Collecting ICT Households indicators during the COVID-19 pandemic: The experience of Cetic.br in Brazil

9TH MEETING OF THE EXPERT GROUP ON ICT HOUSEHOLD INDICATORS (EGH)

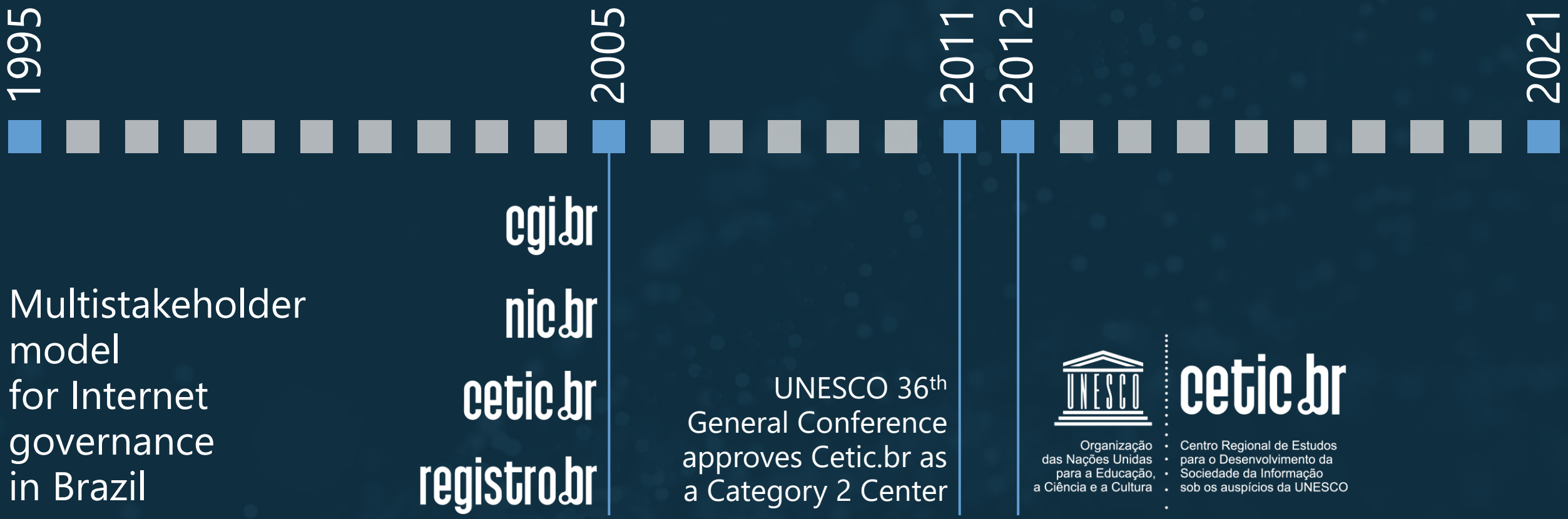
Marcelo Trindade Pitta
Winston Oyadomari

16 September 2021

cetic.br nic.br cgi.br

ABOUT CETIC.BR

15 YEARS PRODUCING ICT DATA FOR POLICYMAKING AND RESEARCH



Fonte: CGI.br/NIC.br (2021)

cetic.br

- ICT statistics production for policy making
- Capacity building
- National and international cooperation
- ICT indicators dissemination and analysis

INTRODUCTION

COVID-19 AND FACE-TO-FACE SURVEYS

Impracticality of traditional face-to-face interviews for most surveys

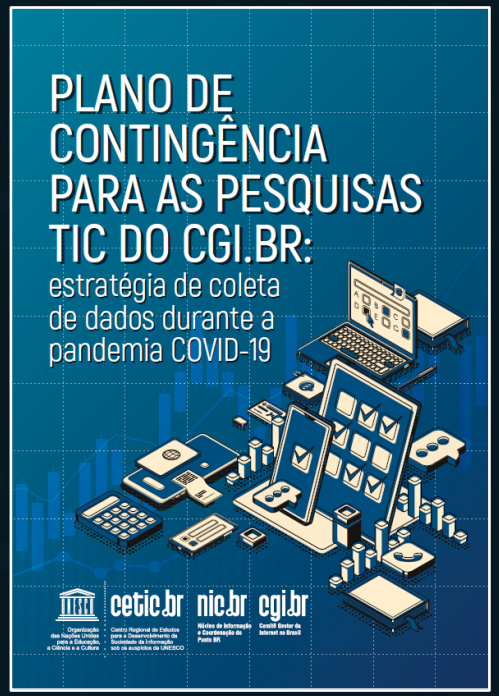
Alternative data collection methods had to be developed for population surveys and other surveys

Incomplete or partial frames as a critical barrier

INTRODUCTION
COVID-19 AND FACE-TO-FACE SURVEYS

Cetic.br developed a contingency plan to collect and publish ICT statistics based in alternative methods of data collection:

- Web panel survey with Internet users
- Telephone data collection for the traditional ICT Households survey



- » New methodological approaches
- » Data collection via CATI and WEB

COLLECTING DATA FROM A WEB PANEL

NONPROBABILITY SAMPLING

ICT Panel COVID-19 (Web panel survey)

Target population

Internet users aged 16+ in Brazil

Target domains

Sex (2), education (3), region (5), age group (5) and socioeconomic status - SES (4) – not cross-classified

Frame

Web panel of individuals obtained from market research companies, complemented by telephone lists (to reach population with lower SES/education)

Sample design

Quota sample based on region, sex, age group, SES, and education

COLLECTING DATA FROM A WEB PANEL

WEIGHTING

Calculating pseudo-weights based on a reference probabilistic survey: ICT Households 2019

Target population

Permanent private households and residents in permanent private households aged 10+ in Brazil

Total sample size

~ 30,000 interviews
(households and individuals)

Frame

IBGE 2010 census tracts database

Sample design

Stratified multi-stage sampling of households and residents

METHODS

APPROACH USED

Update the size of the target population (Internet users aged 16+) using data collected by the 2019 ICT Households survey combined with data from IBGE household survey

Evaluate and identify the population represented by respondents of the web panel survey, among those in the target population, through a predictive model for Internet use

METHODS

APPROACH USED

Estimate pseudo-inclusion probabilities for the non-probability sample units via logistic regression model and use their reciprocals as weights, considering thresholds defined by propensity scores of Internet use (model for Internet use)

Evaluate the results according to calibration factors and experts' knowledge

Estimate variances using bootstrap

RESULTS
EVALUATING PROS & CONS

ADVANTAGES

- Data collected avoiding face-to-face interviews
- The whole survey, from planning to publishing survey results, took less than two months to complete
- Cost of data collection much lower than traditional face-to-face surveys

RESULTS
EVALUATING PROS & CONS

DISADVANTAGES

- Web panel recruitment is not meant to be representative of the target population (Internet users)
- Coverage issues remain, despite using a probability survey as reference
- Approach is model-dependent: good models may not always be available
- Explanation of methodology and its dissemination is complex

ICT HOUSEHOLDS 2020
COLLECTING HOUSEHOLD DATA THROUGH CATI

ICT HOUSEHOLDS 2020

Target population

Permanent private households and residents in permanent private households aged 10+ in Brazil

Frame

All the respondents of ICT Households survey form years 2017, 2018 and 2019 that provided a valid telephone number (53.673 contacts)

Sample design

Stratified, multistage cluster sampling (the same as the past surveys)

ICT HOUSEHOLDS 2020

COLLECTING HOUSEHOLD DATA THROUGH CATI

~7% response rate, with indication of bias towards individuals with higher SES and more connected households and individuals

Attempts to correct the differences not possible by weighting methodologies

Solution: collect face-to-face data based on a small subsample of the frame used

ICT HOUSEHOLDS 2020

COLLECTING HOUSEHOLD DATA THROUGH F2F

Sampling selection of enumeration areas with no respondents in the CATI phase of the data collection

Procced the regular F2F collection method for the selected enumeration areas

Data collection: made in three weeks, w/ appropriate sanitary protocols

Response rate: 72%

ICT HOUSEHOLDS 2020

WEIGHTING

Separately weighting the two collection modes for their respective part of the original frame

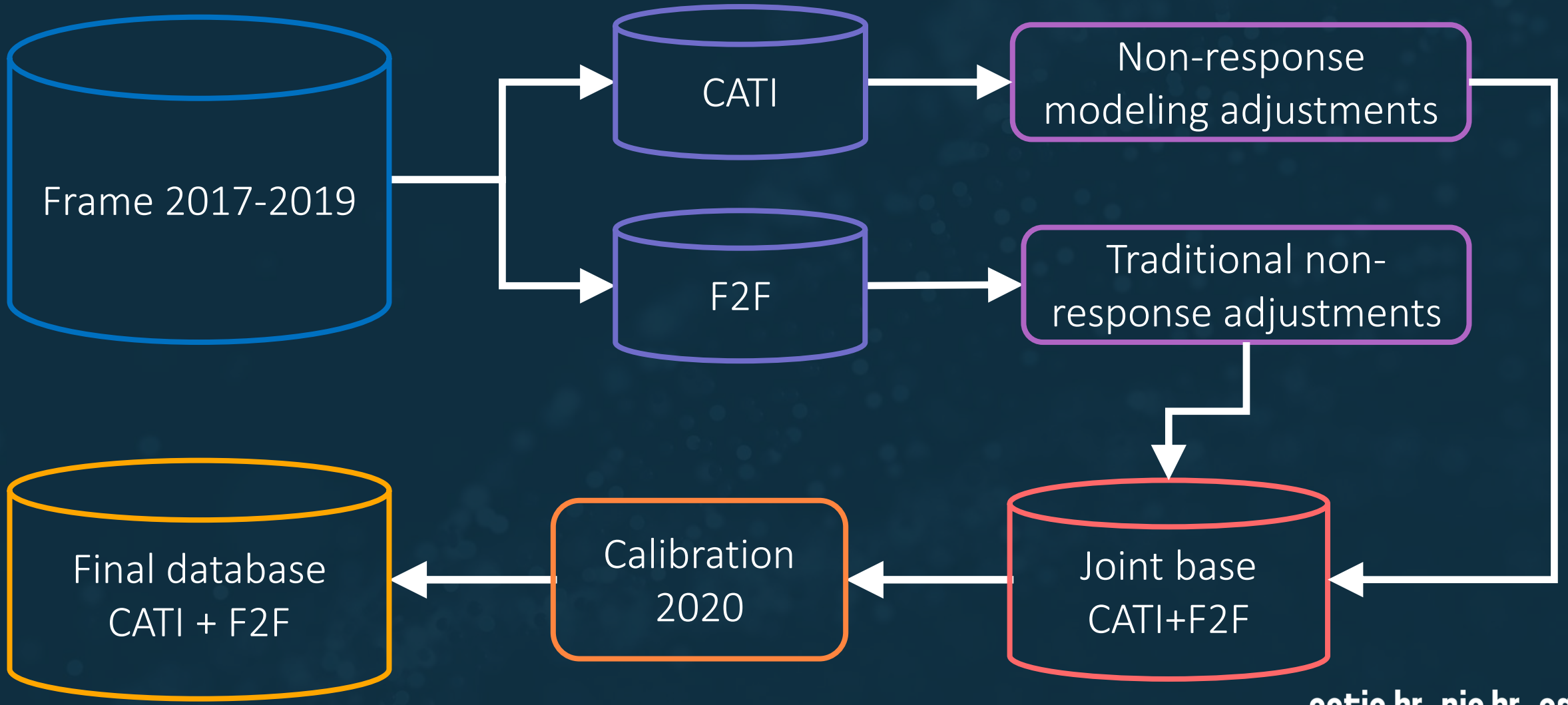
- CATI: weighing using modeling approaches and propensity scores methods
- F2F: weighting using traditional sampling techniques

Joining both modes and calibrating for known totals (IBGE – National household survey)

Estimating variances through bootstrap method

ICT HOUSEHOLDS 2020

WEIGHTING



RESULTS
EVALUATING PROS & CONS

ADVANTAGES

- Data collected minimizing face-to-face interviews
- Cost of data collection cheaper than a traditional face-to-face survey

CAVEATS

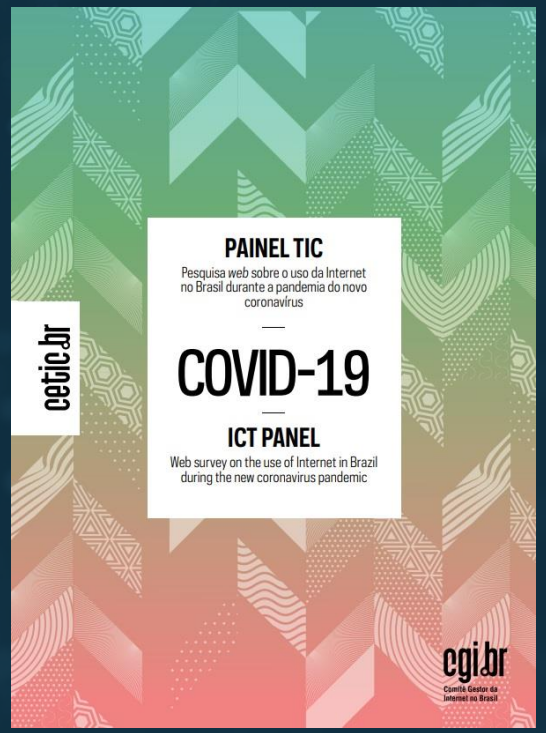
- Requires up-to-date database of telephone contacts (compliant with data privacy regulations)
- It was not possible to evaluate mode effects

RESULTS
EVALUATING PROS & CONS

DISADVANTAGES

- CATI requires shorter questionnaire (less information collected)
- Resulting sample smaller than the traditional sample
- Harder to explain the methodology and to disseminate microdata

Thank you all!



www.cetic.br

marcelopitta@nic.br

Access the survey in English/Portuguese:

https://cetic.br/en/publicacao/painel-tic-covid-19/

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.