

WAYS OF MEASURING INTERNET ACCESS IN BRAZIL BY DIFFERENT HOUSEHOLD SURVEYS

Marcus André Alves Zimmermann Vieira (ENCE)
Denise Britz do Nascimento Silva (ENCE)
Mayra Pizzott Rodrigues dos Santos (NIC.br)

ABSTRACT

Official statistics are mainly based on administrative records or sample surveys and censuses, such as household surveys, and investigate various social phenomena. This article investigates methodological aspects of ICT indicators on internet access (HH6), produced by Brazilian surveys: the Continuous National Household Survey by IBGE – the Brazilian NSO, and ICT Households by CETIC.br - a department of the Brazilian Network Information Center (NIC.br), comparing their methods, concepts, definitions and results. We highlight that operational definitions for internet access are not the same on both surveys, providing HH6 indicator estimates that are statistically different. IBGE adopts the definition of individual mobile service, as the household is classified as having Internet access if at least one resident has the service available. Meanwhile, ICT Households considers internet access as a utility service at home, available to any of the residents. It is interesting to point out that this reading also differs across the world as, for example, countries in the European Union, Japan, Australia and the United States follow the same line as IBGE, while countries in Africa, Mexico and Canada endorse the same definition as implemented by CETIC.br, which in turn ratifies the orientation of the ITU (International Telecommunication Union). Given the divergence of estimates on internet access from the two sources, the paper discusses concepts and definitions, and calculates indicators aiming at comparing survey estimates by bridging the data collection gaps. The final results demonstrate success of this compatibility and harmonizing exercise, allowing the production of proxy estimates and elucidating factors that may influence the measurement process.

Keywords: indicators, internet, survey.

1. INTRODUCTION

The internet is a great tool for integration and social promotion, and its access opens new horizons and opportunities for users. It has an impact on virtually all 17 Sustainable Development Goals (SDGs), but mainly goal 9/target 9c: “Significantly increase access to information and communications technology and strive to provide universal and affordable access to the Internet in least developed countries by 2020.” (UNITED NATIONS, 2023). Thus, the production of public statistics on this target is essential. Furthermore, it is necessary to understand the methodological aspects of different information sources to assess the comparability of the information produced.

Therefore, our objective is to compare the methods and results of household surveys that measure internet access in Brazil. In addition to describing the survey methodology, the operational

definitions and corresponding results of measuring internet access in household surveys in Brazil, we also propose a strategy for making data from Brazilian surveys compatible to compare the indicators according to the different operational definitions.

2. HOUSEHOLD SURVEYS ON ICT IN BRAZIL

We focus on the results for 2017, 2019 and 2021 (the last year of data availability) of the Continuous National Household Survey (PNADC), conducted by the Brazilian Institute of Geography and Statistics (IBGE), and the survey called Information and Communication Technology in Households (ICT Households), conducted by the Regional Center for Studies of Development of the Information Society (CETIC.br). The surveys' investigation methods are compared, as well as their published statistics, to discuss how question formulation impacts the data collection and, consequently, the measurement of the phenomenon.

With regard to capturing information about internet access, the PNADC includes the following question: "Does any resident have internet access at home using a microcomputer, tablet, cellphone, television or other equipment?" (IBGE, 2018, S01029), whose answer options are simply "yes" or "no". In turn, the question about access posed in the first block of the ICT Households survey is: "Does this household have internet access?" (CETIC.br, 2021, A4). In addition to having "yes" and "no", the CETIC.br survey also presents the alternatives "does not know" and "did not answer". However, these two extra options do not account for relevant values.

The two surveys have some general similarities between the elements of the statistical product: (i) there is delimitation of the place of access: the household (which is also the information reference unit); (ii) there is no restriction on the type of device used; and (iii) there is no time limitation (but it is assumed that the question refers to the availability of access at the time of the survey, as recommended by the International Telecommunication Union - ITU).

However, survey estimates present a difference of more than 8 percentage points between the two measurements in 2021. This value extrapolates the margins of error of the estimates, whose confidence intervals for the internet access indicator are [89.8%;90.2%] in the PNADC and [80.4%;82.6%] in the ICT Households. We highlight the fact that the surveys use different operational definitions to define the concept of household with internet access.

IBGE defines internet access at home considering an individual mobile service, since it is enough for at least one resident to have the service available, and register its use at home, to classify the household as having access to the internet, even if the resident does not share access with others. According to Vieira (2020), surveys carried out in the European Union, Japan, Australia and the United States share this same interpretation. CETIC.br, on the other hand, defines internet access as a utility service available to all residents, the same as access to water, sewerage or electricity service. Surveys carried out in Africa, Mexico and Canada also share this same operational definition (VIEIRA, 2020).

Therefore, the question wording and the operational definition applied in each survey affect the measurement of the indicator. According to De Vaus (2002), concepts are tools that fulfill a useful shorthand function: they are abstract summaries of a whole set of behaviors, attitudes and characteristics

that are considered to have something in common, while definitions are approaches with which concepts are used. Thus, nominal definitions are working approaches that are used for research and guidance on the type of information to be collected (DE VAUS, 2002). Further according to the same author, operational definition is the translation of a concept into a set of indicators and respective classification norms to provide the operational definition of the concept. Thus, both surveys indicate the concept of internet access as a home with internet service, but the application of the concept in each survey differs in operational definition.

3. COMPARABILITY AND COMPATIBILITY OF DATA BETWEEN BRAZILIAN SURVEYS

Comparability issues arise when surveys are compared, but differences in the results are not interpretable. Granda and Blasczyk (2016) proposed to establish standardized parameters that allow comparing points fairly, and specify these parameters with definitions, indicators, classifications, training and technical requirements. This is essential for analyzing differences between statistics, as it ensures that no confounders related to the way data were collected, processed and disseminated make it difficult to compare policy effects or analyze economic, social and environmental trends (BALDACC ET AL, 2016).

Harmonization is the development of standardized questions to be used in different surveys, according to De Vaus (2002), before data are collected, during survey planning. According to Granda and Blasczyk (2016), input harmonization also includes other standardized practices, such as sample design, data collection and data imputation. The ITU manual is an example of an initiative undertaken to harmonize data before they are generated. In order for there to be good harmonization, a uniform definition of concepts, questions and classification of codes in different surveys is necessary (DE VAUS, 2002).

Meanwhile, compatibility consists of formulating a method that equates different questions after collecting data in order to adapt them. As framed by Granda and Blasczyk (2016), output harmonization is used to take into account the specific conditions of the data or minimize the financial burden of non-uniformized statistical production. It is a more complex exercise, where the analyst needs to acquire reasonable knowledge about the subject and about the surveys in order to propose comparison strategies. Preference should always be given to the option of harmonization (input), as the second form - compatibility (output) - is a matter involving remediation, besides the fact that it usually requires greater effort and the use of assumptions that are not always optimal, but which may be necessary to make work possible.

In an attempt to reconcile the results of the HH6 indicator (proportion of households with internet access) from the ICT Households survey with the statistics produced by the PNADC, an exercise is carried out in this section. We present alternative ways of calculating estimates for the indicator, based on the information collected for the individual reference unit of the CETIC.br survey, aiming to obtain figures closer to those produced by the IBGE, as follows:

1. The household is considered to have access if the respondent reports that the household has internet access or that he/she has used the internet at home in the last three months;
2. It includes the previous version of the question plus identification of who used the internet via cellphone in the last three months;
3. It includes the previous version plus those who performed some activity at home that requires the internet even though they indicated they do not generally access the internet;
4. It includes the previous version, in addition to the response of a parent or guardian to the ICT Kids Online survey regarding use of the internet at home in the last three months (this survey regards internet use by children in Brazil also conducted by CETIC.br);
5. It includes the previous version plus a child's response to the ICT Kids Online survey on use of the internet at home in the last three.

It is important to point out that the ICT Kids Online survey uses part of the sample from the ICT Households survey. A subsample responds to the ICT Kids Online survey, in which there are two questionnaires: one for children and adolescents and another for parents/guardians. In 2021, 2,939 households were interviewed for the ICT Kids Online survey, according to CETIC.br.

The five listed alternative estimates were calculated using the R software, considering each survey's sampling plan, using household sample weights, based on microdata from the ICT Households and ICT Kids Online surveys. The premise adopted is that the information provided by the respondent represents the household, since there is no information available for all residents of the household due to the sample design. Table 1 presents estimates for 2017, 2019 and 2021.

Table 1: Original and alternative estimates on internet access (HH6) for Brazil in 2017, 2019 and 2021, and corresponding 95% confidence intervals

HH6 estimates	2017	2019	2021
ICT Households	60.8% [59;60.8]	71.4% [70.1;72.7]	81.5 [80.4;82.6]
Version 1	69.5% [68;71]	77.5% [76.4;78.6]	86.1 [85.2;87]
Version 2	72.9% [71.6;74.3]	79.8% [78.7;80.9]	87.2 [86.3;88.1]
Version 3	73.0% [71.7;74.4]	79.8% [78.8;80.9]	87.2 [86.3;88.1]
Version 4	73.9% [72.5;75.2]	80.4% [79.3;81.5]	87.6 [86.7;88.4]
Version 5	74.9% [73.5; 76.2]	80.9% [79.9;81.9]	87.9 [87;88.7]
PNADC	76.4% [76.1 ;76.7]	84.0% [83.7;84.3]	90 [89.8;90.2]

Source: Own elaboration based on PNADC, ICT Households and ICT Kids Online (2017, 2019 and 2021).

It can be observed that, for 2021, differences between ICT Households and PNADC estimates can be reduced from more than 8 percentage points to roughly almost 2, considering version 5. Therefore, this exercise was successful, since the amplitude of this difference fell gradually when certain variables were incorporated to approximate the operational definition of internet access in the two surveys. Indeed,

in 2017 the confidence interval for version 5 estimate actually overlapped the confidence interval of the PNADC statistics.

The fact that this difference is still marked for the last two survey editions reinforces the idea that the way each question is asked, and mainly the operational definition for concept associated with the phenomena of interest, directly impact its measurement. It is important to note that a compatibility exercise may not always be sufficient to equalize results after the data are generated. This supports the argument in favor of data harmonization, that can be much more efficient to reduce dissonance of the results by standardizing the research and survey processes before the actual data collection.

4. CONCLUSIONS

A main point addressed throughout this work was the duality of capturing information about internet access at home. There is no ideal way, to solve the issue of comparability of estimates for the HH6 indicator as captured by the ICT Households and PNADC surveys now, since they have different operational definitions for the calculation of this indicator. In any case, it is worth noting that the ITU, responsible for harmonizing standards and indicators for the ICT sector, suggests implementing the question in the form adopted by CETIC.br for the ICT Households survey.

To ensure the harmonization and comparability of statistics, all elements of the statistical product, including statistical concepts and nomenclatures, must be considered. Among these are the time frame, the reference unit, the spatial restriction, the composition of how the questions is asked and the questionnaire logical flow. When this does not occur in the planning and execution of stages of a survey, the subsequent exercise to guarantee comparability, can be difficult, as demonstrated in this work. Therefore, the best option is to promote consensus, and implementation, of concepts and definitions already established by a harmonizing body, such as the ITU.

REFERENCES

Baldacci, E.; Stoop, I.; Japac, L. Models of harmonization: now and in the future. Topic 7 – Better statistics for a globalized world. **Conference of European Statistics Stakeholders**. 2016. Available at http://www.ksh.hu/cess2016/pdf/cess2016_b9_5470.pdf. Accessed on 02/23/2020.

Regional Center of Studies for the Development of the Information Society. **ICT Households 2021. Questionnaire**. Available at: https://cetic.br/media/microdados/615/tic_domicilios_2021_individuos_domicilios_questionario_v1.0.pdf. Accessed on: 02/24/2023.

Regional Center of Studies for the Development of the Information Society. **ICT Households**. 2017, 2019, 2021. Available at: <https://cetic.br/pesquisa/domicilios/>. Accessed on 02/23/2023.

Regional Center of Studies for the Development of the Information Society. **ICT Children's Online**. 2017, 2019, 2021. Available at: <https://cetic.br/pt/pesquisa/kids-online/>. Accessed on: 02/24/2023.

Regional Center of Studies for the Development of the Information Society. **ICT Children's Online 2021. Questionnaire for children and adolescents**. Available at:

https://cetic.br/media/microdados/641/tic_kids_online_2021_question%C3%A1rio_crian%C3%A7as_v.1_0.pdf. Accessed on: 04/24/2023.

Regional Center of Studies for the Development of the Information Society. **ICT Children's Online 2021. Questionnaire for parents and guardians.** Available at: https://cetic.br/media/microdados/642/tic_kids_online_2021_question%C3%A1rio_pais_v1.0.pdf.

Accessed on: 04/24/2023.

De Vaus, D.A. **Surveys in social research.** 5th ed. ISBN 1 86508 611 8. 2002.

Granda, P.; Blasczyk, E. Data Harmonization. **Cross-Cultural Survey Guidelines.** 2016. Available at: <https://www.ccsr.isr.umich.edu/index.php/chapters/data-harmonization-chapter>. Accessed on 02/23/2020.

IBGE. **Continuous National Household Sample Survey.** Data. 2017, 2019, 2021.

IBGE. **Continuous National Household Sample Survey. Housing Questionnaire.** 2018. Available at: https://biblioteca.ibge.gov.br/visualizacao/instrumentos_de_coleta/doc5539.pdf. Accessed on: 02/23/2023.

United Nations. **Sustainable Development Goals.** Available at: <https://brasil.un.org/pt-br/sdgs>. Accessed on 02/25/2023.

Vieira, M. A. A. Z. **Formas de mensuração de acesso à internet no Brasil e no mundo por diferentes pesquisas domiciliares.** Dissertation (Master Degree in Population, Territory and Public Statistics) – National School of Sciences Statistics. Rio de Janeiro, 2020. Available at: https://ence.ibge.gov.br/images/ence/doc/mestrado/dissertacoes/2020/Dissertacao_MarcusZ.pdf.

Accessed on: 02/25/2023.