



An
Phríomh-Oifig
Staidrimh

Central
Statistics
Office

BSTN Non-probability samples

CSO's planned test of using a non-probability survey

Motivation

- Declining response rates, high cost of data collection, increasing demand, etc.
- Growth in availability of non-probability sources such as administrative data, big data, social media etc.
- Web panel surveys easy and cheap to carry out and usually have fast turnaround
- Availability of a 'quasi' panel of over 12,000 email addresses from a series of online opt-in surveys conducted by CSO over past two years



But ...

- Can non-probability surveys be used in conjunction with other sources to reduce some of the burden of traditional surveys while maintaining a valid statistical inference framework and with an acceptable level of quality?
- There is a growing amount of literature on this area and there are methods to adjust for the bias of a non-probability sample.



Exploring the potential of non-probability samples in the context of official statistics

- CSO does not currently have expertise in this area but are interested in developing it as we can see that it is a growing area
- Potential for collaboration with other NSIs and/or Eurostat to develop a framework or a set of appropriate practices
- Group already set up and has met in September



Next Steps

- CSO comparison using ICT survey
- Sharing experiences such as different approaches:
 - a design-based approach (sample matching)
 - a model-based approach (hierarchical models) and
 - perhaps pseudo-weights also.
- Sharing best practice/potential projects:
 - ISTAT are looking at some issues with respect to coverage of administrative sources
 - Stats NZ considering using some of these methods for business surveys
 - CBS has a lot of administrative sources that require adjustments for coverage etc.



Test of the ICT

- The CSO conducts an annual survey on the ICT activities of households (EU requirements under Regulation (EC) No 808/2004). This survey will be conducted as usual in Q1 and Q2 next year
- In parallel ... a probability sample (target sample) will be selected from our population frame and 'sample matching' will be applied to find a match for every individual in the probability sample from the PULSE panel with respect to a set of auxiliary variables that are available on both data sources



Test of the ICT

- Although the matched sample will look like the target sample, it is important to note that the matched sample is still *not* a probability sample as it can only match the target with respect to the available auxiliary information but not with respect to the vast number of unmeasured characteristics about these individuals.
- This can be especially problematic if some of these unmeasured characteristics are related to the target variables of the survey. However, web panel surveys are easy and cheap to carry out and the availability of the contact details of over 12,000 individuals provides an opportunity for us to test some of the methods available to adjust non-probability surveys.



Test of the ICT

- Data collection for the test will be via CAWI (email invitations) and will happen in parallel with the main survey
- Results from the two surveys will be analysed in Q3 2023
- Limitations:
 - Conducting an ICT survey via CAWI
 - Difference in coverage w.r.t. age – ICT coverage should be 16-74 years whereas PULSE respondents are 18 and over





An
Phríomh-Oifig
Staidrimh

Central
Statistics
Office

Thank you