

WEIGHTING OF A WEB PANEL SURVEY CARRIED OUT IN THE PANDEMIC SCENARIO

Marcelo Trindade Pitta

Pedro Luis do Nascimento Silva

NIC.br Annual workshop on survey methodology | 7 October 2020



Organização das Nações Unidas para a Educação, a Ciência e a Cultura

cetic.br **nic.br** **egi.br**

Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação sob os auspícios da UNESCO

Núcleo de Informação e Coordenação do Ponto BR

Comitê Gestor da Internet no Brasil



INTRODUCTION

THE COVID-19 SCENARIO FOR CARRYING OUT FACE-TO-FACE SURVEYS



The COVID-19 pandemic made the traditional face-to-face interviews impossible for most surveys.

Alternative data collection methods had to be developed for population and other surveys.

Incomplete or partial frames have emerged as a critical barrier.



TRADITIONAL SURVEY

ARE FACING
INCREASING
CHALLENGES



Traditional surveys for public/official statistics:

- › Rely on probability sampling from complete frames;
- › Methods of data collection: CAPI, CATI and/or CAWI;
- › Increasing non-response rates; and
- › Increasing demand for more timely and disaggregated data.

While dealing with decreasing resources.

On the other hand....



ALTERNATIVE METHODS

NON- PROBABILITY SURVEYS



Non-probability surveys:

- › Cheaper to collect;
- › Larger samples can be collected in less time;
- › May introduce some types of bias.

Approach to overcome the difficulties:

- › Probability sample survey as a reference to weight the non-probability sample;
- › Mitigate the bias of estimates: model-dependent which adds complexity to the survey estimation.



COLLECTING DATA FROM A WEB PANEL

NON-
PROBABILITY
SAMPLE



ICT COVID-19 Web Panel Survey

Target population: Internet users aged 16+ in Brazil.

Domains: sex (2), level of education (3), region (5), age group (5) and socioeconomic status - SES (4).

Frame:

- › Web panel of individuals maintained by market research companies.
- › Telephone lists to reach a broader population (least favored social classes and less educated).

COLLECTING DATA FROM A WEB PANEL

NON- PROBABILITY SAMPLE



ICT COVID-19 Web Panel Survey

- › **Sample design:** quota sample established based on the following variables: sex, region, socioeconomic status (SES), age group and level of education.

Mode	Contacts tried	Respondents
Web	18,734	2,539
Telephone	72,794	88
Total	91,528	2,627



COLLECTING DATA FROM A WEB PANEL

PROBABILITY
SAMPLE AS A
REFERENCE



ICT Households Survey 2019: Reference survey

- › **Target population:** Brazilian households and all individuals 10+ years old residing in permanent private Brazilian households.
- › **Frame:** IBGE 2010 census tracts database.
- › **Total sample size:** some 30,000 interviews (households and individuals).
- › **Sample design:** stratified multi-stage sampling of households and residents.

A dark, semi-transparent image of a person's profile looking at a smartphone, serving as a background for the title.

METHODS FOR WEIGHTING

REQUIREMENTS



Common support

Using the probability sample to weight the web panel survey requires the two samples to:

- › Be carried out for the same reference period;
- › Collect a range of variables in the same way (overlapping questions);
- › Address the same target population;
- › Have no intersection of respondents.

METHODS FOR WEIGHTING

REQUIREMENTS



Common support

Using the probability sample to weight the web panel survey requires the two samples to:

- ? › Be carried out for the same reference period;
- OK › Collect a range of variables in the same way (overlapping questions);
- ? › Address the same target population;
- ? › Have no intersection of respondents.

The probability sample must represent the whole target population, being adjusted for non-response and expanding to known population totals (Valliant 2019).

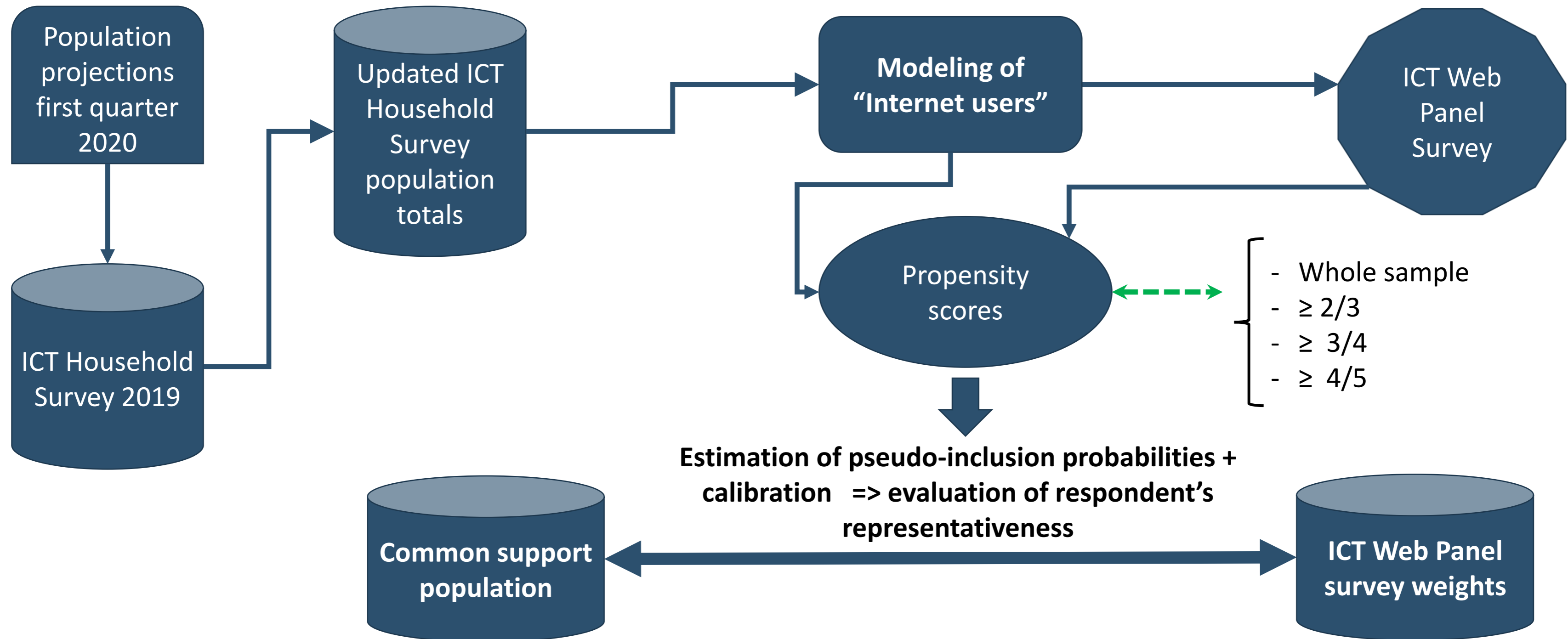
METHODS

APPROACH USED

- › Estimate the size of the target population – Internet users 16+ years old – using data collected by 2019 ICT Households Survey.
- › Estimate pseudo-inclusion probabilities for the non-probability sample units via logistic regression model and use their reciprocals as weights.
- › Evaluate and identify the population represented by respondents of the web panel survey, among those in the target population.
- › Estimate variance using bootstrap.

METHODS

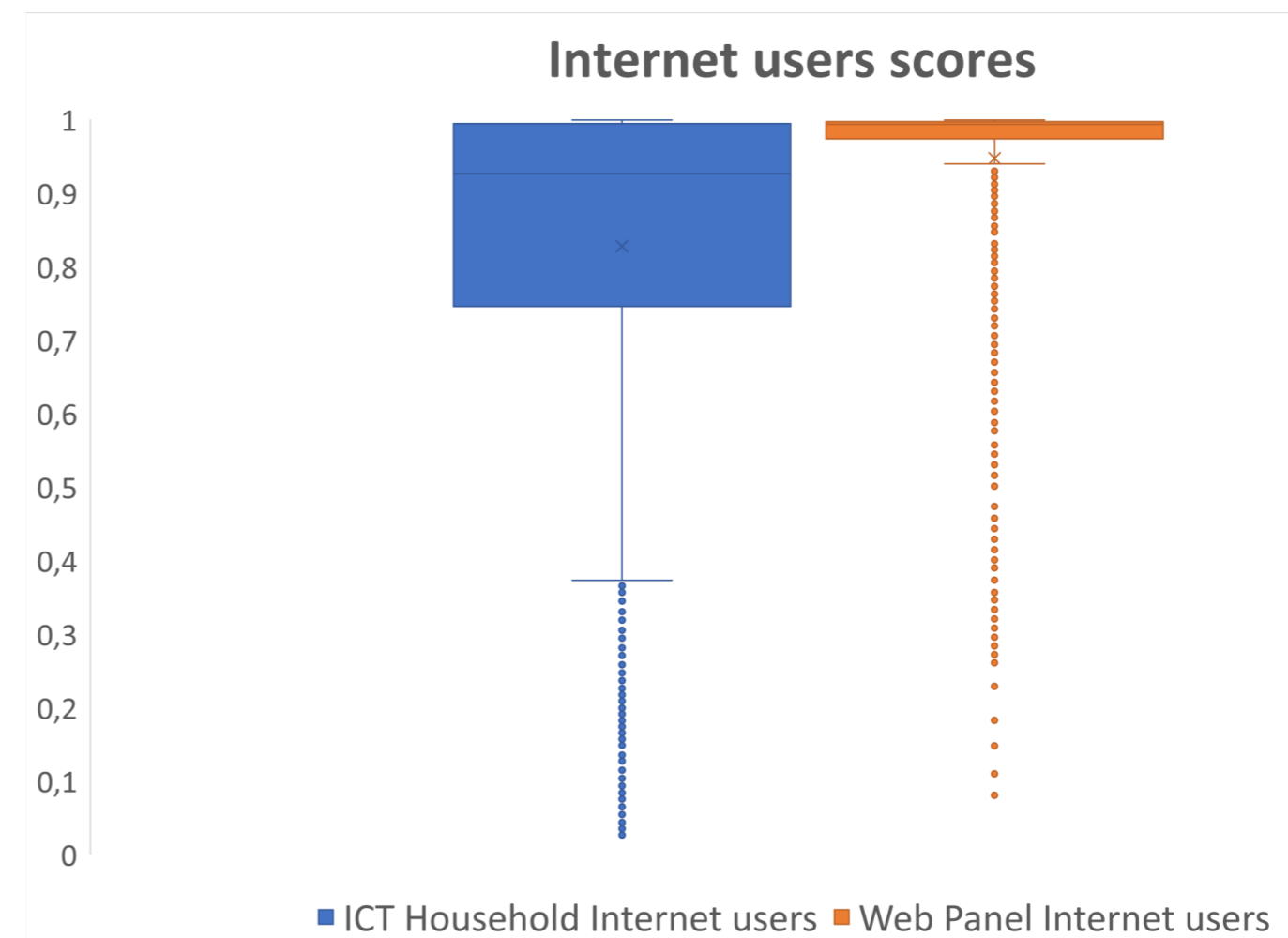
APPROACH USED



RESULTS

ICT COVID-19 WEB PANEL

- › Internet users model statistics:
 $R^2 = 0.431$; Correctly classified individuals = 83%.
- › Comparison of scores between ICT Household Survey and Panel Survey.



RESULTS

ICT COVID-19 WEB PANEL

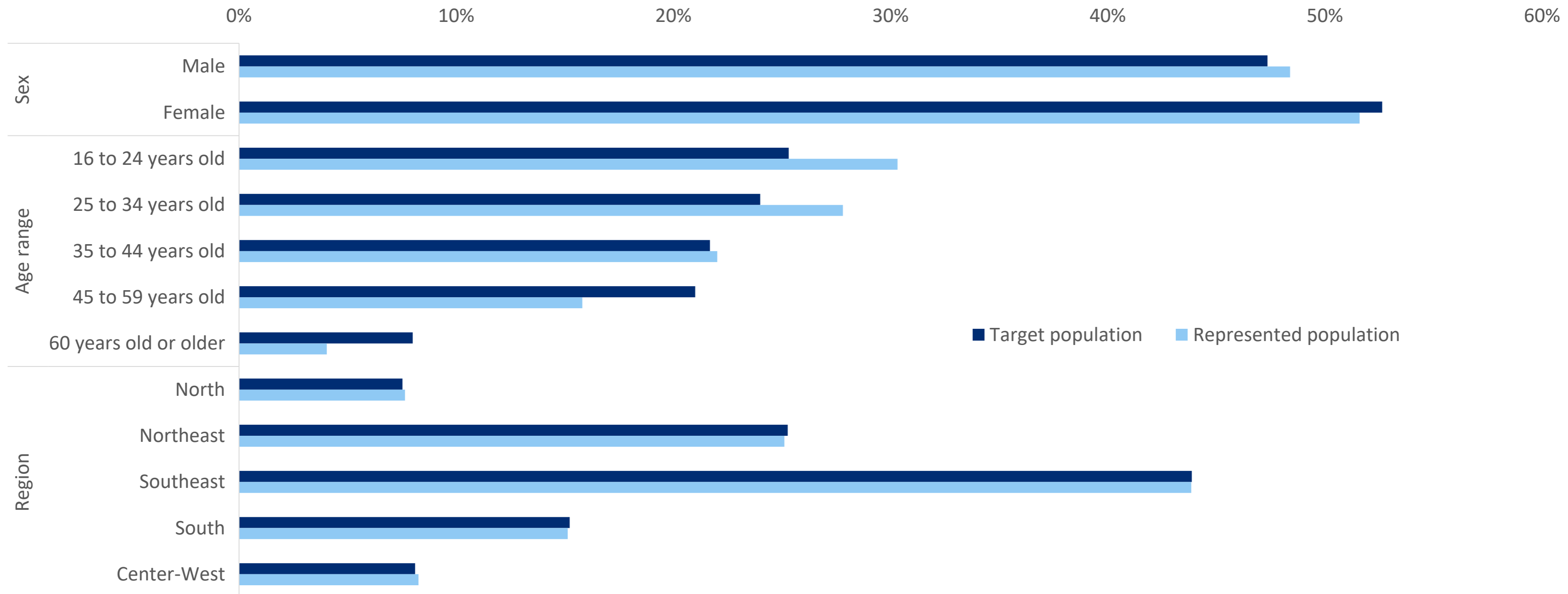
- › Estimation and evaluation of pseudo-inclusion probabilities representativeness

Probability sample population used for calibration

<i>Web Panel weight calibration factor statistics</i>	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Whole panel sample	0.120	0.283	0.369	0.581	0.630	5.840
Sample of Internet users with scores from 2/3 and up	0.078	0.484	0.637	0.988	1.038	12.518
Sample of Internet users with scores from 3/4 and up	0.063	0.623	0.767	0.918	0.930	4.380
Sample of Internet users with scores from 4/5 and up	0.087	0.447	0.554	0.729	0.791	5.604

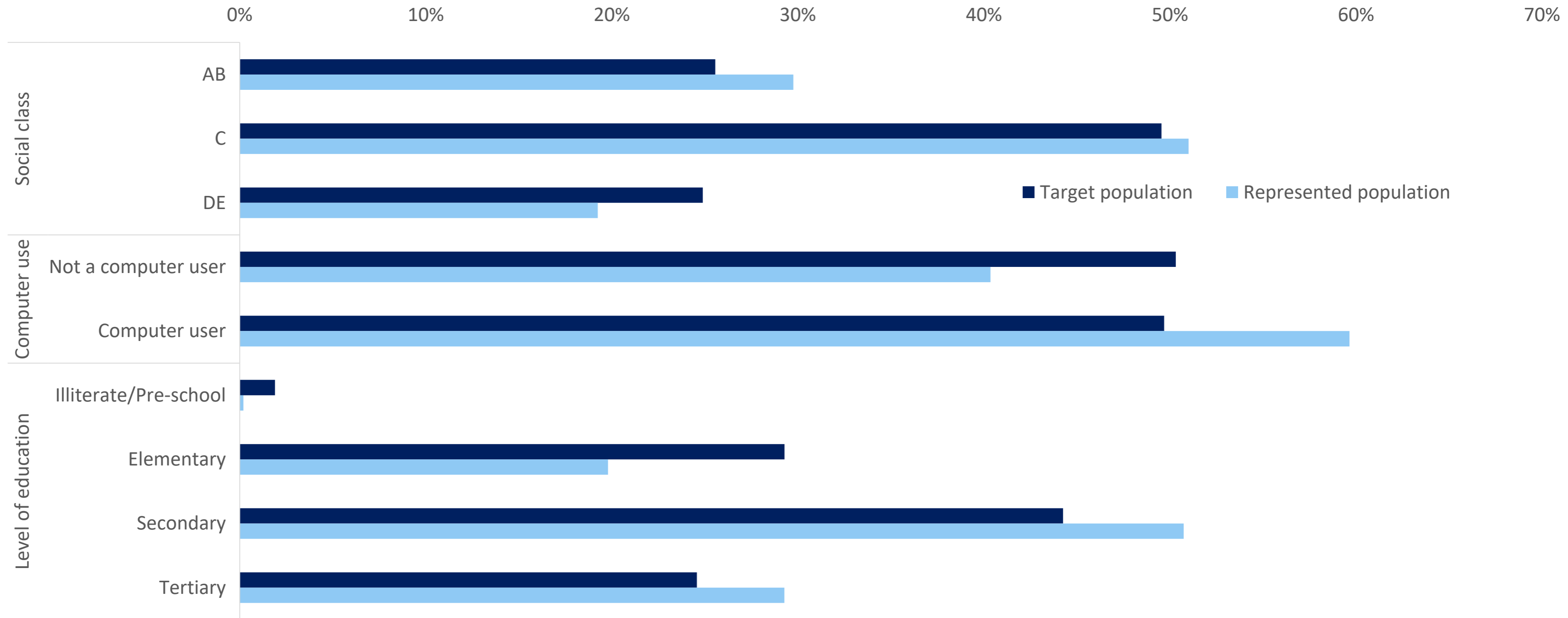
RESULTS: Represented population (101M) and target population (121M)

ICT COVID-19 WEB PANEL



RESULTS: Represented population (101M) and target population (121M)

ICT COVID-19 WEB PANEL



RESULTS: EVALUATING PROS & CONS

ICT COVID-19 WEB PANEL

Advantages

- › Data were collected avoiding face-to-face interviews.
- › The whole survey, from planning to data release, took less than two months to complete.
- › Cost of data collection much cheaper than for a traditional face-to-face survey.

Disadvantages

- › Web panel frame recruitment is not meant to address a specific population.
- › Coverage issues remain, despite using a probability sample as reference.
- › Approach is model-dependent – good models may not always be available.
- › Explanation of methodology and its dissemination is complex.

CONCLUSIONS

ICT COVID-19 WEB PANEL

- › Use of non-probability sampling for surveys is increasing.
- › Use of reference probability sample survey to weight the non-probability sample is attempt to produce more reliable estimates.
- › Probability sample surveys will continue to be needed, if they are to serve as reference samples for weighting more frequent non-probability surveys.
- › Approaches to weight the non-probability samples are still being developed, much to be done.

Thank you all!

www.cetic.br
marcelopitta@nic.br
pedronsilva@gmail.com



Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

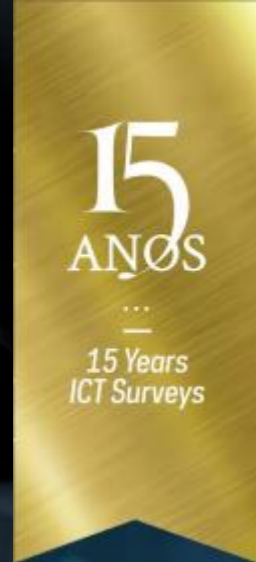
cetic.br **nic.br** **cgi.br**

Centro Regional de Estudos
para o Desenvolvimento da
Sociedade da Informação
sob os auspícios da UNESCO

Núcleo de Informação
e Coordenação do
Ponto BR

Comitê Gestor da
Internet no Brasil

References



- › Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. and Tourangeau, R. 2013. *Summary Report of the AAPOR Task Force on Non-Probability Sampling*. *Journal of Survey Statistics and Methodology* 1 (2): 90–105. <https://doi.org/10.1093/jssam/smt008>.
- › Coelho, I. B., Pitta, M. T. and Silva, P. L. d. N. 2020. *Estimating State Level Indicators from ICT Household Surveys in Brazil*. *Statistical Journal of the IAOS* 36 (2): 495–508. <https://doi.org/10.3233/SJI-190511>.
- › Dever, J. A. 2018. *Combining Probability and Nonprobability Samples to Form Efficient Hybrid Estimates: An Evaluation of the Common Support Assumption*. In 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference, 15.
- › Elliott, M. R. 2009. *Combining Data from Probability and Non-Probability Samples Using Pseudo-Weights*. *Survey Practice* 2 (6): 1–7. <https://doi.org/10.29115/sp-2009-0025>.

References



- › Elliott, M. R., and Valliant, R. 2017. *Inference for Nonprobability Samples*. *Statistical Science* 32 (2): 249–64. <https://doi.org/10.1214/16-STS598>.
- › ITU, International Telecommunication Union. 2014. *Manual for Measuring ICT Access and Use by Households and Individuals*. https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITCMEAS-2014-PDF-E.pdf.
- › Little, R. J. A., and Rubin, D. B. 2002. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics.
- › Valliant, R. 2019. *Comparing Alternatives for Estimation from Nonprobability Samples*. *Journal of Survey Statistics and Methodology*, <https://doi.org/10.1093/jssam/smz003>.
- › Valliant, R., and Dever, J. A. 2011. *Estimating Propensity Adjustments for Volunteer Web Surveys*. *Sociological Methods and Research*. Vol. 40. <https://doi.org/10.1177/0049124110392533>.