

# UNECE Expert Meeting on Statistical Data Collection

Geneva, 22-24 May 2024

# Experience on multimode data collection in the NSI Spain: challenges and opportunities.

*Francisco Hernández  
INE, Spain*



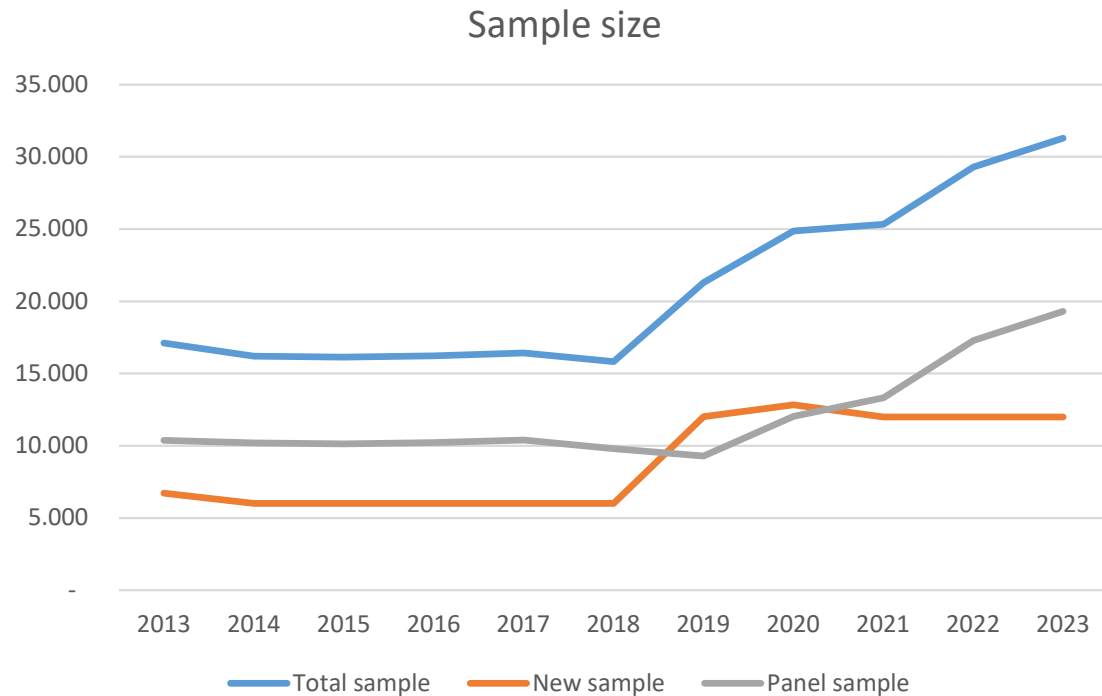
## EU-SILC: Objective and Regulation

- The main objective of the survey is to provide information on the income, level and composition of poverty and social exclusion in Spain and to allow comparisons with other European Union countries.
- Regulation (EU) No 2019/1700 establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples:
  - Combination of statistical and administrative information: National Tax Agency and Social Security.

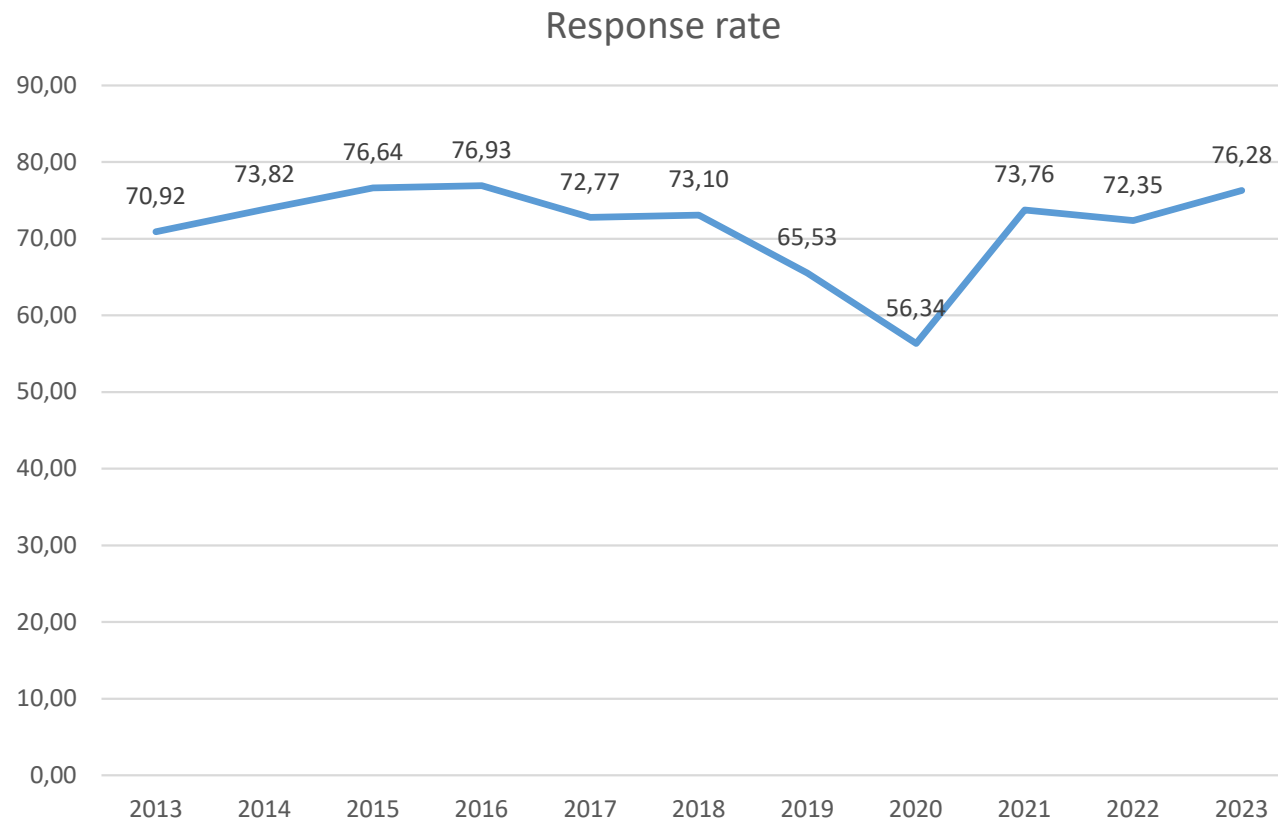


# EU-SILC: Sample design

- **Rotating panel design:**
  - **Two-stage sample design** with stratification of the first stage units.
  
- **Sample size**
  - 2018: 15.824 households.
  - 2023: 31.297 households.



# EU-SILC: Response rate Effective sample/Total sample



# EU-SILC: Review of Data collection

## ➤ Data collection mode:

- Until 2020: CAPI
  - 2017: Pilot test of multimode data collection:
    - Carried out by outsourcing.
    - Non-satisfactory results.
- 2020: Pandemic → CATI
- Since 2021: Multimode data collection: CAWI-CATI-CAPI




## ➤ Outsourcing of data collection:

- Price adjusted by data collection mode.
- Less possibilities for controlling interviewers' work.
- Change of data collection company submitted to administrative criteria.

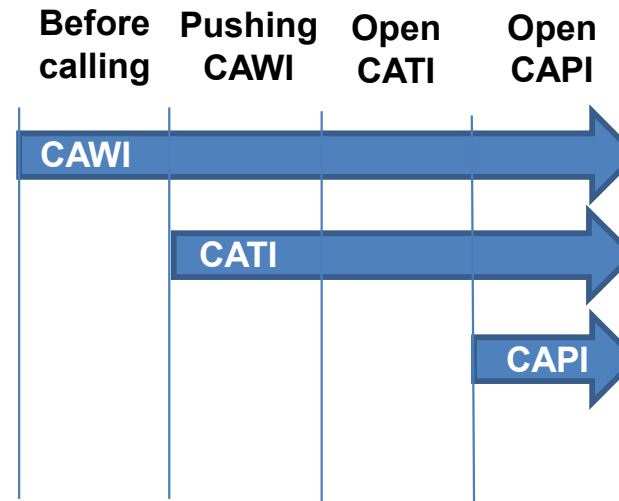


## Data collection platform: IRIA

- The data collection plan is conditioned by the platform to be used.
- NSI Spain has its own platform: The logo for IRIA (Instituto de Estadística de España) features the word "iria" in a bold, lowercase, black sans-serif font. To the right of the text is a stylized graphic consisting of a red left square bracket and a red right square bracket. Inside the space between these brackets are three vertical bars of different colors: a green bar on top, an orange bar in the middle, and a blue bar on the bottom. The top two bars have small red dots above them.
  - All collection modes are integrated in a flexible way.
  - Storing the information in centralised data bases.
    - When outsourcing the data collection, NSI Spain provides the platform, including communications.
  - IRIA: One platform for economic and household surveys
    - Good practice: Staff in the data collection unit deeply knows the tool and they usually programme the questionnaire.



## Should we close the other modes when opening a new mode? When should we add a new mode? How much can costs be reduced?



### ➤ Consecutive modes:

- Continuous process in which introducing a new mode doesn't entail closing previous modes.
- Each mode has its own protocol to make sure that every household has been properly worked in that mode.
- The main goal is to reduce costs:
  - For example, minimizing the use of CAPI.



## Relevance of the consecutive system

### Should we close the other modes when opening a new one?

- Good practice: Pushing the CAWI response via telephone:

|      | Before calling (t) | Pushing CAWI (t+7) | Open CATI (t+15) | Open CAPI (t+42) | TOTAL |
|------|--------------------|--------------------|------------------|------------------|-------|
| CAWI | 16%                | 33%                | 48%              | 3%               | 100%  |
| CATI | 0%                 | 22%                | 64%              | 14%              | 100%  |
| CAPI | 0%                 | 0%                 | 0%               | 100%             | 100%  |

- The EU-SILC sample is divided in 3 similar blocks, each of them entering the data collection with a 1-week delay over the previous one.
- Some remarks:
  - Closing CAWI when opening CATI => 50% less CAWI.
  - Not allowing CATI responses when pushing CAWI => 22% less CATI.
  - Closing other modes when opening CAPI => 3% less CAWI and 14% less CATI.

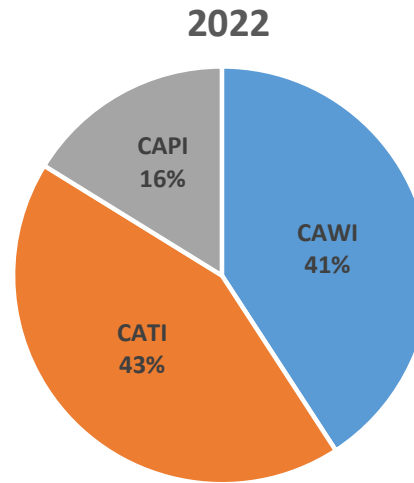
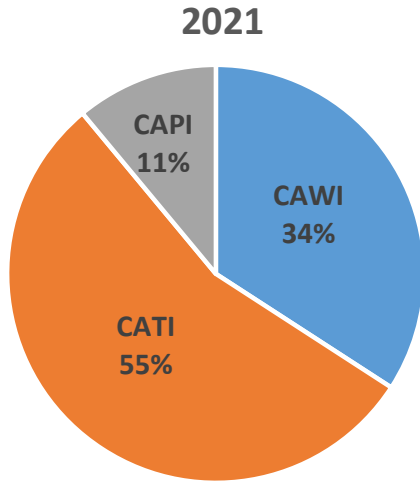




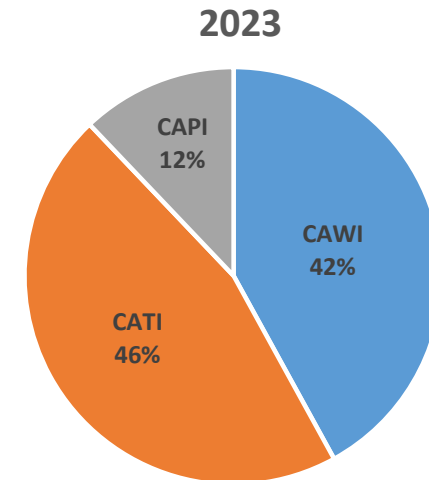
# EU-SILC: Response rate

Effective sample/Total sample

## How much can costs be reduced?



| Year 2023 | Units answering |
|-----------|-----------------|
| CAWI      | 10.023          |
| CATI      | 10.977          |
| CAPI      | 2.864           |



## EU-SILC: Longitudinal analysis of response modes

| 2021 | 2022  | 2023  |
|------|-------|-------|
| CAWI | 71,5% | 82,4% |
| CATI | 63,4% | 81,9% |
| CAPI | 29,9% | 47,3% |

- Households staying in the sample from 2021 to 2023 and answering:
  - Households have a certain degree of adherence to a specific data collection mode.
  - Data for CAPI are not as eloquent because it is residual.



# Same questionnaire for all data collection modes?

- Flexibility:
  - Same questionnaire in different modes?
  
- Two different situations:
  - EU-SILC and ICT usage in households:
    - Same questionnaire, with minor linguistic adjustments.
  
  - HBS:
    - Different questionnaires depending on the COICOP knowledge:



- CATI: Interviewers know the classification structure.

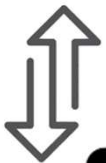
- CAWI: Respondents are not familiar with the classification.



## CATI

# When should we add a new mode?

- IRIA has its own algorithm for calling the respondents:
  - Data collection staff knew that algorithm and this allows to make the data collection more efficient by adjusting some parameters.
- Protocol: Time period and Number of calls (new or panel sample: 16 or 12).
  - Total number of calls made: 253.058 calls.
  - Total questionnaires: 21.000 questionnaires.
- Recording of the telephone calls: Chance for a quick and very efficient monitoring for taking measures related to quality.
  - Automatic dialing was not used without an interviewer available and ready.



# Availability of telephone numbers of respondents

- What if telephone numbers of **sample units** are not available for the institution?
  - Different situations: Rotating panel vs One-time survey



- Situation of INE Spain:
  - Agreement: Telephone numbers were available for the sample units (near by 90% of sampled households).
    - Not available for all the units in the sampling frame.
    - Land lines and mobile phones.
  - E-mail addresses are not available for household surveys:
    - In some cases, they are requested during the first collaboration.
- Good practice: Use of sms for claimings.



## Face-to-face stage: CAPI

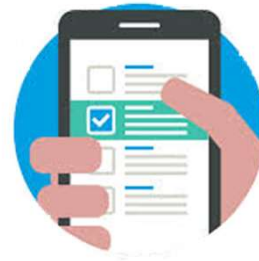
### Should we close the other modes when opening a new mode?

- IRIA includes CAPI data collection.
- The interviewers used tablets 10.1.
- Take advantage of the better connectivity in CAPI data collection:
  - Work directly connected to the centralised data bases.
  - Some problems in rural areas.
- Monitoring of the interviewers.
- Use of telephone for setting appointments and completing information.



## Next steps

- Smartphones should be made available as a usable device in household surveys.



- CAWI: Editing based on paradata, e.g. questionnaire completing time, forward and backward...



- Paradata: Used for strategic planning to more efficiently mix the different modes – Use of AI.



# Thank you!!



[francisco.hernandez.jimenez@ine.es](mailto:francisco.hernandez.jimenez@ine.es)





## Non-sampling errors

- Eurostat quality reports:

<https://ec.europa.eu/eurostat/web/income-and-living-conditions/quality>

- EU-SILC methodology in NSI Spain's web site:

[https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica\\_C&cid=1254736176807&menu=metodologia&idp=1254735976608](https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176807&menu=metodologia&idp=1254735976608)

