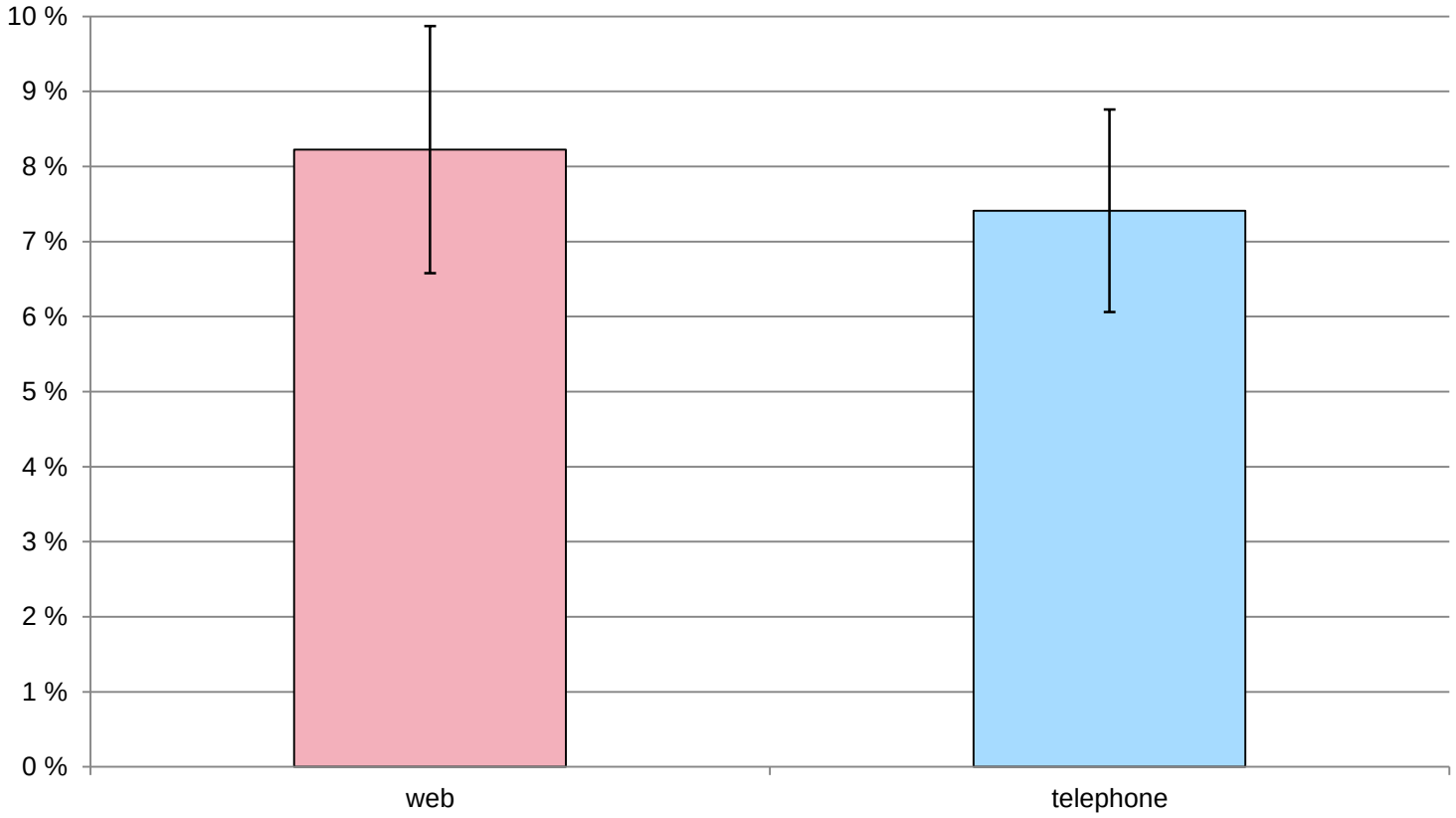
- Redesign of LFS to web or telephone leads to mode effects on employment. These are a mix of selection and measurement bias.
- At Statistics Netherlands, mode-specific selection bias on employment can be largely removed through administrative data.
- Mode effects vary over relevant population strata.

- BUT: Experiment does not allow for easy translation to mixed-mode method effects and to proxy interviewing without additional assumptions.
- For this purpose LFS parallel runs are used.

Mode effects study DE

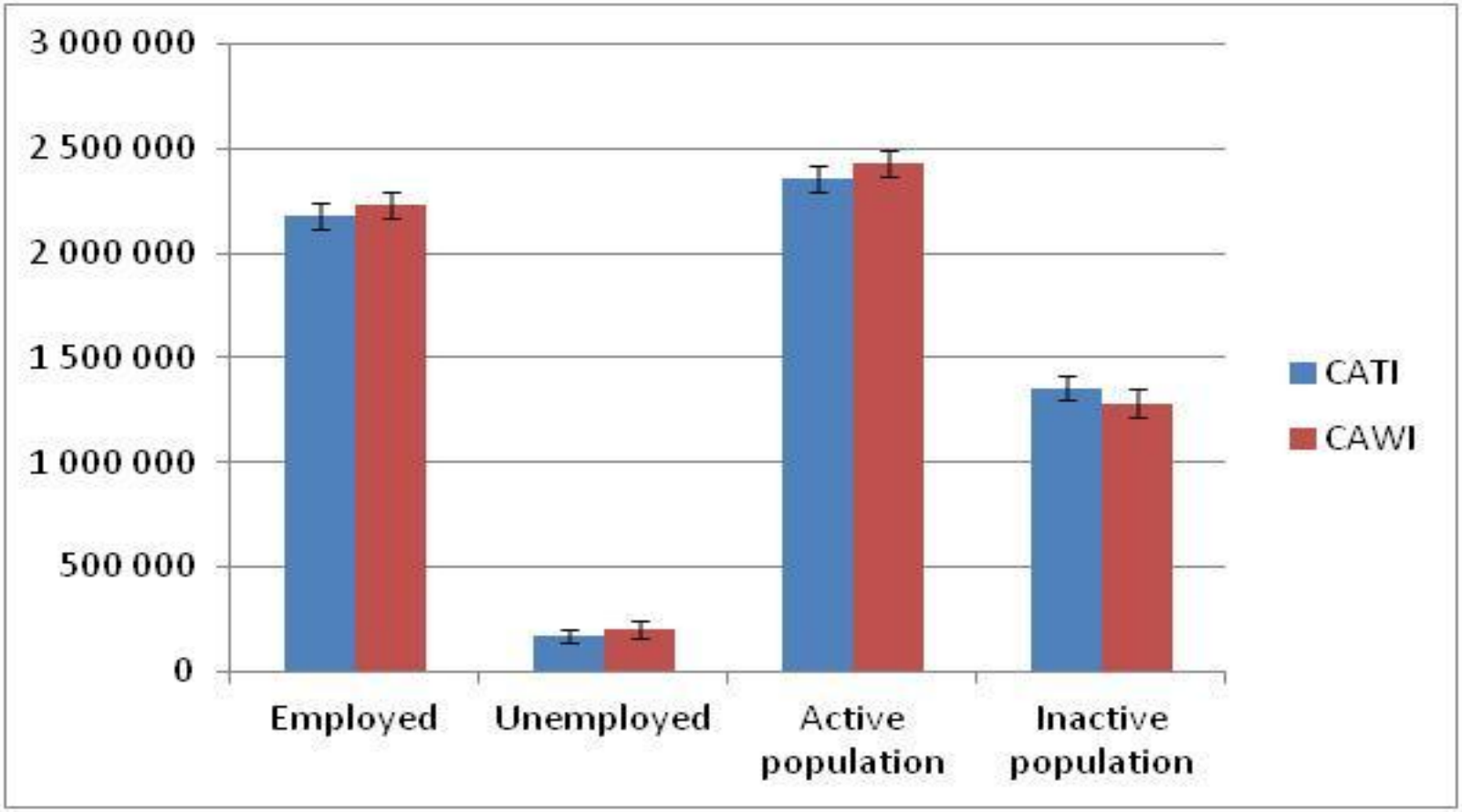
- Presentatie DE

The unemployment rate by web (pilot) and by telephone interview (LFS), and CI (95%), October 2013.



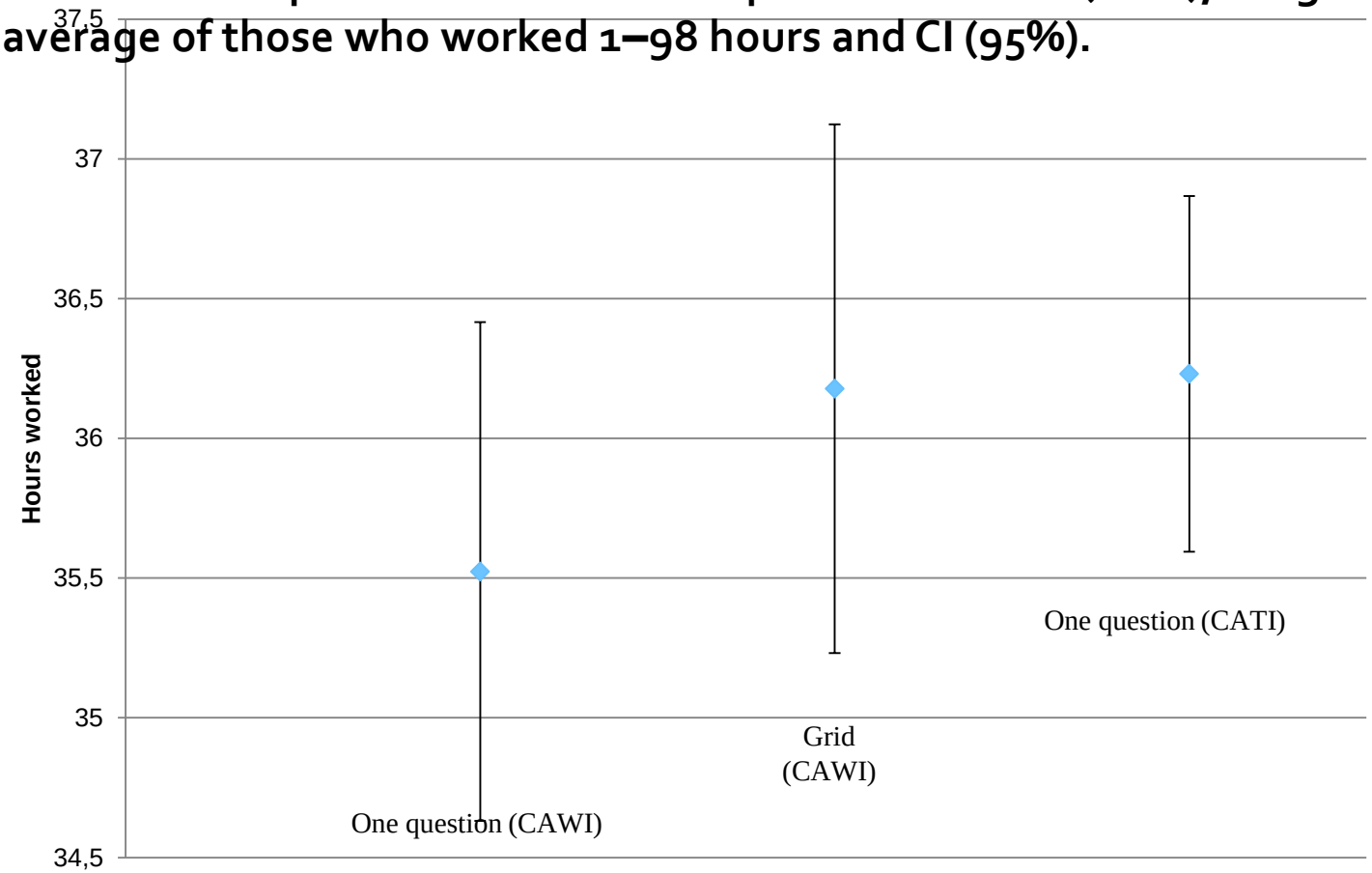
Statistics Finland

The main indicators of LFS: Estimates and CI (95 %) for employed, unemployed, active and inactive population in LFS data collected by CATI and CAWI, October 2013.



Statistics Finland

Hours worked during the reference week among web data collection (CAWI) to two different question forms and telephone interview (CATI), weighted average of those who worked 1–98 hours and CI (95%).



Unimode project

Thank you for your attention

More information:

– www.cros-portal.eu/content/data-collection-social-surveys-using-multiple-modes-dcss