DOI: 10.1590/1980-549720190004.supl.2

ORIGINAL ARTICLE / *ARTIGO ORIGINAL*

Laboratory exams of the National Health Survey: methodology of sampling, data collection and analysis

Exames laboratoriais da Pesquisa Nacional de Saúde: metodologia de amostragem, coleta e análise dos dados

Célia Landmann Szwarcwald[I], Deborah Carvalho Malta[II], Paulo Roberto Borges de Souza Júnior[I], Wanessa da Silva de Almeida[I], Giseli Nogueira Damacena[I], Cimar Azeredo Pereira[III], Luiz Gastão Rosenfeld[IV]*

ABSTRACT: ***Introduction:*** This article aims at describing the National Health Survey (*Pesquisa Nacional de Saúde* — PNS) methodology of collecting laboratory exams data. ***Methodology:*** A subsample of 25% of the census tracts was selected, according to the stratification of the PNS sample, with a probability inversely proportional to the difficulty of collection. The collection of blood and urine was done in the households by a laboratory agent, among residents selected for individual interview. Due to the difficulties found in the field work, the sample did not reach the minimum expected number in some strata, and a post-stratification procedure was proposed for the data analysis. ***Results:*** The collection of biospecimens was performed in 8,952 individuals. Laboratory tests were: glycated hemoglobin; total cholesterol; LDL cholesterol; HDL cholesterol; serology for *dengue*; red blood cell count (erythrogram) and white series count (leukogram); high performance liquid chromatography (HPLC) for diagnosis of hemoglobinopathies; creatinine. The excretion of potassium, salt and sodium and creatinine was estimated in the urine. The database of laboratory exams was weighed and made publicly available on the Oswaldo Cruz Foundation's PNS website and can be accessed without prior authorization. ***Conclusion:*** The total subsample of laboratory exams is of great value, since it allowed us to establish national reference parameters adequate to sociodemographic and geographic characteristics of the Brazilian population, providing relevant and complementary information for the analysis of the health situation of Brazil.

Keywords: Clinical laboratory techniques. Health surveys. Urine specimen collection. Blood specimen collection. Sampling studies.

[I]Institute for Communication and Scientific and Technological Information in Health, *Fundação Oswaldo Cruz* – Rio de Janeiro (RJ), Brazil.
[II]Department of Maternal Child Nursing and Public Health, School of Nursing, *Universidade Federal de Minas Gerais* – Belo Horizonte (MG), Brazil.
[III]Brazilian Institute of Geography and Statistics – Rio de Janeiro (RJ), Brazil.
[IV]Hematology Center of São Paulo – São Paulo (SP), Brazil.
*in memoriam.
Corresponding author: Célia Landmann Szwarcwald. Avenida Brasil, 4.365, 2º andar sala 225, CEP: 21040-900, Manguinhos, RJ, Brasil. E-mail: celia.szwarcwald@icict.fiocruz.br
Conflict of interests: nothing to declare – **Financial support:** *Secretaria de Vigilância em Saúde* (SVS), Ministry of Health.

RESUMO: ***Introdução:*** O artigo teve o objetivo de descrever a metodologia de coleta dos dados dos exames laboratoriais da Pesquisa Nacional de Saúde (PNS). ***Metodologia:*** Foi selecionada uma subamostra de 25% dos setores censitários, obedecendo à estratificação da amostra da PNS, com probabilidade inversamente proporcional à dificuldade de coleta. A coleta de sangue e urina dos moradores selecionados para entrevista individual foi realizada nos domicílios por um agente de laboratório. Por conta das dificuldades encontradas no trabalho de campo, a amostra não atingiu número suficiente em alguns estratos da pesquisa, então para a análise dos dados foi proposto procedimento de pós-estratificação. ***Resultados:*** A coleta de material biológico foi realizada em 8.952 indivíduos. Os exames realizados foram: hemoglobina glicada; colesterol total; colesterol LDL; colesterol HDL; sorologia para dengue; hemograma série vermelha (eritograma) e série branca (leucograma); cromatografia líquida de alta eficiência (HPLC) para diagnóstico de hemoglobinopatias; e creatinina. Na urina, estimativa de excreção de potássio, sal, sódio e creatinina. A base de dados dos exames laboratoriais foi ponderada e disponibilizada para os usuários no site da PNS da Fundação Oswaldo Cruz, sem necessidade de autorização prévia para uso. ***Conclusão:*** A subamostra total coletada é de grande valia, pois permitiu estabelecer parâmetros de referência nacionais adequados às características sociodemográficas e geográficas da população brasileira, fornecendo informações relevantes e complementares para a análise da situação de saúde do Brasil.

Palavras-chave: Exames e diagnósticos laboratoriais. Inquéritos de saúde. Coleta de urina. Coleta de amostras de sangue. Amostragem.

INTRODUCTION

As part of the Ministry of Health's project, focusing on the health conditions of the Brazilian population and on assessing the performance of the national health system, the National Health Survey (*Pesquisa Nacional de Saúde* – PNS) had the main objective of producing national data on the health situation and lifestyles of the Brazilian population, as well as health care, as regards access to and use of services, preventive actions, continuity of care, and financing of care[1].

The development of the research involved the collection of information along three main points: health conditions; health care (access, use, continuity of care and financing); non-communicable chronic disease (NCD) surveillance and associated risk factors[2]; in addition to issues related to social determinants such as the socioeconomic, demographic, cultural, behavioral and environmental factors that influence health[3].

The PNS is a nationwide household-based survey conducted by the Ministry of Health in partnership with *Fundação Oswaldo Cruz* and the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística* – IBGE) in 2013 and 2014. As part of the Integrated System Household Survey (Sistema Integrado de Pesquisas Domiciliares – SIPD), it benefited from the National Household Sample Survey Master Sample (*Pesquisa Nacional por Amostra de Domicílios*; continuous PNAD), with greater geographical spread and improved accuracy of estimates. Designed specifically to collect health information, the PNS was conceived to estimate various indicators at different levels of geographical breakdown[4].

An important advance of the PNS was the systematic measurement of weight, height and waist circumference, key markers for monitoring one of the most serious problems facing the epidemic of overweight and obesity today in Brazil[5]. In addition to blood pressure measurement, it is noteworthy that, for the first time, a nationally representative survey collected biological samples at the patient's home for additional tests.

This paper describes the methodology for collecting biological material (blood and urine) in its various stages: sampling, collection and analysis of data in a PNS subsample; and introduces the various articles in this supplement, which contain analyses of various biological markers, carried out on the occasion of the dissemination of the basis of PNS laboratory tests by IBGE.

METHODOLOGY

SAMPLING

PNS is part of IBGE's SIPD and used the same sample infrastructure as other IBGE household surveys. The methodology employed was cluster sampling in three selection stages, with stratification of the primary sampling units (PSU). As part of the SIPD, in the first stage, the selection of PSU (census tracts or sector composition) was performed by simple random sampling (SRS), thus maintaining the same stratification as the master sample. In the second stage, SRS selected a fixed number of permanent private households (10 to 14) in each PSU selected in the first stage. In the third stage, within each household in the sample, a resident aged 18 years old or older was also selected by SRS from a list of eligible residents put together at the time of the interview[4].

Weight, height, waist circumference and blood pressure of the selected adult resident were measured, as well as, in subsample, blood collection for laboratory tests was carried out. Urine collection was performed to obtain data on renal function and salt intake. At the end of the analysis of all laboratory tests, blood samples were stored, without identification of the subjects, in a serum bank located at *Fundação Oswaldo Cruz*.

Subjects selected for the survey provided informed consent for all procedures, including: questionnaire; weight, height and waist circumference measurements; blood pressure measurement; blood and urine collection; and sample storage without identification.

In order to collect biological material from the residents selected in the third stage of the PNS, a subsample of 25% of the surveyed census sectors was planned. To facilitate the logistics of biological material collection, census tracts were selected with inversely proportional probability to the difficulty of collection. In order to establish the difficulty of collection by numerical parameter, all municipalities with 80 thousand inhabitants or more in all Federative Units (FU) were identified. In each FU, the distances between all municipalities with less than 80 thousand inhabitants selected in the sample and the large population

municipalities (80 thousand inhabitants or more were calculated using the geographic coordinates of the centroids of the municipalities. Minimum distance was adopted as the numerical parameter to measure the difficulty of collection.

BIOLOGICAL MATERIAL COLLECTION

The PNS biological material collection stage was carried out with resources from the Institutional Development Support Program of the Unified Health System (*Programa de Apoio ao Desenvolvimento Institucional do Sistema Único de Saúde* – Proadi-SUS). A consortium of private laboratories, led by the DASA network, established through a partnership between the Ministry of Health and *Hospital Sírio Libanês*, under Proadi-SUS, was responsible for both the collection of PNS biological material and laboratory tests analysis. The choice of laboratories was made among those that met the quality control criteria of the Ministry of Health and for which the current norms for collection, transportation and processing of biological material were ensured.

To perform the collection, consortium laboratories received the following materials: multiple vacuum collection needle, venipuncture compression cotton, multiple collection needle adapter, tourniquet, occlusive dressing, collection site antiseptic alcohol swab, procedure gloves, *gelox*, styrofoam holder, styrofoam and thermal box. Collection kit consisted of the following materials: three tubes containing ethylenediaminetetraacetic acid (EDTA), two *sorogel* tubes, isolated sample urine collection kit, an inviolable password card for the research subject to retrieve their results via the internet, and the Informed Consent Form (ICF). For the analysis of laboratory tests, techniques and reagents standardized by SUS laboratories were used.

In the first stage of the PNS, during individual interviews in the households of the sectors selected for biological material collection, all participants were informed by the IBGE agent about the collection. Upon completion of all industry interviews, IBGE released the industry to begin laboratory data research. Upon permission of the interviewed residents, information about the individual's address, identification and characteristics such as gender and age was transmitted to the DASA network coordinator. Individual identification codes were created to ensure the confidentiality of the information.

Considered as a second phase of the PNS, laboratory test research was performed at the participants' homes by laboratory agents, and took place in 2014 and 2015. At the home visit, the laboratory technician explained to the participant the procedure to be performed. After signing the ICF, the collection kit was presented and instructions were given on how to receive the report, containing the results of the laboratory tests. Both blood and urine samples were randomly collected during home visit at different times of the day.

After analysis of laboratory tests, results were transmitted to IBGE in electronic media containing only the identification codes of the individuals.

Participants received test results directly from the contracted laboratory. Individuals with altered exams were referred for medical attention. Cases with results considered critical were identified immediately by the accredited laboratory, and assistance was provided in the SUS network.

DATA ANALYSIS

In the first phase of the PNS, 81,254 households were selected, of which 69,994 were occupied ones. The survey was conducted in 64,348 households and 60,202 adult individuals were interviewed, selected in each household with equal probability. Considering that 25% of the census tracts were selected for laboratory tests, and assuming a non-response rate of 20%, the expected number of individuals with laboratory data was approximately 12,000 individuals.

However, several factors caused a loss greater than 20% in the subsample of individuals indicated for laboratory tests. Among which, the following stand out: difficulty of locating the address by the hired laboratory; refusal of the selected resident to collect biological material; long time between the application of the questionnaire and the visit of the laboratory agent; and operational difficulties to transport biological material. In Brazil as a whole, laboratory tests were performed on 8,952 individuals, representing about 60% of residents interviewed individually in the initial phase of the PNS, in the sectors selected for laboratory research.

In addition to the higher than expected non-response rate, the loss distribution was not uniform and some strata had no sample representation, restricting the weighting of the data by natural expansion factors. Weightings were calculated for use of the database related to laboratory exams regarding post-stratification procedures by gender, age, race/color and level of education according to large region, from the total sample of the PNS.

To calculate the weights, we considered data from 60,202 residents selected for individual interviews in the initial stage of the PNS. Selected residents were aggregated into 4 age groups (18 to 29; 30 to 44; 45 to 59; 60 years old or older) and 3 levels of education (incomplete elementary; complete elementary and incomplete high school; complete high school or more). Regarding race/color, individuals of black, brown, white color/race and others were considered, resulting in 96 categories in each of the geographic macro-regions.

Mathematically, the N_h number of observations of selected PNS residents in each category h and n_h was calculated, as well as the number of corresponding observations in the subsample of laboratory tests. Then, the weighting factors (Wh) were calculated by Equation 1.

$$W_h = \frac{N_h}{n_h} \quad (1)$$

Post-stratification weights were calculated by Equation 2.

$$W_h = \frac{N_h}{n_h} x \frac{\Sigma n_h}{\Sigma N_h} \quad (2)$$

DISSEMINATION OF RESULTS

The collection of biological material was performed in 8,952 individuals. Laboratory tests performed on blood samples were: glycated hemoglobin; total cholesterol; low density lipoprotein cholesterol (LDL); high density lipoprotein cholesterol (HDL); *dengue* — immunoglobulin G (IgG) serology; red blood count (erythogram) and white blood count (leukogram); high performance liquid chromatography (HPLC) for diagnosis of hemoglobinopathies; and creatinine. And in urine, estimated potassium, salt, sodium and creatinine excretion.

Data provided contain results of laboratory tests and PNS information collected at the interview stage with the selected resident: gender, age, race/color, educational level, having a private health insurance, searching for a health professional in the past two weeks prior to the survey, hospitalization in the last 12 months, self-reported *dengue*, informing pregnancy at the time of the survey, eating habits, smoking, alcohol use, reported morbidity (hypertension, diabetes, cholesterolemia, heart disease, arthritis, insufficiency kidney disease, cancer), use of medicines for hypertension and diabetes, self-rated health, and physical measurements of weight, height, and blood pressure.

The post-stratification weights calculated by Equation 2 were used to analyze the data presented in the articles of this supplement. The laboratory test database with the weights calculated is available on the *Fundação Oswaldo Cruz* PNS website (www. pns.fiocruz.br) and may be used without prior authorization.

DISCUSSION

The total subsample collected, of almost 9 thousand individuals, is large and of great value, since it allowed to establish national reference parameters appropriate to the sociodemographic and geographical characteristics of the Brazilian population, providing relevant and complementary information for the analysis of the health situation in the country and by large region.

The analysis of laboratory tests performed in the other articles of this supplement established population diagnoses of cholesterolemia, diabetes, chronic kidney disease, anemia, prevalence of hemoglobinopathies such as sickle cell anemia, *dengue* serology and urinary sodium dosages, supporting the formulation, the monitoring and evaluation of public health policies, in addition to supporting the fight against the *dengue* epidemic. It is important to highlight that laboratory tests are complementary to those of reported morbidity, since

they depend on access to diagnosis[6], and can be used as markers of effective management of these diseases[7].

In addition, the findings form a baseline for monitoring global indicators from the World Health Organization and the United Nations, and the achievement of the goals of the Chronic Disease Management Plan and the Sustainable Development Goals, agenda 2030[8], such as how to stop the growth of diabetes and how to reduce sodium consumption in all social classes[9,10].

Equally of great relevance to Brazilian public health, the creation of the blood sample serum bank will allow future research with these data, to be defined by the Ministry of Health, such as the introduction of the Zika virus and Chikungunya, which brought a considerable burden to the health of the Brazilian population and huge challenges to the health system[11]. The most likely hypothesis of Zika virus entry into Brazil is related to the transmission of the virus by asymptomatic foreign tourists during the 2014 World Cup[12]. Speculation, however, about the period of Zika virus entry, based on mathematical models for the spread of the epidemic, suggest that the virus was introduced before the sports event[13]. Given the time of PNS biological material collection, hypotheses about the circulation of viruses from other countries may be better investigated with this data.

Due to difficulties in the collection of biological material, the sample represented 60% of the residents interviewed individually in the first stage of the PNS in the households of the sectors selected for laboratory research, and did not reach a sufficient number in some strata of the research. Limitations were mainly due to the fact that the conduction of the laboratory test research was not concomitant with the PNS home interviews, causing a series of operational difficulties. Among them are those particularly found in the northern region. Despite the selection with inversely proportional probability to the difficulty of collection, as the selection of the subsample followed the same stratification of the PNS, there was the inclusion of sectors of small and difficult to reach municipalities. Specifically in this region, operational problems consisted of traveling long distances, delaying collection time in the North and delaying the start of collection in other regions. Additionally, although households were identified with GPS coordinates, laboratory agents sometimes could not find the households surveyed in the previous phase of the PNS. It should be noted, however, that in addition to the operational difficulties presented, the most critical point of this operation was the nature of this survey, which requires more than information reported by the informant and which admittedly has a high rate of refusals[14].

Given that non-response rates were different according to the PNS strata, it was necessary to use a post-stratification procedure based on the total PNS sample. This technique corrects for non-representation of the sample in all strata and has been frequently used in household surveys, such as continuous PNAD, to compensate for loss due to non-response and to adjust for population totals generated by population projections in data disclosure domains[15].

Data calibration or post-stratification has also been used in several health surveys, such as the World Health Survey, which had different non-response rates by gender and age range[16].

In the Chronic Disease Risk and Protection Surveillance Telephone Survey (*Inquérito Telefônico de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas* – Vigitel), post-stratification weights are applied to compensate for low fixed telephony coverage in some state capitals[17].

Variables used to construct post-stratification weights were chosen to compensate for the losses in some strata and to adjust for sample totals estimated with data from the first stage of the PNS, analyzed after imputation procedures, and weighted by factors calculated by the inverse probability of selection according to the complex sampling design of the survey. Although post-stratification weights were applied to minimize the bias of differentiated stratum losses, a limitation of this procedure is the choice of variables and the construction of analysis categories. Changes in the set of variables or categories used may modify the estimates of the indicators of interest[18]. In the present study, categories of age, race/color, gender and macro-region of residence were used, but different age categorizations, for example, could have implications for the results.

CONCLUSION

The database of laboratory tests represents an important advance for the knowledge of characteristics and health conditions of the Brazilian population, allowing to draw the biochemical profile of clinical or preclinical conditions of the Brazilian population and to improve surveillance and effective management of non-communicable chronic diseases in the country, as well as supporting the fight against the *dengue* epidemic.

REFERENCES

1. Szwarcwald CL, Malta DC, Pereira CA, Vieira MLFP, Conde WL, Souza Júnior PRB, et al. Pesquisa Nacional de Saúde no Brasil: concepção e metodologia de aplicação. Ciên Saúde Colet 2014; 19(2): 333-42. http://dx.doi.org/10.1590/1413-81232014192.14072012
2. Damacena GN, Szwarcwald CL, Malta DC, Souza Júnior PRB, Vieira MLFP, Pereira CA, et al. O processo de desenvolvimento da Pesquisa Nacional de Saúde no Brasil, 2013. Epidemiol Serv Saúde 2015, 24(2): 197-206. http://dx.doi.org/10.5123/S1679-49742015000200002
3. Victora CG, Barreto ML, Leal MC, Monteiro CA, Schmidt MI, Paim J, et al. Health conditions and health-policy innovations in Brazil: the way forward. Lancet 2011; 377(9782): 2042-53. https://doi.org/10.1016/S0140-6736(11)60055-X
4. Souza Júnior PRB, Freitas MPS, Antonaci GA, Szwarcwald CL. Desenho da amostra da Pesquisa Nacional de Saúde 2013. Epidemiol Serv Saúde 2015; 24(2): 207-16. http://dx.doi.org/10.5123/S1679-49742015000200003
5. Malta DC, Bernal RTI, Vieira Neto E, Curci KA, Pasinato MTM, Lisbôa RM, et al. Tendências de fatores de risco e proteção de doenças crônicas não transmissíveis na população com planos de saúde no Brasil de 2008 a 2015. Rev Bras Epidemiol 2018; 21 (Supl. 1): e180020. http://dx.doi.org/10.1590/1980-549720180020.supl.1
6. Fang J, Yang Q, Ayala C, Loustalot F. Disparities in access to care among US adults with self-reported hypertension. Am J Hypertens 2014; 27(11): 1377-86. https://doi.org/10.1093/ajh/hpu061
7. Wang L, Mo N, Pang R, Deng Q, Liu Y, Hu Y, et al. Should quality goals be defined for multicenter laboratory testing? Lessons learned from a pilot survey on a national surveillance program for diabetes. Int J Qual Health Care 2016; 28(2): 259-63. https://doi.org/10.1093/intqhc/mzv121
8. GBD 2016 SDG Collaborators. Measuring progress and projecting attainment on the basis of past trends of the health-related Sustainable Development Goals in 188 countries: an analysis from the Global Burden of Disease Study 2016. Lancet 2017; 390(10100): 1423-59. https://doi.org/10.1016/S0140-6736(17)32336-X

9. Malta DC, Silva MMA, Moura L, Morais Neto OL. A implantação do Sistema de Vigilância de Doenças Crônicas Não Transmissíveis no Brasil, 2003 a 2015: alcances e desafios. Rev Bras Epidemiol 2017; 20(4): 661-75. http://dx.doi.org/10.1590/1980-5497201700040009
10. Pullar J, Allen L, Townsend N, Williams J, Foster C, Roberts N, et al. The impact of poverty reduction and development interventions on non-communicable diseases and their behavioural risk factors in low and lower-middle income countries: A systematic review. PLoS One 2018; 13(2): e0193378. https://doi.org/10.1371/journal.pone.0193378
11. Zanotto PMA, Leite LCC. The Challenges Imposed by Dengue, Zika, and Chikungunya to Brazil. Front Immunol 2018; 9: 1964. https://doi.org/10.3389/fimmu.2018.01964
12. Salvador FS, Fujita DM. Entry routes for Zika virus in Brazil after 2014 world cup: New possibilities. Travel Med Infect Dis 2016; 14(1): 49-51. https://doi.org/10.1016/j.tmaid.2015.10.004
13. Massad E, Burattini MN, Khan K, Struchiner CJ, Coutinho FAB, Wilder-Smith A. On the origin and timing of Zika virus introduction in Brazil. Epidemiol Infect 2017; 145(11): 2303-12. https://doi.org/10.1017/S0950268817001200
14. Gindi RM, Zipf G, Galinsky AM, Miller IM, Nwankwo T, Terry AL. Comparison of in-home collection of physical measurements and biospecimens with collection in a standardized setting: The Health Measures at Home Study. Vital Health Stat 2014; (164): 1-16.
15. Instituto Brasileiro de Geografia e Estatística. Diretoria de Pesquisas. Coordenação de Trabalho e Rendimento. Pesquisa Nacional por Amostra de Domicílios Contínua - Notas técnicas Versão 1.0. Rio de Janeiro: IBGE; 2017.
16. Vasconcellos MTL, Silva PLN, Szwarcwald CL. Sampling design for the World Health Survey in Brazil. Cad Saúde Pública 2005; 21 (Supl. 1): S89-S99. http://dx.doi.org/10.1590/S0102-311X2005000700010
17. Bernal RTI, Iser BPM, Malta DC, Claro RM. Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico (Vigitel): mudança na metodologia de ponderação. Epidemiol Serv Saúde 2017; 26(4): 701-12. http://dx.doi.org/10.5123/s1679-49742017000400003
18. Silva PLN. Calibration estimation. When and why, how much and how. Rio de Janeiro: IBGE; 2004. (Textos para Discussão, Diretoria de Pesquisas, nº 15).

Received on: 01/02/2019
Final version presented on: 02/19/2019
Approved on: 02/20/2019

Authors' contributions: C.L.S. and D.C.M. prepared the text; C.L.S. and P.R.B.S.J. proposed the data analysis methodology; G.N.D. and W.S.A. applied the proposed methodology; C.A.P. and L.G.R. coordinated research on laboratory tests. All authors contributed to the elaboration of the text.