CSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

Comparison of sampling designs from the two editions of the *Brazilian National Health Survey*, 2013 and 2019

Comparação de desenhos amostrais das duas edições da *Pesquisa Nacional de Saúde*, 2013 e 2019

Comparación entre diseños muestrales de las dos ediciones de la *Encuesta Nacional de Salud* Brasil, 2013 y 2019

Paulo Roberto Borges de Souza Júnior [1]
Celia Landmann Szwarcwald [1]
Wanessa da Silva de Almeida [1]
Giseli Nogueira Damacena [1]
Marcel de Moraes Pedroso [1]
Carlos Augusto Moreira de Sousa [2]
Igor da Silva Morais [1]
Raphael de Freitas Saldanha [1]
Jefferson Lima [1]
Sheila Rizzato Stopa [3]

doi: 10.1590/0102-311X00164321

Abstract

Our objective is to describe the differences in the sampling plans of the two editions of the Brazilian National Health Survey *(PNS 2013 and 2019) and to evaluate how the changes affected the coefficient of variation (CV) and the design effect (Deff) of some estimated indicators. Variables from different parts of the questionnaire were analyzed to cover proportions with different magnitudes. The prevalence of obesity was included in the analysis since anthropometry measurement in the 2019 survey was performed in a subsample. The value of the point estimate, CV, and the Deff were calculated for each indicator, considering the stratification of the primary sampling units, the weighting of the sampling units, and the clustering effect. The CV and the Deff were lower in the 2019 estimates for most indicators. Concerning the questionnaire indicators of all household members, the Deffs were high and reached values greater than 18 for having a health insurance plan. Regarding the indicators of the individual questionnaire, for the prevalence of obesity, the Deff ranged from 2.7 to 4.2, in 2013, and from 2.7 to 10.2, in 2019. The prevalence of hypertension and diabetes per Federative Unit had a higher CV and lower Deff. Expanding the sample size to meet the diverse health objectives and the high Deff are significant challenges for developing probabilistic household-based national survey. New probabilistic sampling strategies should be considered to reduce costs and clustering effects.*

Health Surveys; Sampling Studies; Epidemiologic Research Design

Correspondence
P. R. B. Souza Júnior
Departamento de Informações para a Saúde, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz.
Av. Brasil 4365, Biblioteca de Manguinhos, sala 225, Rio de Janeiro, RJ 21040-360, Brasil.
paulo.borges@icict.fiocruz.br

[1] Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.
[2] Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.
[3] Ministério da Saúde, Brasília, Brasil.

Introduction

Population-based health surveys are an essential source of information for planning and evaluating health policies and programs [1]. When performed periodically, they can be used for the monitoring and the surveillance of the population's health conditions and of the indicators on morbidity, risk factors, and health system performance [2,3].

Health was incorporated into the *Brazilian National Household Sample Survey* (PNAD) of the Brazilian Institute of Geography and Statistics (IBGE) for the first time in 1981. A new health supplement on access to and use of health services was included in 1998 in the PNAD; which was subsequently applied in 2003, with minor modifications, beginning a five-year series of population-based health survey [4]. In 2008, motivated to continue the series, the third survey brought some changes and additions while keeping the essential aspects, allowing for the monitoring of health indicators and the comparison of results over the years [5,6].

Given the growing need for information to formulate policies in health promotion, surveillance, and care at national level, the development of a national survey designed specifically to collect information on health became necessary [7,8]. The *Brazilian National Health Survey* (PNS) is a household-based survey with a probabilistic and representative sample of the Brazilian population, conducted by the Brazilian Ministry of Health in partnership with the IBGE, which can produce estimates of various indicators at national and subnational levels, such as Federative Units (UF), capitals, and Metropolitan Areas [9,10]. This survey is part of IBGE's Integrated System of Household Surveys (SIPD) as an independent survey, with its own sampling design, which includes the random selection of a single eligible resident per household to answer a part of the questionnaire (more details are presented in the section of instruments) – different form the PNAD, in which all residents answer the Health Supplement questionnaire.

The PNS was carried out for the first time in 2013, based on three fundamental axes: the national health system performance; health conditions; and surveillance of diseases and health problems and associated risk factors [10,11]. Given the significant growth of chronic noncommunicable diseases (NCDs) in Brazil – responsible for more than 70% of premature deaths and loss of quality of life [12,13] – NCDs deserved specific attention in the survey, as with the associated risk factors, such as tobacco and alcohol use, physical activity, and eating habits [14]. Complementarily, in its first edition, the PNS 2013 included anthropometric and blood pressure measurements, as well as blood and urine collection to enhance knowledge about some biological markers in the Brazilian population; establishing national benchmarks specific to sociodemographic and geographic features of the Brazilian population [15].

The second edition of the PNS was held in 2019 and continued most of the modules covered in the first edition, involving a larger sample of households [16]. Some of the differences in the PNS 2019 include the change in age group, in which residents aged 15 years or over were considered for individual interviews, and the inclusion of new modules required by technical areas of the Brazilian Ministry of Health, namely: communicable diseases – addressing the symptoms of tuberculosis and leprosy and sexually transmitted infections (STIs); sexual behavior; and medical care, with a focus on access and quality of primary health care [17].

The possibility of monitoring the indicators estimated with the data from the two editions of the PNS and the complex sampling plan of the survey motivated the production of this study. We aim to describe the differences in the sampling plans of the two editions of the survey and to assess how changes in the sampling design of the PNS 2019 affected the coefficient of variation and the design effect of selected indicators estimates.

Methods

Survey design

The PNS is a cross-sectional, nationwide household-based survey carried out by the IBGE in partnership with the Brazilian Ministry of Health. The population surveyed corresponds to residents of permanent private households in Brazil, except those located in special census enumeration areas

(barracks, military bases, accommodation, camps, vessels, penitentiaries, penal colonies, prisons, jails, asylums, orphanages, convents, and hospitals).

In 2013, at the end of the fieldwork, 69,994 households were visited; and 64,348 household interviews were conducted, as well as 60,202 individual interviews. In 2019, 100,541 households were visited; and 94,114 household interviews were conducted, as well as 90,846 individual interviews. The non-response rates were, respectively, 8.1% and 6.4% [16].

The PNS was approved by the Brazilian National Ethics Research Commission (CONEP) in June 2013 (opinion n. 328.159) for the 2013 edition, and in August 2019 (opinion n. 3.529.376) for the 2019 edition.

Sampling

The PNS is part of the SIPD, in which the sampling structure is the Master Sample, consisting of a set of census enumeration areas or aggregates of census enumeration areas selected to support the household surveys carried out by the IBGE. The primary units of the Master Sample are stratified by four criteria: Administrative (UF, capitals, Metropolitan Areas, Integrated Development Regions (RIDE) and other UF census enumeration areas); Geographic (subdivisions of capitals and other large municipalities in districts, sub-districts, and neighborhoods); Situation (urban and rural); and Statistical, which subdivides the strata based on the three criteria aforementioned into homogeneous strata, according to information on total household income and number of private households [18].

The PNS sample is a sub-sample of the IBGE Master Sample. In the first stage of selection, the primary sampling units (PSU) are obtained by simple random sampling among those previously selected for the Master Sample, respecting PSU's stratification. In the second stage, a fixed number of permanent private households is selected by simple random sampling in each PSU selected in the first stage. The selection of households is made based on the National Address List for Statistical Purposes (CNEFE) in its last update before completing this stage of the sampling plan. In the third stage, a resident from each household in the sample is randomly selected, from a list of eligible residents at the time of the interview, to answer the third part (individual) of the questionnaire [18].

To calculate the PNS sample size needed to estimate parameters of interest at different levels of geographic disaggregation, the following aspects were considered: estimated proportions with the desired level of precision at 95% confidence intervals (95%CI); the design effect (Deff), since it is a multi-stage cluster sampling; the number of households selected per PSU; the proportion of households with people in the age group of interest [18]. The PNS sample estimates the leading indicators at the UF and capital levels; some indicators of interest, however, can be published at lower levels of geographic disaggregation: Metropolitan Region (excluding the capital) and UF (excluding the Metropolitan Area).

Data weighing

Sampling weights were calculated by the inverse product of the inclusion probabilities at each stage, including an adjustment factor for losses. Note that, regarding the information obtained in the questionnaires referring to household characteristics and the set of all household members, the weighting method correspond to the first two stages of the selection. Calibration was carried out based on population projections for Brazil and its UF [19] to modify the sample natural weights.

The IBGE calibrated the PNS 2013 sample weights considering the revised population projection of UF by gender and age for 2010-2060, released in 2018, to allow for comparisons between the 2013 and 2019 PNS editions (version released in August 24th, 2020). This same population projection was used to calibrate the PNS 2019 weights, thus ensuring comparability between the two editions [19].

Instruments

The PNS questionnaire is subdivided into three parts: the overall household, all household members, and the individual. The overall household and all household members' questionnaires are answered by a household resident who can provide information on the socioeconomic and health status of all

household members. The individual questionnaire is answered by an eligible resident selected with equal probability among all household residents. The PNS 2013 considered residents aged 18 or over as eligible to respond the individual interview, whereas the PNS 2019 included residents aged 15 or over [10,16]. IBGE tested all questions before the start of fieldwork to verify if it could be understood by the population throughout the different UF.

Anthropometrics

In the PNS 2013, the residents who were selected in the third stage had their weight, height, waist circumference, and blood pressure measured by field researchers using standardized equipment. Anthropometric measurements (waist circumference, weight, and height) and blood pressure were taken from 59,402 individuals, excluding pregnant women, refusing participants, and those for whom it was impossible to take these measurements. In the PNS 2019, weight and height measurements were performed in a PSU subsample in individuals aged 15 years or older selected in the third stage. For both surveys, the procedures for anthropometric measurements were developed by the Oswaldo Cruz Foundation (Fiocruz) with the Laboratory of Population Nutritional Assessment (LANPOP) of the School of Public Health, University of São Paulo (USP).

The sub-sample for anthropometric measurements was defined and proportionally allocated to the PNS strata, keeping a minimum of two PSU per stratum. Both primary units and households within them were selected by simple random sampling, and, in the selected households, measurement was carried out on the resident selected to answer the individual questionnaire [17]. Anthropometric measurements were implemented on 6,730 individuals aged 15 years or older (6,571 aged 18 years or older).

Biological material collection

The PNS 2013 also included collecting biological material (blood and urine), which was carried out in 2014 and 2015 by a consortium of private laboratories after the end of individual interviews. A subsample of 25% of the census enumeration areas was selected for the biological material collection, with a probability inversely proportional to the difficulty of collection, measured by the distance to a municipality with a large population (≥ 80,000). Since the sample did not reach a sufficient number in some strata due to fieldwork difficulties, post-stratification was proposed for data analysis [20], but it was not possible to consider the clustering effects.

The laboratory tests performed on 8,952 individuals include: glycated hemoglobin; total cholesterol; low-density lipoprotein (LDL) cholesterol; high-density lipoprotein (HDL) cholesterol; dengue serology; red blood cell count (erythrogram) and white blood cell count (leukogram); high-performance liquid chromatography (HPLC), for diagnosing hemoglobinopathies; and the estimated excretion of potassium, salt and sodium, and creatinine in urine [21].

PNS website

A website was created (https://www.pns.icict.fiocruz.br/) containing the research construction history, the outline of the two PNS editions, the questionnaires, the anthropometric and biological material collection instructions, the IBGE publications, and all supplements to studies on the PNS. Databases of the two PNS editions were included, as well as the database of laboratory tests, which was weighed and made available to users with no need of prior authorization.

The Panel of Indicators on the PNS website, using the Institute od Scientific and Technological Communications and Information in Health (ICICT/Fiocruz) Data Science Platform Applied to Health, was developed to characterize the socio-spatial trends of chronic diseases and other health problems, the lifestyles of the Brazilian population, and health care, regarding the use of health services in the 2013-2019 period. The panel presents health indicators by demographic, socioeconomic, and geographic features, presented in tables, charts, and maps available for download.

Each health indicator is documented with the definition form and calculation method. The indicators and the respective confidence intervals were built with R programming (http://www.r-project.

org), also publicly available on the PNS website. Since the data from the two PNS editions (2013 and 2019) were collected with a complex sampling design – which combines stratification of census enumeration areas, clustering, and unequal selection probabilities – the design effects were considered in the estimation of the standard errors of all indicators.

Comparing PNS 2013 and 2019 sampling designs

The analysis used data from PNS 2013 and 2019 after calibration by population projection. The estimates of variances were calculated by combining the primary cluster method and the linearization methods [22,23]. The CV shows the extent of variability in relation to the mean and was calculated as follows:

$$CV(\hat{p}) = \frac{SE_C(\hat{p})}{\hat{p}} X100$$

In which, $\hat{p}$ is the weighted mean estimate calculated for a given indicator and *SE* is the standard error of $\hat{p}$, estimated under a complex design.

The Deff was calculated by the ratio between the variance of $\hat{p}$ estimated under a complex design ($Var_C(\hat{p})$) and the variance of the $\hat{p}$ under a simple random sample ($Var_{SRS}(\hat{p})$) [24], as follows:

$$Deff = \frac{Var_C(\hat{p})}{Var_{SRS}(\hat{p})}$$

The stratification of the primary sampling units, the weighting of the sampling units, and the clustering effect in the PSU were considered to calculate Deff. For the indicators by UF, the Deff was calculated by comparing the survey's estimated variance with the simple random sample variance for each UF. The "survey" module of the software Stata, version 14.0 (https://www.stata.com), was used.

Some health indicators that were built with data from the two PNS editions were chosen. The selection of indicators considered variables from different parts of the questionnaire (all household members; individual) from different questionnaire modules and addressing percentages with different magnitudes. The value of the point estimate, the CV, and the Deff were calculated for each indicator.

The indicators considered in this study – those calculated with the data from the questionnaire of all household residents – correspond to health service used, usual source of care, and having a health plan.

The indicators calculated with the data from the individual questionnaire were related to self-rated health, oral health, self-reported diagnosis of at least one NCDs, self-report of high cholesterol, and chronic health problem. Indicators related to self-reported diagnosis of high blood pressure and diabetes were also analyzed by UF. The indicators of healthy behavior were also calculated, related to eating habits, physical activity, sedentary lifestyle, and alcohol and tobacco use.

The indicators of nutritional status were calculated using the body mass index (BMI), based on the measured weight and height data from the two PNS editions, considering the prevalence of overweight (BMI ≥ 25kg/m^2) and obesity (BMI ≥ 30kg/m^2).

To allow comparison between the two editions of the PNS, the indicators based on the individual questionnaire and nutritional status for 2019 were restricted to data of those aged 18 years or over.

Results

Table 1 shows the characteristics of the two PNS editions sample designs. Regarding the changes in sample design, there was an increase in the sample size of primary sampling units and a reduction in the sample of individuals subjected to anthropometric measurements. While weight and height were measured in 59,402 individuals aged 18 or over in 2013, a sub-sample of 6,730 individuals aged 15 or over was considered in the 2019 edition.

Table 1

Characteristics of the samples from the two editions of the *Brazilian National Health Survey* (PNS). Brazil, 2013 and 2019.

Sample characteristic	PNS 2013		PNS 2019	
	Questionnaire for all residents of the household	Selected resident questionnaire	Questionnaire for all residents of the household	Selected resident questionnaire
Number of selection stages	2	3	2	3
Sample size in the first stage *	6,062	6,055	8,028	8,027
Average of households with interviews conducted per PSU	10.6	9.9	11.7	11.3
Average of people interviewed per PSU **	33.9	9.9	34.8	11.3 (11.0 **)
Sample size of households with an interview conducted	64,348	60,202	94,114	90,846
Sample size of people interviewed	205,546	60,202	279,382	90,846
Sample size of people aged 18 or over with an interview conducted	145,580	60,202	207,845	88,531
Sample size of people aged 18 or over, with performed anthropometric measurements ***	-	59,402	-	6,571

PSU: primary sampling units.

* Number of selected PSU and with interviews carried out among people aged 18 years and over;

** People aged 18 and over;

*** Individuals whose weight and height measurements were taken during the survey fieldwork.

Table 2 shows the results corresponding to the indicators calculated with data from all household members. For most indicators, both the CV and the Deff were lower for the 2019 estimates. The proportion of individuals with health insurance and the proportion of individuals with the usual source of care had the highest Deffs in 2013 (18.6 and 18.1, respectively) and 2019 (18.7 and 13.8, respectively). On the other hand, the lowest Deffs were estimated for the proportion of individuals admitted to a hospital in the last 12 months (3.9 in 2013 and 3.8 in 2019) and the proportion of individuals who sought care in the last two weeks (5.3 in 2013 and 5.0 in 2019).

Table 3 presents the estimates for some indicators calculated with data from residents aged 18 or over, selected for the individual interview. The proportion of individuals aged 18 years or over with medical diagnosis of at least one chronic disease had the lowest CV (0.6%) in 2019. For 2013, the proportion of people aged 18 or over who reported a medical diagnosis of stroke had the highest CV, 5.7%. In 2019, the highest CV was 6.8% for the proportion of people aged 18 or over with obesity. Most indicators were more accurate for 2019, with lower CV.

The Deff ranged from 2.7 (proportion of individuals with poor or very poor self-assessment) to 4.2 (proportion of adult individuals who consume alcohol more than once a week) in 2013. In 2019, the Deff ranged from 2.7 (proportion of individuals with poor or very poor self-assessment and proportion of adults with a medical diagnosis of stroke) to 10.2 (proportion of obese people aged 18 years or older) (Table 3).

Table 4 shows the indicators for self-reported diagnosis of diabetes and hypertension per UF. Most UFs had a lower CV in 2019 for both indicators when compared to 2013. The prevalence of diabetes had higher CVs, ranging from 4.8% to 19.5% for Rio de Janeiro (2019) and Maranhão (2013). Bahia had a Deff of 4.4 for the prevalence of hypertension in 2013. Most UFs had a Deff ranging from 1 to 2 for the two indicators in 2019.

Table 2

Indicators of health services use among individuals aged 18 years and over in the two editions of the *Brazilian National Health Survey* (PNS). Brazil, 2013 and 2019

Indicator	PNS 2013 (n = 205,546)			PNS 2019 (n = 279,382)		
	%	CV	Deff	%	CV	Deff
Medical visit in the last 12 months	71.2	0.4	7.4	76.2	0.2	5.4
Dental visit in the last 12 months	44.4	0.7	8.8	49.4	0.5	7.4
Search for health care in the last two weeks	15.3	1.2	5.3	18.6	0.9	5.0
Usual source of care	77.8	0.5	18.1	76.5	0.4	13.8
Hospitalization for 24 hours or more in the last 12 months	6.0	1.7	3.9	6.6	1.4	3.8
Use of some integrative and complementary practice in the last 12 months	3.8	4.1	13.8	4.6	2.9	11.6
Having a private medical or dental health plan	27.9	1.5	18.6	28.5	1.3	18.7

%: estimate; CV: coefficient of variation of the estimate; Deff: design effect; n: unweighted count.

Table 3

Indicators of self-rated health and chronic noncommunicable diseases among individuals aged 18 years and over in the two editions of the *Brazilian National Health Survey* (PNS). Brazil, 2013 and 2019.

Indicator	PNS 2013				PNS 2019			
	n	%	CV	Deff	n	%	CV	Deff
Poor or very poor self-rated health	60,202	5.8	2.7	2.7	88,531	5.8	2.3	2.7
Medical diagnosis of hypertension	60,202	21.4	1.5	3.6	88,531	23.9	1.1	3.2
Medical diagnosis of diabetes	60,202	6.2	2.8	3.1	88,531	7.7	2.0	2.8
Medical diagnosis of high cholesterol	60,202	12.5	1.9	3.2	88,531	14.6	1.5	3.5
Medical diagnosis of cerebrovascular accident or stroke	60,202	1.5	5.7	3.1	88,531	2.0	4.0	2.7
Chronic back problem	60,202	18.5	1.7	4.1	88,531	21.6	1.3	4.1
Medical diagnosis of work-related musculoskeletal disease	60,202	2.5	5.1	4.0	88,531	2.5	5.0	5.5
Medical diagnosis of at least one chronic disease	60,202	45.0	0.9	3.6	88,531	50.8	0.6	3.7
Regular consumption of fruits and vegetables	60,202	28.7	1.3	4.0	88,531	32.4	1.0	3.9
Perception of high salt consumption	60,202	14.2	2.0	3.9	88,531	12.7	1.8	4.3
Alcohol consumption once or more a week	60,202	23.9	1.5	4.2	88,531	26.4	1.1	3.9
Regular consumption of alcohol	60,202	2.7	4.5	3.3	88,531	2.5	3.9	3.3
Alcohol abuse in the last 30 days	60,202	13.6	1.9	3.6	88,531	17.1	1.4	3.7
Driving a car or motorcycle soon after drinking in the last 12 months	9,537	24.4	3.4	3.6	21,735	17.0	2.7	3.3
Adequate leisure physical activity	60,202	22.7	1.4	3.7	88,531	30.1	1.0	4.0
TV time of 3 hours or more per day	60,202	29.0	1.3	4.1	88,531	21.8	1.2	3.5
Current use of tobacco byproducts	60,202	14.9	1.7	3.1	88,531	12.8	1.6	3.3
Current tobacco use	60,202	14.7	1.7	3.1	88,531	12.6	1.6	3.3
Former smokers	60,202	17.5	1.7	3.6	88,531	26.6	1.0	3.3
Current cigarette smoker	60,202	14.4	1.7	3.0	88,531	12.3	1.7	3.3
Total loss of teeth	60,202	11.0	2.2	3.5	88,531	8.9	2.0	3.2
Overweight	59,402	57.0	0.7	3.9	6,571	60.3	1.8	3.1
Obesity	59,402	20.8	1.5	3.5	6,571	25.9	6.8	10.2

%: estimate; CV: coefficient of variation of the estimate; Deff: design effect; n: unweighted count.

Table 4

Characteristics of the samples related to medical diagnosis of hypertension and diabetes in the two editions of the *Brazilian National Health Survey* (PNS) by Federation Units (UF). Brazil, 2013 and 2019.

UF	Sample characteristic		Diagnosis of high blood pressure						Diagnosis of diabetes					
	PNS 2013	PNS 2019	PNS 2013			PNS 2019			PNS 2013			PNS 2019		
	n	n	%	CV	Deff	%	CV	Deff	%	CV	Deff	%	CV	Deff
Rondônia	1,694	2,108	18.1	6.8	1.7	18.8	5.7	1.5	5.0	16.3	2.4	5.3	10.0	1.1
Acre	1,814	2,283	16.0	6.7	1.5	19.2	5.5	1.6	3.3	17.2	1.8	4.3	10.3	1.0
Amazonas	2,586	3,370	13.7	6.0	1.5	16.0	4.8	1.4	4.5	11.3	1.6	5.4	8.6	1.4
Roraima	1,591	2,135	13.9	7.5	1.5	15.7	6.2	1.5	3.8	15.0	1.4	5.0	10.9	1.3
Pará	2,004	3,696	12.9	7.4	1.6	15.3	4.9	1.5	3.9	17.4	2.5	5.6	8.6	1.6
Amapá	1,332	1,473	12.9	9.3	1.7	18.2	8.0	2.0	4.9	14.8	1.5	4.3	15.7	1.6
Tocantins	1,515	1,872	19.6	6.9	1.8	22.5	5.9	1.8	5.3	13.8	1.6	6.3	9.8	1.2
Maranhão	1,774	4,889	13.8	9.7	2.7	19.3	3.9	1.7	5.4	19.5	3.9	5.5	7.5	1.5
Piauí	1,804	2,674	19.3	7.3	2.3	23.6	4.6	1.7	5.1	14.1	1.9	6.8	9.0	1.5
Ceará	2,560	4,141	18.9	5.7	2.0	21.3	4.0	1.7	4.9	11.4	1.7	8.5	6.2	1.4
Rio Grande do Norte	1,691	2,877	20.7	6.1	1.6	21.9	4.0	1.3	5.5	13.5	1.8	8.8	9.0	2.2
Paraíba	1,943	3,068	21.6	4.9	1.3	25.1	4.4	2.0	4.5	12.8	1.5	7.5	7.6	1.4
Pernambuco	2,591	3,992	21.5	4.5	1.4	23.4	3.2	1.2	6.2	10.4	1.9	7.1	7.0	1.4
Alagoas	1,748	2,898	19.3	6.4	1.7	23.9	4.0	1.4	6.7	11.8	1.8	7.8	7.5	1.3
Sergipe	1,553	2,563	20.7	5.4	1.2	22.5	4.6	1.5	6.1	11.1	1.2	6.8	8.8	1.4
Bahia	2,641	3,600	20.1	8.1	4.4	25.2	4.1	1.9	5.0	11.6	1.9	6.7	8.3	1.7
Minas Gerais	3,779	5,128	23.8	5.5	3.5	27.7	3.6	2.4	6.5	10.9	3.1	8.0	6.2	1.7
Espírito Santo	1,724	3,463	20.5	7.0	2.2	25.5	3.7	1.5	6.0	14.5	2.3	6.8	8.5	1.8
Rio de Janeiro	3,486	4,849	24.0	3.9	1.7	28.1	2.9	1.6	6.3	7.0	1.1	9.3	4.8	1.1
São Paulo	5,305	5,995	23.0	3.6	2.1	24.2	3.2	1.9	7.7	6.5	1.9	8.6	5.6	1.7
Paraná	3,012	3,893	21.1	5.9	2.8	22.9	4.6	2.4	5.8	12.9	3.1	7.7	7.8	1.9
Santa Catarina	1,623	3,676	21.7	8.7	3.4	23.6	3.9	1.7	5.5	13.0	1.6	6.9	7.1	1.3
Rio Grande do Sul	2,913	3,707	25.0	4.7	2.2	26.6	3.8	1.9	7.1	9.2	1.9	8.8	6.9	1.6
Mato Grosso do Sul	1,809	2,805	21.1	5.5	1.4	24.5	4.5	1.7	7.8	8.3	1.1	7.8	8.9	1.8
Mato Grosso	1,476	2,423	20.7	5.6	1.2	21.6	5.2	1.7	6.2	12.3	1.5	6.6	13.5	3.0
Goiás	2,423	2,648	21.8	5.3	1.9	23.4	5.1	2.0	6.2	10.8	1.9	7.4	8.6	1.5
Federal District	1,811	2,305	19.8	5.8	1.5	16.6	5.9	1.6	5.8	10.9	1.3	6.6	9.4	1.4

%: estimate; CV: coefficient of variation of the estimate; Deff: design effect; n: unweighted count.

Discussion

The PNS is the main health survey in Brazil and is the gold standard for population estimates produced by sample surveys. Publicly available information serves as a reference for other research, such as the *Risk and Protection Factors Surveillance for Chronic Non-Comunicable Diseases Through Telephone Interview* (Vigitel) [25]. The PNS also provides information for monitoring global indicators, including those of the *Sustainable Development Goals* [26], the Global Action Plan for the Prevention and Control of Chronic Noncommunicable Diseases 2013-2020 [27], and the Strategic Action Plan to Tackle Noncommunicable Diseases in Brazil 2011-2022 [28].

The questionnaires from both PNS editions was designed to allow for a comparison with the data from the Health Supplement of the PNAD from previous editions (1998, 2003, and 2008) and with the data collected in Vigitel (2006-2019), continuing the spatiotemporal monitoring of a set of health indicators. In this sense, the Health Indicators Panel is a tool that enables the monitoring and surveillance of chronic diseases and their risk and protective factors, fulfilling the purpose of supporting priority health policies.

The inclusion of young people aged 15-17 years in the PNS 2019 allowed investigating health issues among Brazilian adolescents in this age group. The information obtained by a household survey, such as the PNS, has the advantage of being more comprehensive than the *Brazilian National Survey of School Health* (PeNSE) information, which only includes teenagers who attend school [29].

In the 2019 edition, there was a need to increase the sample size of households by expanding the age group of the individual interview to 15 years or older, which allowed analyzing more indicators by small-sized population groups, for which the estimates obtained by the PNS 2013 sample did not have adequate precision. Indeed, the coefficients of variation for less prevalent events, such as a diagnosis of stroke and work-related musculoskeletal disease (WMSD), decreased from 2013 to 2019 [30].

Another issue related to the increase in the sample size is the increase in survey costs, limiting other aspects addressed in the survey, such as anthropometric measurements. In 2019, weight and height measurements were taken in a reduced sub-sample, decreasing the precision of estimates, hindering statistical inference and estimation of temporal trends of anthropometric indicators in some population subgroups. This analysis shows that carrying out anthropometric measurements in the sub-sample in the PNS 2019 resulted in significant increases in the coefficient of variation of overweight and obesity indicators and the design effect.

The release of Deff is essential, as it provides parameters that can be used in other research with complex samples, allowing a more adequate sample size calculation [31,32]. A study on the nutritional status of several countries suggested another use for Deff. Since the clustering effects of healthy eating by geographic area were high, intraclass correlations can be used to focus on preventive interventions for overweight and obesity [33]. This study outcomes show sizeable design effects for the prevalence of obesity and could be used in initiatives to stop the growth of obesity in the country.

Conversely, very high clustering effects indicate that other sampling strategies must be discussed and developed since significant variances widen the confidence intervals and interfere with statistical tests [34,35]. In this analysis, some indicators – calculated with data from the questionnaire of all household members – had extremely high Deffs, such as the proportion of individuals who usually look for the same place, the same physician, or the same health service when they need health care (the usual source of care), showing a high intraclass correlation for this indicator. This is expected since the primary care units located near the census enumeration areas sampled in the survey tend to become a benchmark for the census enumeration areas's residents [36]; similar to the indicator for having a health plan, since households in a census enumeration areas have similar socioeconomic characteristics [37].

Although the Deffs were lower for the indicators calculated from the selected resident's information, they were higher than three for most indicators. In other words, the indicators related to NCDs and healthy behaviors also correlate within the census enumeration areas, indicating an association with the census enumeration areas's sociodemographic characteristics. In the classic statistical approach used in the PNS, the variance estimators are calculated considering only the primary sampling units. Separating the variance into more components, however, could decrease the clustering effects and provide better estimates of variance [38]. Additionally, the Deff reflects the clustering effect on the PSUs and the entire complex sampling design. In the case of the PNS, the Deff is influenced by stratification, clustering, unequal selection probabilities, weights adjusted by non-response rates, and calibrations by population projections, and even the imputation of missing data, which has been considered more recently [39]. Therefore, it is essential to verify the influence of the various elements involved in the effect of the sampling plan to guide the choice of more efficient designs [40].

Notably, the differences between the estimated CV for the indicators of the two editions of the PNS cannot be attributed solely to the differences between the sampling designs of the two surveys, since the indicators may have become more dispersed from one edition to another.

Expanding the sample size to meet the diversity of health objectives and the high Deffs are currently significant challenges for developing a probabilistic household-based national health survey. Reducing the number of selected households per PSU, thus increasing the number of PSUs, can be an alternative to reduce the Deff and improve estimates accuracy. However, new probabilistic sampling strategies should be considered to reduce costs and to design effects.

Contributors

P. R. B. Souza Júnior and C. L. Szwarcwald participated in the study conception, drafting, and data analysis and interpretation. W. S. Almeida, G. N. Damacena, M. M. Pedroso, C. A. M. Sousa and S. R. Stopa participated in the data analysis, writing of the article, and discussion of results. I. S. Morais, R. F. Saldanha and J. Lima participated in the methodology and data analysis. All the authors approved the final version of the manuscript for publication.

Additional informations

ORCID: Paulo Roberto Borges de Souza Júnior (0000-0002-8142-4790); Celia Landmann Szwarcwald (0000-0002-7798-2095); Wanessa da Silva de Almeida (0000-0002-5164-8603); Giseli Nogueira Damacena (0000-0002-7059-3353); Marcel de Moraes Pedroso (0000-0002-7323-2107); Carlos Augusto Moreira de Sousa (0000-0003-3279-5728); Igor da Silva Morais (0000-0003-2006-5250); Raphael de Freitas Saldanha (0000-0003-0652-8466); Jefferson Lima (0000-0002-7560-1068); Sheila Rizzato Stopa (000-0001-8847-665X).

Acknowledgments

We would like to thank the reviewers for the detailed review and new suggestions. To the Department of Health Surveillance, Brazilian Ministry of Health (TED 18/2019) for the financing.

References

1. Gouveia GC, Souza WV, Luna CF, Souza-Júnior PRB, Szwarcwald CL. Health care users' satisfaction in Brazil, 2003. Cad Saúde Pública 2005; 21 Suppl 1:S109-18.
2. Malta DC, Silva MMAD, Moura L, Morais Neto OL. The implantation of the Surveillance System for Noncommunicable Diseases in Brazil, 2003 to 2015: successes and challenges. Rev Bras Epidemiol 2017; 20:661-75.
3. Theme-Filha MM, Szwarcwald CL, Souza-Júnior PRB. Socio-demographic characteristics, treatment coverage, and self-rated health of individuals who reported six chronic diseases in Brazil, 2003. Cad Saúde Pública 2005; 21 Suppl 1:S43-53.
4. Travassos C, Viacava F. Utilização e financiamento de serviços de saúde: dez anos de informação das PNAD. Ciênc Saúde Colet 2011; 16:3646.
5. Lima-Costa MF, Matos DL, Camargos VP, Macinko J. Tendências em dez anos das condições de saúde de idosos brasileiros: evidências da Pesquisa Nacional por Amostra de Domicílios (1998, 2003, 2008). Ciênc Saúde Colet 2011; 16:3689-96.
6. Viacava F. Dez anos de informação sobre acesso e uso de serviços de saúde. Cad Saúde Pública 2010; 26:2210-1.
7. Malta DC, Leal MC, Costa MFL, Morais-Neto OL. Inquéritos Nacionais de Saúde: experiência acumulada e proposta para o inquérito de saúde brasileiro. Rev Bras Epidemiol 2008; 11 Suppl 1:159-67.
8. Ministério da Saúde. Portaria nº 1.002, de 13 de abril de 2017. Institui, no âmbito no Ministério da Saúde, o Comitê Gestor da Pesquisa Nacional de Saúde 2018 – PNS/2018. Diário Oficial da União 2017; 17 apr.
9. Instituto Brasileiro de Geografia e Estatística. Sistema Integrado de Pesquisas Domiciliares (SIPD). Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2007. (Textos para Discussão, 24).
10. Szwarcwald CL, Malta DC, Pereira CA, Vieira ML, Conde WL, Souza Júnior PR, et al. National Health Survey in Brazil: design and methodology of application. Ciênc Saúde Colet 2014; 19:333-42.
11. Damacena GN, Szwarcwald CL, Malta DC, Souza-Jr PRB, Vieira MLFP, Pereira CA, et al. O processo de desenvolvimento da Pesquisa Nacional de Saúde no Brasil, 2013. Epidemiol Serv Saúde 2015; 24:197-206.
12. Schmidt MI, Duncan BB, Silva GA, Menezes AM, Monteiro CA, Barreto SM, et al. Chonic non-comunicable diseases in Brazil: burden and current challenges. Lancet 2011; 377:1949-61.
13. Cardoso LSM, Teixeira RA, Ribeiro ALP, Malta DC. Premature mortality due to noncommunicable diseases in Brazilian municipalities estimated for the three-year periods of 2010 to 2012 and 2015 to 2017. Rev Bras Epidemiol 2021; 24 Suppl 1:e210005.

14. Buss PM, Hartz ZMA, Pinto LF, Rocha CMF. Health promotion and quality of life: a historical perDeffctive of the last two 40 years (1980-2020). Ciênc Saúde Colet 2020; 25:4723-35.
15. Malta DC, Szwarcwald CL, Pereira CA. National Health Survey, laboratory analyses and monitoring of noncommunicable diseases reduction targets. Ciênc Saúde Colet 2021; 26:1190.
16. Stopa SR, Szwarcwald CL, Oliveira MM, Gouvea ECDP, Vieira MLFP, Freitas MPS, et al. National Health Survey 2019: history, methods and perpectives. Epidemiol Serv Saúde 2020; 29:e2020315.
17. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Saúde 2019. Atenção primária à saúde e informações antropométricas. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2020.
18. Souza-Jr PRB, Freitas MPS, Antonaci GA, Szwarcwald CL. Desenho da amostra da Pesquisa Nacional de Saúde 2013. Epidemiol Serv Saúde 2015; 24:207-16.
19. Instituto Brasileiro de Geografia e Estatística. Nota técnica – informações referentes à revisão do plano tabular da PNS 2013. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2020.
20. Szwarcwald CL, Malta DC, Souza Júnior PRB, Almeida WDS, Damacena GN, Pereira CA, et al. Laboratory exams of the National Health Survey: methodology of sampling, data collection and analysis. Rev Bras Epidemiol 2019; 22 Suppl 2:e190004.
21. Malta DC, Szwarcwald CL, Silva Júnior JBD. First results of laboratory analysis in the National Health Survey. Rev Bras Epidemiol 2019; 22 Suppl 2:e190001.
22. Hansen MH, Hurwitz WN, Madow WG. Sample survey methods and theory. New York: John Wiley & Sons; 1953.
23. Pessoa DG, Silva PLN. Análise de dados amostrais complexos. São Paulo: Associação Brasileira de Estatística; 1998.
24. Kish L. Survey sampling. New York: John Wiley & Sons; 1965.
25. Departamento de Análise em Saúde e Vigilância de Doenças Não Transmissíveis, Secretaria de Vigilância em Saúde, Ministério da Saúde. Vigitel Brasil 2019: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico. Estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2019. Brasília: Ministério da Saúde; 2020.
26. Organização das Nações Unidas. Objetivos de desenvolvimento sustentável. https://nacoesunidas.org/pos2015/agenda2030/ (accessed on 13/Apr/2020).
27. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva: World Health Organization; 2013.
28. Departamento de Vigilância de Doenças e Agravos Não Transmissíveis e Promoção da Saúde, Secretaria de Vigilância em Saúde, Ministério da Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022. Brasília: Ministério da Saúde; 2011.
29. Reis AACD, Malta DC, Furtado LAC. Challenges for public policies aimed at adolescence and youth based on the National Scholar Health Survey (PeNSE). Ciênc Saúde Colet 2018; 23:2879-90.
30. Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci 2012; 5:7-13.
31. Szwarcwald CL, Damacena GN. Amostras complexas em inquéritos: planejamento e implicações na análise estatística de dados. Rev Bras Epidemiol 2008; 11 Suppl 1:38-45.
32. Bell BA, Onwuegbuzie AJ, Ferron JM, Jiao QG, Hibbard ST, Kromrey JD. Use of design effects and sample weights in complex health survey data: a review of published articles using data from 3 commonly used adolescent health surveys. Am J Public Health 2012; 102:1399-405.
33. Masood M, Reidpath DD. Intraclass correlation and design effect in BMI, physical activity and diet: a cross-sectional study of 56 countries. BMJ Open 2016; 6:e008173.
34. Franco C, Little RJA, Louis TA, Slud EV. Comparative study of confidence intervals for proportions in complex sample surveys. J Surv Stat Methodol 2019; 7:334-64.
35. Johnson DR, Lisa AE. Sampling design effects: do they affect the analyses of data from the national survey of families and households? J Marriage Fam 1998; 60:993-1001.
36. Escorel S, Giovanella L, Mendonça MHM, Senna MCM. The Family Health Program and the construction of a new model for primary care in Brazil. Rev Panam Salud Pública 2007; 21:164-76.
37. Instituto Brasileiro de Geografia e Estatística. Notas metodológicas. Estatísticas de gênero. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2020
38. Graubard BI, Korn EL. Modelling the sampling design in the analysis of health surveys. Stat Methods Med Res 1996; 5:263-81.
39. He Y, Shimizu I, Schappert S, Xu J, Beresovsky V, Khan D, et al. A note on the effect of data clustering on the multiple-imputation variance estimator: theoretical addendum to Lewis et al. (2014), *JOS*. J Off Stat 2016; 32:147-64.
40. Kalton G, Brick JM, Lê T. Estimating component of design effects for use in sample designs. In: United Nations, editor. Household sample surveys in developing and transition countries. 2nd Ed. New York: United Nations; 2005. p. 95-121.

Resumo

Nosso objetivo é descrever as diferenças nos desenhos amostrais das duas edições da Pesquisa Nacional de Saúde *(PNS 2013 e 2019) e avaliar como suas mudanças afetaram o coeficiente de variação (CV) e o efeito do desenho (Deff) de alguns dos indicadores avaliados. Variáveis de diferentes partes do questionário foram analisadas para avaliar proporções com diferentes magnitudes. A prevalência de obesidade foi incluída na análise uma vez que a medição de antropometria na pesquisa de 2019 foi realizada em uma subamostra. Os valores do estimador pontual, CV e Deff foram calculados para cada indicador considerando a estratificação das unidades amostrais primárias, a ponderação das unidades amostrais, e o efeito do agrupamento. Para a maioria dos indicadores, CV e Deff foram menores nas estimativas de 2019. Em relação aos indicadores para todos os membros familiares, Deffs foram elevados e atingiram valores superiores a 18 para a posse de um plano de saúde. Quanto aos indicadores no questionário individual, Deff variou de 2,7 a 4,2 em 2013 e de 2,7 a 10,2 em 2019 para a prevalência de obesidade. A prevalência de hipertensão arterial e diabetes por Unidade Federativa apresentou CV maior e Deff menor. A expansão do tamanho da amostra para atender aos diversos objetivos de saúde e Deff altos são desafios expressivos para o desenvolvimento de uma pesquisa nacional domiciliar probabilística. Novas estratégias de amostragem probabilística devem ser consideradas para reduzir custos e efeitos do agrupamento.*

Inquéritos Epidemiológicos; Amostragem; Desenho de Pesquisa Epidemiológica

Resumen

Nuestro objetivo es describir las diferencias en los diseños muestrales de las dos ediciones de la Encuesta Nacional de Salud *(PNS 2013 y 2019) y evaluar cómo sus cambios afectaron el coeficiente de variación (CV) y el efecto de diseño (Deff) de algunos de los indicadores evaluados. Se analizaron variables de diferentes partes del cuestionario para evaluar proporciones con diferentes magnitudes. La prevalencia de obesidad se incluyó en el análisis, ya que la medición de la antropometría en la encuesta de 2019 se realizó en una submuestra. Los valores del estimador puntual, CV y Deff se calcularon para cada indicador considerando la estratificación de las unidades de muestreo primarias, la ponderación de las unidades de muestreo y el efecto de agrupamiento. Para la mayoría de los indicadores, CV y Deff fueron más bajos en las estimaciones de 2019. En cuanto a los indicadores para todos los miembros de la familia, los Deff fueron altos y alcanzaron valores superiores a 18 por tener un plan de salud. En cuanto a los indicadores del cuestionario individual, Deff osciló entre 2,7 y 4,2 en 2013, y entre 2,7 y 10,2 en 2019 para la prevalencia de obesidad. La prevalencia de hipertensión arterial y diabetes por Unidad Federativa tuvo mayor CV y menor Deff. Un mayor tamaño de la muestra para cumplir con los diversos objetivos de salud y un alto valor de Deff son desafíos importantes para el desarrollo de una encuesta nacional domiciliar probabilística. Se deben considerar nuevas estrategias de muestreo probabilístico para reducir los costos y efectos de agrupamiento.*

Encuestas Epidemiológicas; Muestreo; Diseño de Investigaciones Epidemiológicas

Submitted on 01/Jul/2021
Final version resubmitted on 28/Mar/2022
Approved on 11/Apr/2022