The prevalence and potential bias of online surveys

Zita Fellner[1], Blanka Szeitl[1]

Survey Methods Room Budapest, Department of Statistics, Faculty of Social Sciences, University of Eötvös Loránd, Hungary[1]

Workshop "Survey Climate and Trust in Scientific Surveys – Recent Developments and Controversial Issues"
October 4-5, 2022

Outline

1. Motivation & backgrounds
2. Research methods
3. Results
   - Prevalence of online surveys
   - Potential bias of online surveys
4. Conclusion

Outline

Motivation

- the classic face-to-face data collection method has faced **difficulties:** decreasing of response rates, rising costs, and a hard-to-control interviewer network;
- changing **technological environment** of survey data collection: increase in internet penetration, smartphone usage, and availability of survey programs;
- an **increasing amount** of online data collection, the **quality** of which is questionable;
- relevant differences at **country level**;

Motivation

- **case study** of Hungary;
- how the **prevalence and quality** of online surveys can be measured?
- how potential **bias** of online surveys can be estimated?
- taking into account **population patterns** (administrative data, face-to-face survey data);
- **why this is important? - perception & trust**

Backgrounds

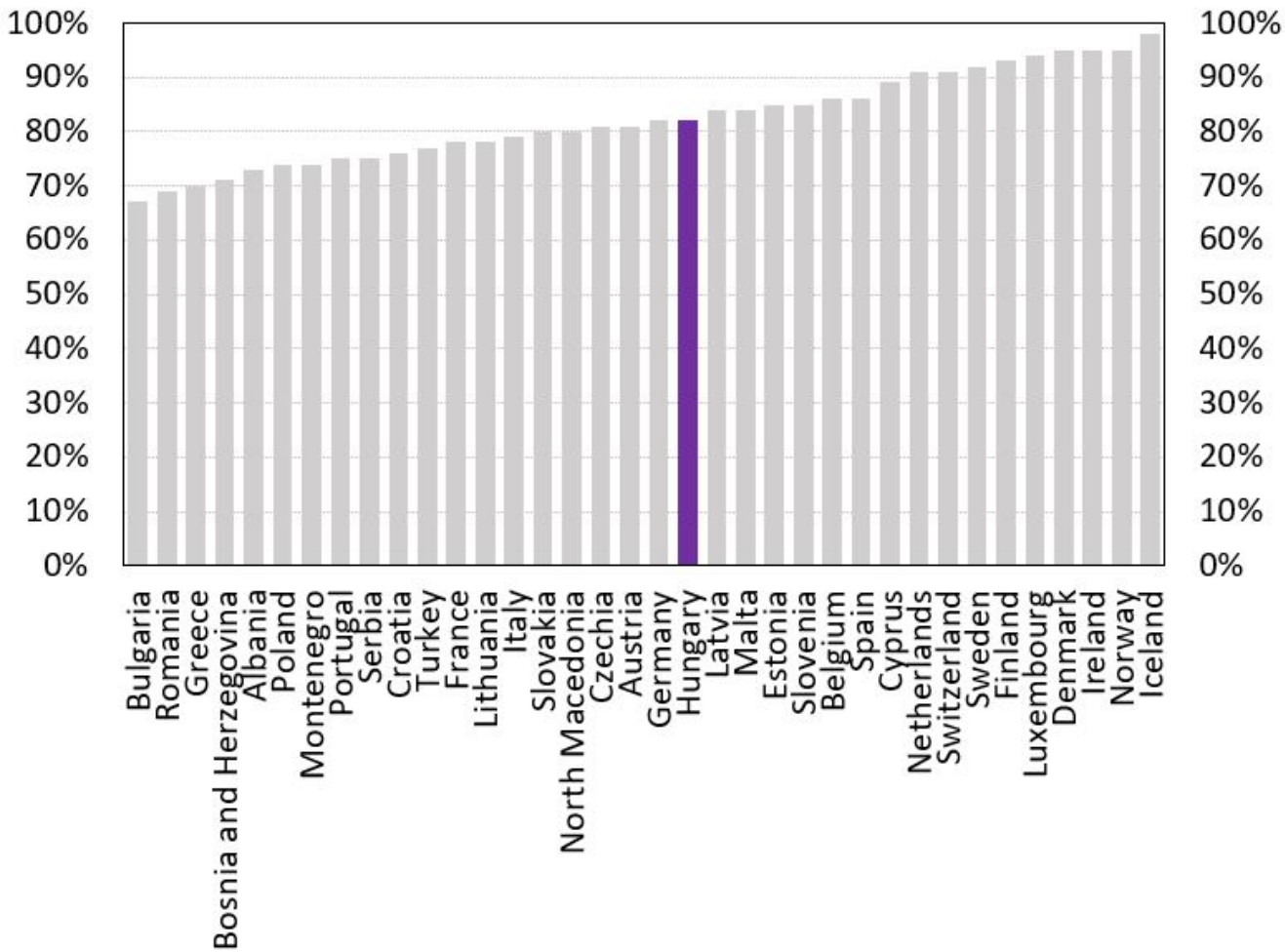

Figure: Share of daily internet users, 2021 (Source: Eurostat, %)

Backgrounds

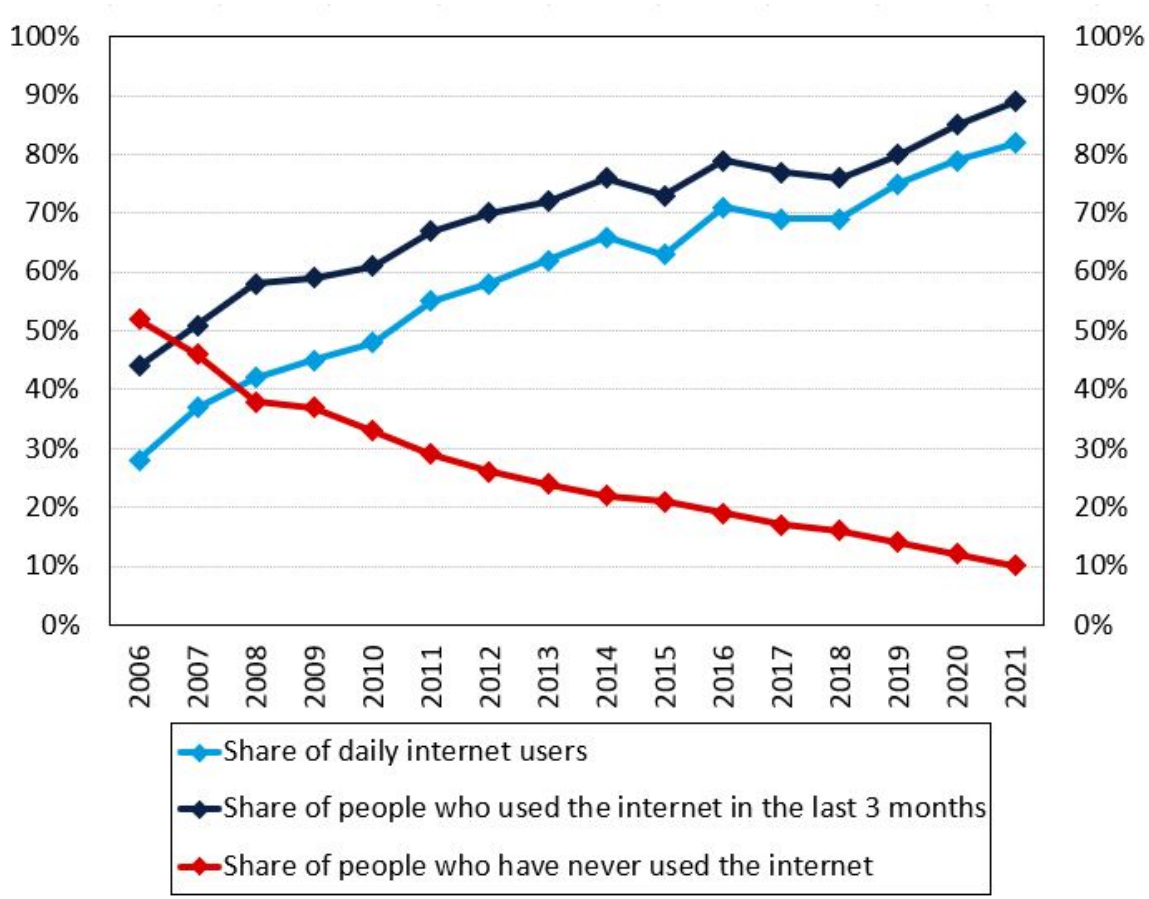

Figure: The frequency of internet use in Hungary between 2006-2021 (Source: Eurostat, %)

Backgrounds

		Total population	Population with internet access	Population of daily internet users
Age	16-24	13%	15%	16%
	25-34	16%	18%	19%
	35-44	19%	21%	22%
	45-54	20%	21%	21%
	55-64	16%	14%	13%
	65-74	16%	11%	9%
Education	Primary	19%	16%	15%
	Secondary	56%	56%	55%
	Higher	25%	27%	29%

Figure: Total population, internet users and daily internet users by age, and education in Hungary (Source: Central Statistical Office of Hungary, 2021, %)

Backgrounds

- the internet penetration increases, and as wider and wider groups of people become involved can be **the target of online surveys**;
- online surveys conducted in different periods cannot be easily compared because of the **continuous change in the population base**;
- there are relevant differences between online populations and total population **along main demographic characteristics**.

Outline

Research methods

Prevalence with content analysis:

- simple random sample of online articles (n=4,000);
- sample of scientific papers (n=150);
- assessing these cases based on previously defined methodological criteria: sampling method, mode of data collection, stratification, interpretation.

Research methods

Potential bias with simulation:

- modelling different populations: total population; population of those with internet access and population of daily internet users;
- modelling population parameters in all the 3 populations;
- modelling the estimates from samples of these populations.

Supplemented with: comparisons based on the European Social Survey (ESS).

Outline

Prevalence of online surveys

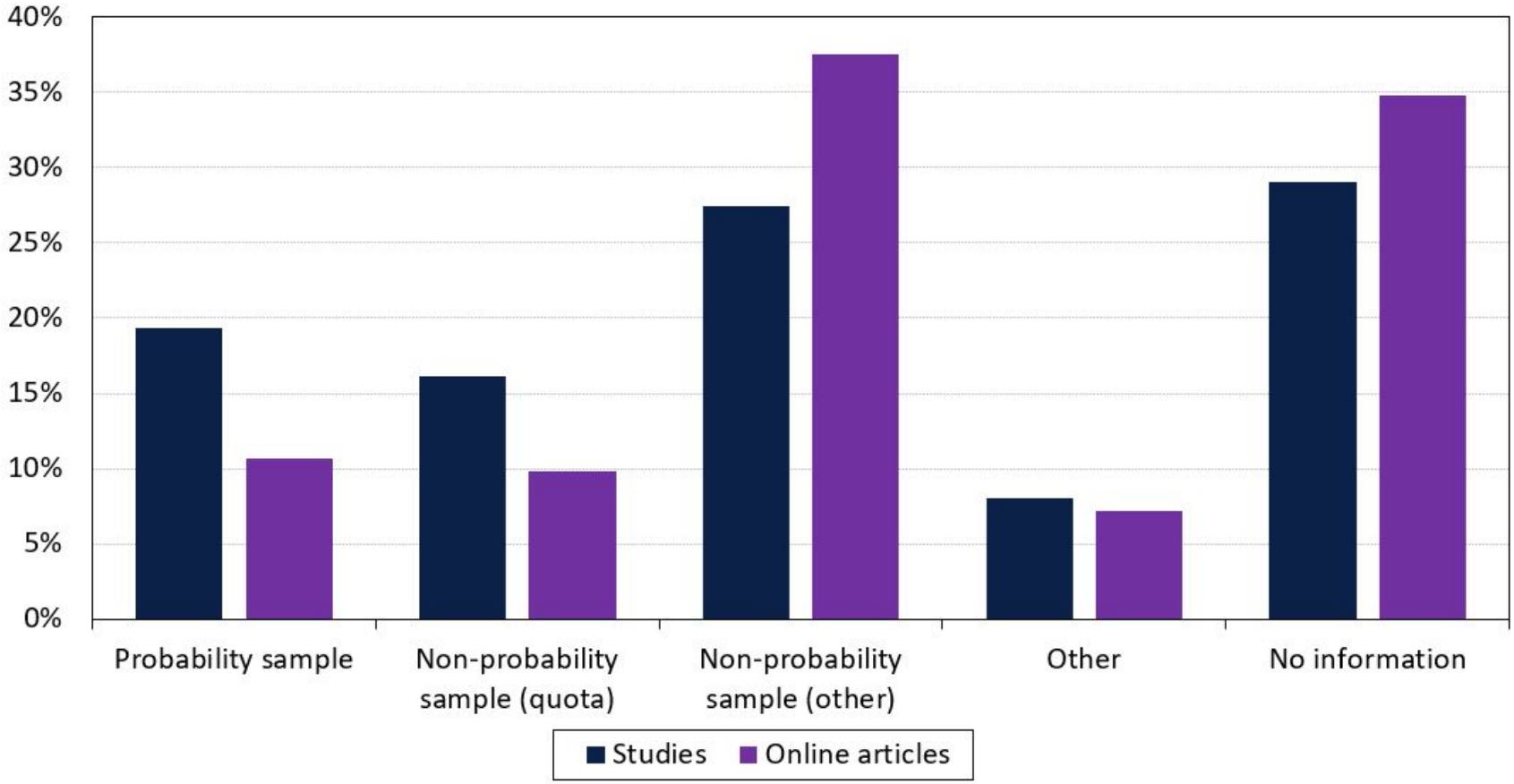

Figure: Studies and online articles analysed by sampling method

Prevalence of online surveys

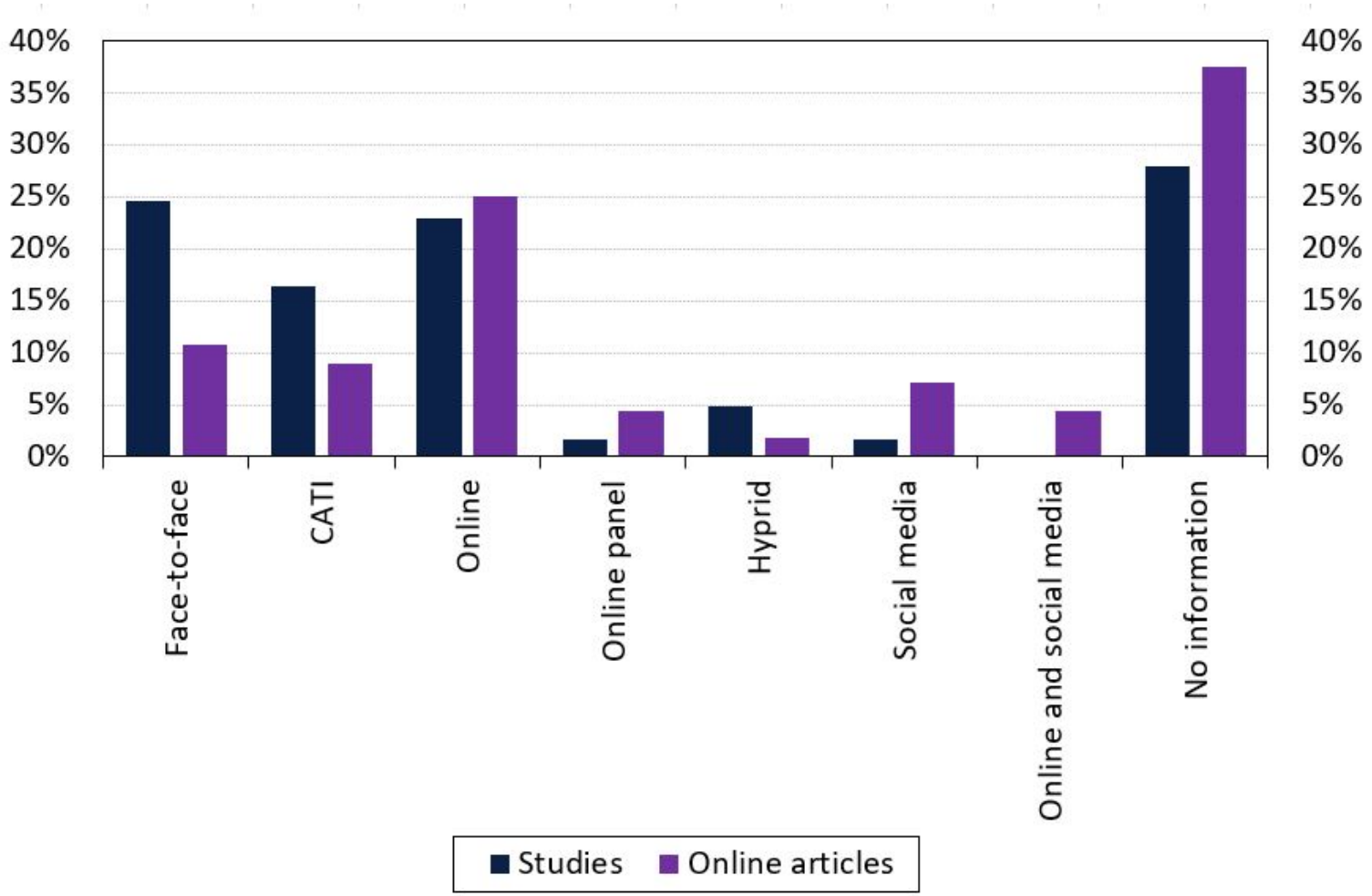

Figure: Studies and online articles by method of data collection

Potential bias of online surveys

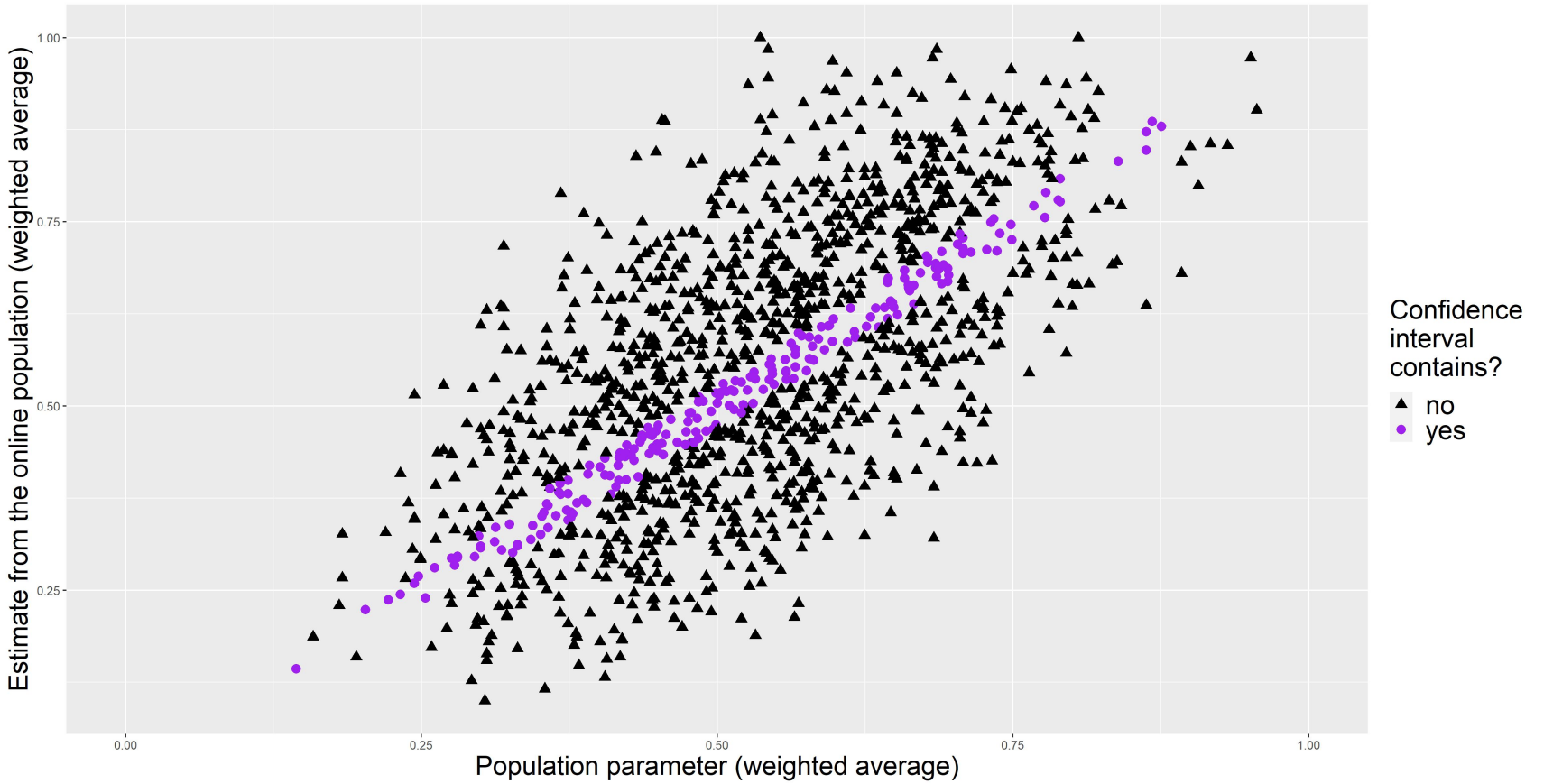

Figure: Difference between the distribution estimated from the online sample and the population distribution

Potential bias of online surveys

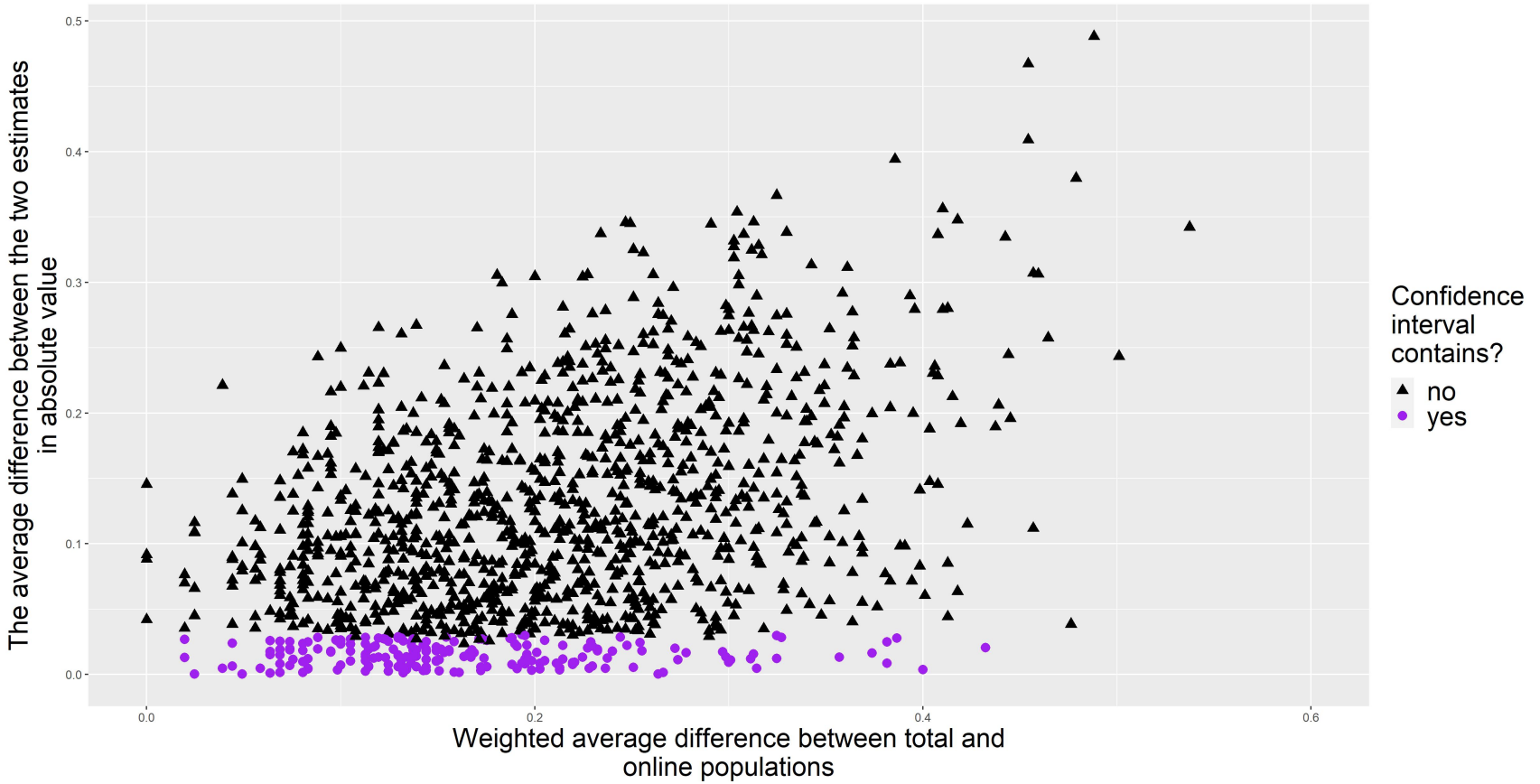

Figure: Difference between the distribution estimated from the online sample and the population distribution based on the total population parameter value

Potential bias of online surveys

Real estimates (based on the European Social Survey):

- political attitude: when considering online pop., we significantly underestimate the proportion of those who do not feel close to any party (0.55 compared to 0.61 estimated on the basis of the full sample);
- health status: we would find a significantly better health status based on an internet sample (0.3 for those reporting very good health compared to a rate of 0.17 for the full sample).

Outline

Conclusion

- a phenomenon of **the rise of online data collection**;
- an online survey cannot be considered a random sample, even with an apparently high internet coverage;
- estimates calculated from such data **may be biased** even if they are representative of the total population in certain demographic aspects;
- there is a **big share** of online surveys in scientific papers and in online articles;
- **results are often incorrectly interpreted as random samples**;
- the results are not generalisable to other countries, our aim is to draw attention to **the need of start local studies**.

Thank you!

Blanka Szeitl
Assistant Lecturer
Eötvös Loránd University
Faculty of Social Sciences
Department of Statistics

phone: +36 30 963 3164
e-mail: szeitl.blanka@tatk.elte.hu

Zita Fellner
Assistant Lecturer
Eötvös Loránd University
Faculty of Social Sciences
Department of Statistics

phone: +36 30 652 2755
e-mail: fellner.zita@tatk.elte.hu