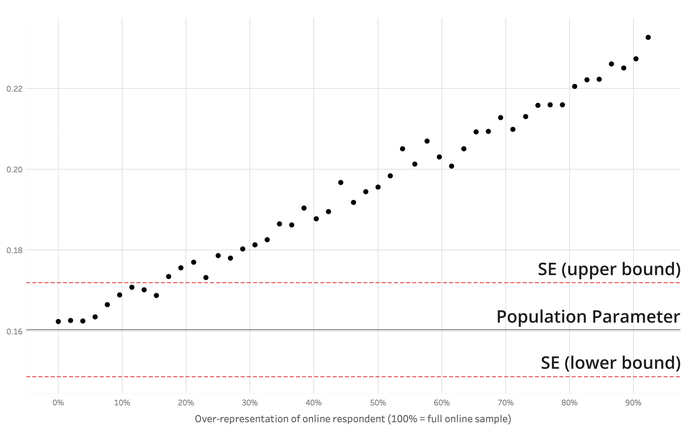
Mixing survey models in empirical social research

Blanka Szeitl, Benedek Sarnyai, Eszter Katona,
Anna Kovács, Bernadett Nagy, Nhi Nguyen Thi Phuong, Rebeka Tóth

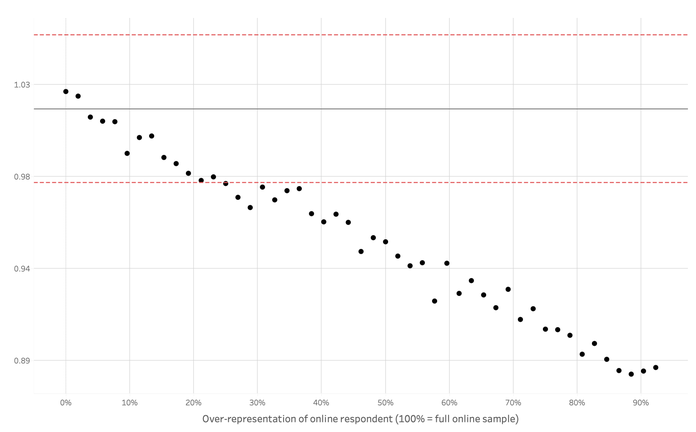
ELTE Survey Methods Room Budapest
Prepared in the "National Laboratory for Social Innovation" project (RRF-2.3.1-21-2022-00013), within the framework of and with the support of the Recovery and Resilience of the European Union

We present how estimations based on online surveys may be biased and how mixing online and offline (CAPI) survey modes may improve the results. A quasi-population was built based on 29 nationnally representative face-to-face (CAPI) surveys from Hungary between 2016 and 2022 with a total quasy-population size of 30,700 individuals and their characteristics. We simulated probability-based sampling procedures from this quasi-population with changing proportions of daily internet users: it started with samples with 100% daily internet users and was gradually reduced by 1% until the internet penetration rate in the quasi-population (54%). Each sample contained 1,000 individuals selected with stratification based on gender, age, type of settlement and level of education. 6 variables are selected to present how overrepresentation of daily internet users (x-axis) result in biased estimates (y-axis) and what is the minimum proportion of offline respondents which may improve the results.

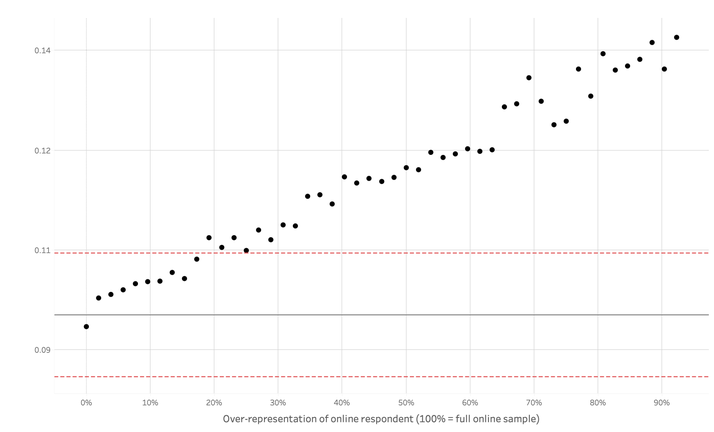
Level of education (university degree, proportion)
When estimating the proportion of people with university degree starting from 18% of over-representation level of daily internet users (64%) result in biased estimate. A full online sample greatly overestimates the proportion of highly educated groups.



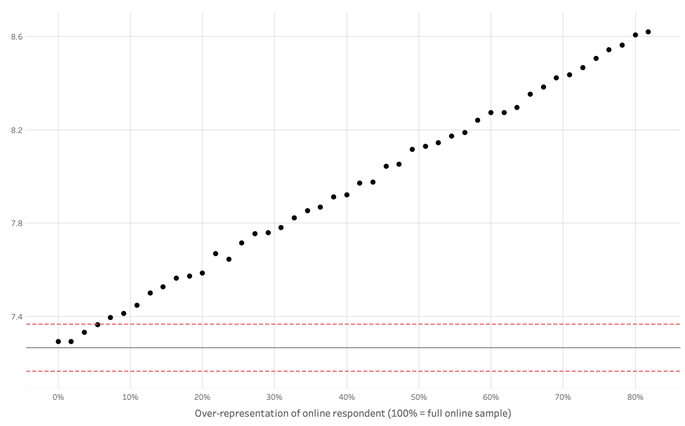
Number of children in the household (average)
Starting from 29,6% percent of over-representation level of daily internet users (70%) result in biased estimate. A full online sample greatly underestimates the average number of children in the household (a



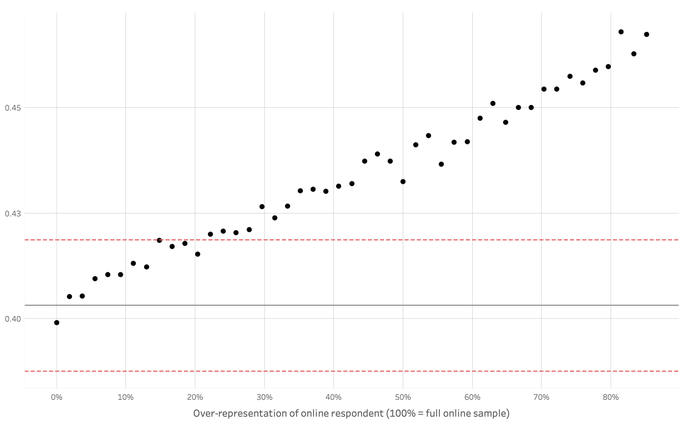
Subjective health assessment („very good", proportion)
In the case of subjective health, starting from 18% of over-representation level of daily internet users (62%) result in biased estimate. A full online sample moderately overestimates the proportion of those having „very good" subjective health.



Household supplied with consumer goods (average, proportion)
In the case of consumer goods, starting from 11% of over-representation level of daily internet users (59%) result in biased estimate. A full online sample severely overestimates the number of consumer goods in the household.



Subjective *current* financial situation of the household („very good", proportion)
In the case of subjective current financial situation, starting from 22% of over-representation level of daily internet users (66%) result in biased estimate. A full online sample moderately overestimates the proportion of those having „very good" subjective financial situation.



Subjective *future* financial situation of the household („positive", proportion)
When estimating "positive" expectations on subjective future financial situation the over-representation level of daily internet users is still 66%, and a full online sample overestimates the proportion of those having "positive" expectations regarding future financial situation of the household.