



ORIGINAL ARTICLE



Comparison of calibration methods in the analysis of 2013 Brazilian National Health Survey data

Comparaç o de m todos de calibraç o na an lise de dados da Pesquisa Nacional de Sa de de 2013

Juliana Sena de Souza^I , M rcia Helena Barbian^{II} , Rodrigo Citton Padilha dos Reis^{I,II}

^IUniversidade Federal do Rio Grande do Sul, Graduate Program in Epidemiology – Porto Alegre (RS), Brazil.

^{II}Universidade Federal do Rio Grande do Sul, Department of Statistics – Porto Alegre (RS), Brazil.

ABSTRACT

Objective: This study aims to compare calibration methods for weights in the subsample of Laboratory Exams from the 2013 Brazilian National Health Survey (PNS), seeking to assess their representativeness and precision. **Methods:** Two alternative proposals for constructing calibrated weights were performed based on post-stratification and raking methods. A comparison between the weights provided for the Laboratory Exams subsample and the two suggested weights was conducted through parameter estimates using the 2013 PNS subsample data. Additionally, seven measures were used to assess the performance of the proposed weighting systems. **Results:** The alternative post-stratification and raking weights produced generalizable estimates for the target population of the 2013 PNS, while the original weights did not. The alternative methods showed similar performance to the original method, with a slight advantage for raking in some evaluation measures. **Conclusion:** It is recommended that basic design weights be documented and included in the public-use data files of the PNS. Furthermore, it is suggested to cross-reference information between the sample and subsample of the 2013 PNS to enable the exploration of methods such as data imputation, aiming to obtain more accurate and representative estimates. These improvements are essential to ensure the quality and usefulness of PNS data in epidemiological and public health studies.

Keywords: Population forecast. Sampling studies. Epidemiologic studies. Population studies in public health. Statistics as topic.

CORRESPONDING AUTHOR: Rodrigo Citton Padilha dos Reis. Programa de P s-graduaç o em Epidemiologia, Universidade Federal do Rio Grande do Sul. Rua Ramiro Barcelos, 2.400, sala 215, CEP: 90035-003, Porto Alegre (RS), Brasil. E-mail: citton.padilha@ufrgs.br

CONFLICT OF INTERESTS: nothing to declare.

HOW TO CITE THIS ARTICLE: Souza JS, Barbian MH, Reis RCP. Comparison of calibration methods in the analysis of 2013 Brazilian National Health Survey data. Rev Bras Epidemiol. 2025; 28: e250005. <https://doi.org/10.1590/1980-549720250005>

ASSOCIATED EDITOR: Jos  Cazuza de Farias J nior

SCIENTIFIC EDITOR: Cassia Maria Buchalla , Juraci Almeida Cesar

This is an open article distributed under the CC-BY 4.0 license, which allows copying and redistribution of the material in any format and for any purpose as long as the original authorship and publication credits are maintained.

Received on: 05/16/2024

Reviewed on: 10/09/2024

Accepted on: 11/07/2024



INTRODUCTION

The Brazilian National Health Survey (*Pesquisa Nacional de Saúde – PNS*), conducted by the Ministry of Health in collaboration with Fundação Oswaldo Cruz (Fiocruz) and the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística – IBGE*), represents the most extensive study ever undertaken in Brazil on health conditions and their determinants¹. It enables the assessment of access to diagnosis and healthcare services² while generating comprehensive data on the lifestyle patterns of the Brazilian population.

A significant innovation in the 2013 PNS was the inclusion of biological material collection (blood and urine samples) from a subsample comprising 25% of participants who completed the initial phase of the survey. This advancement enabled laboratory analyses and studies on the prevalence of anemia, total cholesterol levels, kidney failure, diabetes, and other health conditions, along with their associated factors within the Brazilian population³⁻⁷. Furthermore, this marked the first time such a study was conducted on a nationwide scale⁸.

Despite the innovative approach, challenges in fieldwork led to a loss of over 20% of the subsample designated for laboratory testing, resulting in 8,952 participants providing biological material. To address these losses and ensure the validity of the results, a post-stratification weighting method was applied for data analysis⁹.

Alternative techniques to post-stratification have been proposed in the literature, with raking¹⁰ and two-phase sampling calibration^{11,12} emerging as prominent methods. The evaluation of estimator performance, taking into account sampling weights and other design features, remains a critical area of focus in research on complex survey sampling designs^{13,14}.

In this context, the study aimed to compare various weight calibration methods to address distortions in the subsample of laboratory tests from the 2013 PNS and to enhance the accuracy and reliability of the resulting estimates. Performance evaluation measures¹⁵ were employed to identify the most effective calibration strategies, ensuring the optimal use of the data collected by the PNS and contributing to a more comprehensive understanding of public health in Brazil.

METHODS

Calibration in the subsample of Laboratory Tests of the 2013 PNS

Given that the Laboratory Tests subsample of the 2013 PNS was designed based on the distance between the sector selected (for the first phase of the survey) and municipalities with large populations (those with 80,000 or more inhabitants) within the state of the sector⁹, it was anticipated that a weighting system for a two-phase design ($\{d_k^*; k \in S_2\}$,

where S_2 represents the Laboratory Tests subsample) would be provided to statisticians, epidemiologists, and other researchers using the 2013 PNS data. This information could be utilized to derive estimates using the double expansion estimator, as outlined in Equation 3 of Supplementary Material 1, or even by employing the calibration estimator. This would enable data users to construct calibrated weighting systems based on the basic weights of the two-phase design d_k^* .

On the other hand, “post-stratification” weights were provided alongside the data from the Laboratory Tests subsample (*weight_lab*). These weights, denoted as $W_k^{(lab)}$, are defined in the article describing the subsample methodology⁹. In calculating the “post-stratification” weights, data from the 60,202 participants selected for individual interviews during the first phase of the 2013 PNS were considered. The following auxiliary variables were used to define the strata: gender (two levels: male and female); age (four groups: 18 to 29 years, 30 to 44 years, 45 to 59 years, and 60 years old or older); race/color (four categories: white, black, brown, and other); level of education (three levels: incomplete elementary school, completed elementary school and/or incomplete high school, and completed high school or higher); and geographic macroregion (five levels: South, Southeast, Central-West, North, and Northeast). This resulted in a total of 480 post-strata. The “post-stratification” weights were then defined by⁹ (Equation 1):

$$w_k^{(lab)} = \frac{N_h}{n_h} \times \frac{\sum n_h}{\sum N_h}, \text{ for } k \text{ belonging to stratum } h, \quad (1)$$

Where:

N_h : the number of residents selected from the 2013 PNS in each stratum,

h and n_h : the number of corresponding observations in the Laboratory Tests subsample⁹.

It is worth noting that, in the notation for a two-phase design, N_h and n_h would be represented as n_{1h} and n_{2h} , respectively.

Concerning the construction and application of the proposed “post-stratification” weights, three key observations are highlighted:

- Although it is not possible to confirm with certainty, the quantity $\sum n_h / \sum N_h$ appears to approximate $n_1/n_2 = 8,952/60,202$, representing the “sampling fraction” relative to the sample from the first phase of the 2013 PNS. Additionally, it is noted that the weights (1) do not seem to be genuine post-stratification weights, as they fail to incorporate the basic design weights;
- The estimates produced using data from the Laboratory Tests subsample and weights $W_k^{(lab)}$ are generalizable to the 2013 PNS sample of 60,202 participants, but not to the target population of the PNS, which comprises adults residing in permanent private households;

- When analyzing data from the Laboratory Tests subsample, the weights defined in (1) are applied in conjunction with the subsample design specification. Typically, the subsample is treated as having unequal inclusion probabilities, where the basic weights are represented by $W_k^{(lab)}$. This approach leads to variance calculations for the estimators based on these weights $W_k^{(lab)}$.

Regarding this last observation, it is suggested that the variance of the estimators should be calculated using the weights d_k , or a weighting system that closely approximates this expression. Additional details on the fundamental definitions of sampling designs, the formulation of basic weights, and the construction of calibrated weighting systems can be found in Supplementary Material 1.

Alternative calibration methods

Since the variables used to construct the weights $W_k^{(lab)}$ are available in the 2013 PNS microdata files for both the full sample of 60,202 participants and the Laboratory Tests subsample of 8,952 participants, an alternative method for analyzing the data from the PNS 2013 subsample is proposed. This approach involves the following design and weight construction: it is assumed that the Laboratory Tests subsample was selected through simple random sampling without replacement from the first-phase sample of the 2013 PNS. Accordingly, the basic weights take the form $d_k = n_1/n_2 = 60,202/8,952 \approx 6.72$ for all 8,952 participants in the second phase of the PNS 2013. Post-stratification weights were then constructed using the same auxiliary variables as the original weights $W_k^{(lab)}$ but adopting the "Population projection" variable (available in the PNS 2013 microdata as variable V00282) as the reference. Using this projection, the total population of Brazilian adults is estimated at 145,572,210. The post-stratification weights were calculated using Equation 2:

$$w_k^{PSAAS} = d_k \frac{N_h}{N_{h,\pi}} = \frac{n_1}{n_2} \times \frac{N_h}{N_{h,\pi}} = \frac{n_1}{n_2} \times \frac{N_h}{n_{2h}} \times \frac{n_2}{n_1} = \frac{N_h}{n_{2h}}, \quad (2)$$

for k belonging to stratum h , where n_{2h} corresponds to the number of participants in stratum h in subsample S_2 . The second-to-last equality in (2) is justified because $\hat{N}_{h,\pi} = \sum_{k \in S_2 \cap U_h} d_k = n_{2h} \times (n_1/n_2)$, given that we assume $d_k = n_1/n_2$.

Note that expression (2) is identical to the first equation in the article on the methodology of the Laboratory Tests subsample⁹, with the distinction that, in (2), N_h represents the number of residents in each stratum h in the Brazilian population, rather than the number of residents selected from the 2013 PNS in each stratum h . Thus, we propose a weighting system designed to produce estimates generalizable to the target population of the 2013 PNS.

A second alternative for creating a calibrated weights system, still assuming that the basic weights correspond to simple random sampling, was implemented using the

method known as raking. This iterative process involves sequentially post-stratifying each set of variables and repeating the procedure until the weights converge to a stable solution¹⁶. Raking enables the use of multiple grouping variables without requiring the construction of a complete cross-classification. The auxiliary variables used to construct these weights were the same as those used for the weights $W_k^{(PSAAS)}$, and the "Population projection" variable was also incorporated. The weights generated through the raking method will be denoted as $W_k^{(rakeAAS)}$.

Ethical aspects

This study utilized publicly available and anonymized data from the IBGE PNS. Consequently, approval from a Research Ethics Committee was not required.

Statistical analysis

To evaluate the underrepresentation and overrepresentation of the groups defined by the auxiliary variables (used to construct the calibrated weights), population projections (based on the 2013 PNS data) were assumed to represent the true population values. The relative frequencies estimated using the three weight systems $W_k^{(lab)}$, $W_k^{(PSAAS)}$, and $W_k^{(rakeAAS)}$ were then compared with the corresponding designs previously described.

To compare the proposed calibrated weights with the weights provided for the Laboratory Tests subsample, population parameters were estimated for selected variables of interest, as outlined in Table 1. Variables with codes beginning with the letter "Z" were collected during the Laboratory Tests phase of the 2013 PNS. In contrast, variables with codes starting with "Q" or "J" were collected during the first phase of the 2013 PNS and are available for the full sample of 60,202 participants.

To evaluate the estimates generated using the proposed weighting systems in $W_k^{(PSAAS)}$ and $W_k^{(rakeAAS)}$, the seven measures outlined in Supplementary Material 2 were calculated. The analyses were conducted using the R software¹⁷, version 17, along with the survey package¹⁸ to account for the sampling design. The R code specifying the design object (sydesign) used in the analyses is provided in Supplementary Material 3.

RESULTS

Weight distribution

The weight distributions of $W_k^{(lab)}$ and $W_k^{(PSAAS)}$ exhibit a similar shape; however, the values each weight system assumes differ significantly, reflecting the distinct objectives of each system. On one hand, $W_k^{(lab)}$ aims to represent the sample from the first phase of the 2013 PNS (Figure 1A). On the other hand, $W_k^{(PSAAS)}$ is designed to represent the target population of the survey, namely, Brazilian adults living in permanent private households (Figure 1B). The distribution of raking weights shows a distinct shape compared to

Table 1. Characteristics of interest in the Laboratory Tests subsample of the 2013 PNS selected for calibration estimate evaluation.

| Variable Code | Variable Description | Parameter Type |
|---------------|--|-------------------|
| Q030 | Diabetes (<i>yes, no, only during pregnancy</i>) | Total prevalence |
| Q002 | Hypertension (<i>yes, no, only during pregnancy</i>) | Total prevalence |
| Q060 | High cholesterol (<i>yes, no</i>) | Total prevalence |
| Q063 | Heart disease [heart attack, angina, heart failure, or other] (<i>yes, no</i>) | Total prevalence |
| Q068 | Stroke or cerebral hemorrhage (<i>yes, no</i>) | Total prevalence |
| Q120 | Cancer (<i>yes, no</i>) | Total prevalence |
| Q124 | Chronic kidney failure – CKD (<i>yes, no</i>) | Total prevalence |
| J058 | Dengue (<i>yes, no</i>) | Total prevalence |
| Z025 | Creatinine (mg/dL) | Total and average |
| Z031 | Total cholesterol (mg/dL) | Total and average |
| Z034 | Glycated hemoglobin – HbA1c (%) | Total and average |

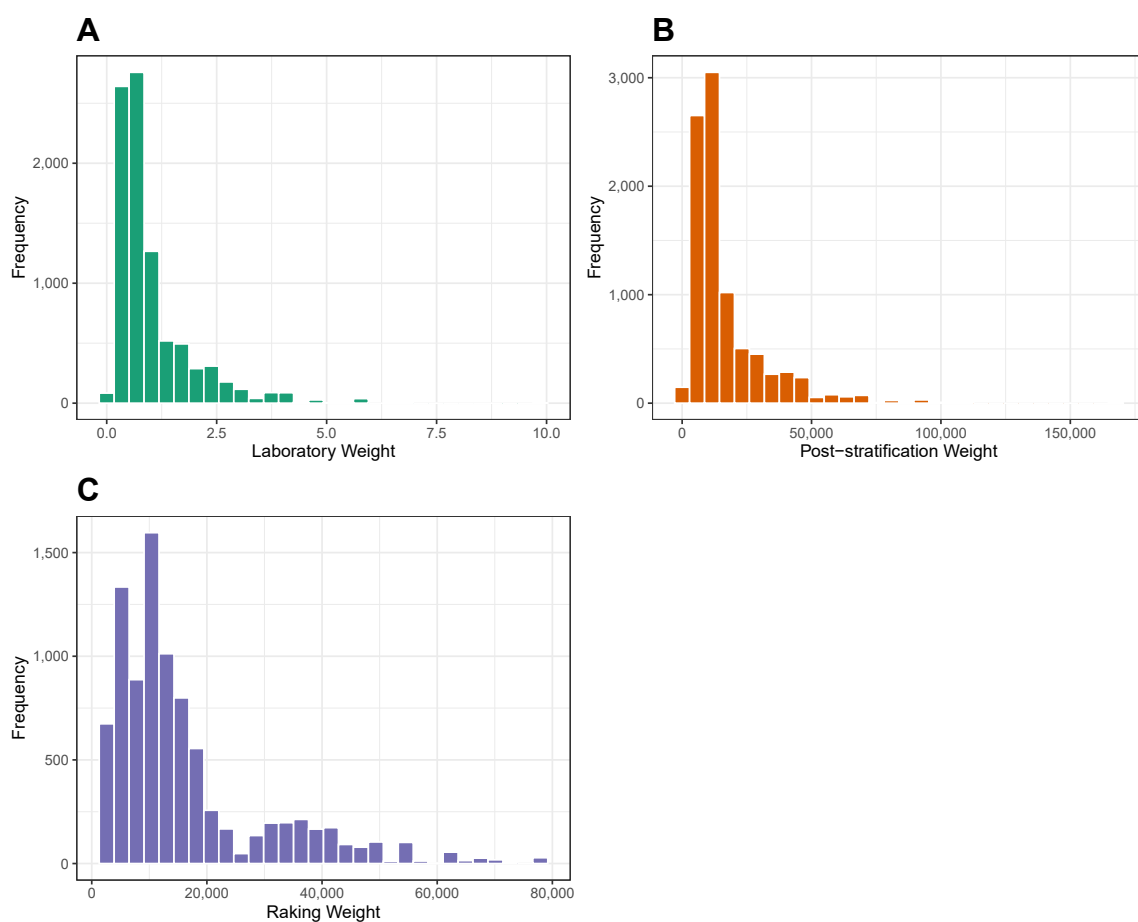


Figure 1. Distribution of weights (A) post-stratification weights provided with the Laboratory Tests subsample data from the 2013 PNS; (B) post-stratification weights; and (C) raking weights constructed from the population projection – Laboratory Tests subsamples, 2013 PNS.

the other weight distributions of $W_k^{(lab)}$ and $W_k^{(PSAAS)}$, but the values are similar to those of $W_k^{(PSAAS)}$. Therefore, the raking weights $W_k^{(rakeAS)}$ also yield generalizable estimates for the target population of the 2013 PNS (Figure 1C).

Representation of post-strata

The weights $W_k^{(lab)}$ produce estimates of the proportions for the auxiliary variables that are close to the population

projections (Figure S1, Supplementary Material). Generally, the estimates based on $W_k^{(lab)}$ show a difference (absolute error) of no more than 0.15%, except for the categories of brown race/color (0.22%), other race/color (0.42%), and the Northeast region (-0.26%). The estimates obtained using the weights $W_k^{(PSAAS)}$ also present an absolute error typically no greater than 0.15%, with the exceptions of white race/color (0.21%), brown race (-0.20%), and other race/color

(0.43%). For the weights calibrated using the raking method, $W_k^{(rakeAAS)}$, the estimates of the proportions for the auxiliary variables align precisely with the population projections. Similar results were obtained when comparing the estimates of the population totals for the auxiliary variables (Table S1, Supplementary Material 4). While the relative error (RE) for $W_k^{(lab)}$ was considered, it became evident that the population estimates derived from these weights diverge significantly from the population projections, resulting in an exceptionally high RE — therefore, these results were not presented.

Calibration assessment measures

The seven evaluation measures for the calibration methods are presented in Table 2. The mean absolute calibration RE was $M_1 = 2.16$ for the post-stratification method and $M_1 = 0$ for raking. These results were anticipated, as the estimated totals for the auxiliary variables were generally lower than the population totals. The mean coefficients of variation for the totals of the auxiliary variables $M_2 = 0\%$ indicate that the estimates produced by both methods are unbiased.

The measures of M_3 and M_4 (proportion of extreme weights) indicate the presence of extreme g weights for both the post-stratification and raking methods. This outcome was anticipated, considering the difference between the design weights d_k and the constructed weights $W_k^{(PSAAS)}$ and $W_k^{(rakeAAS)}$ ($M_6 = 13,440.07$ for post-stratification and $M_6 = 11,447.09$ for raking). The coefficient of variation for the g weights further reflects this characteristic of the calibrated weight construction and was high for both methods ($M_5 = 91.37\%$ for post-stratification and $M_5 = 83.94\%$ for raking), with a slight advantage observed for the raking method.

The average efficiency of the estimates (M_7) by the alternative calibration methods for the set of variables presented in Table 1 indicates a slight advantage of the post-stratification method over raking.

Accuracy of estimates of parameters of interest

Table 3 presents the estimates for the parameters related to the characteristics of interest (listed in Table 1), along with the coefficients of variation obtained from the

three calibrated weighting systems. It can be observed that the point estimates (totals, prevalences, and means) produced by the three methods are generally very similar. The exception is the estimate of population totals from the post-stratification weights of the Laboratory Tests subsample itself, as indicated by the weights $W_k^{(lab)}$. As previously noted, these results are not generalizable to the target population of the 2013 PNS. Lastly, the estimates of the coefficients of variation for the estimates show greater precision than those of the alternative calibration methods proposed in this study. This outcome was anticipated, as the estimation of the standard error of the estimates incorporates aspects of the assumed sampling design.

To evaluate the performance of the calibration methods in estimating population subgroups, estimates of diabetes prevalence were obtained based on several characteristics of interest (Table 4). Once again, it can be observed that the point estimates are very similar across the three weighting systems, with the apparent advantage of the alternative calibration methods being seen in the precision of the estimates. The 95% confidence intervals for the prevalence of diabetes in population subgroups, produced by the post-stratification ($W_k^{(PSAAS)}$) and raking ($W_k^{(rakeAAS)}$) methods, are slightly narrower than those produced by the weight $W_k^{(lab)}$.

DISCUSSION

The 2013 PNS included the collection of a subsample of laboratory tests, which represents a significant contribution to studies on the health of the Brazilian population. Sampling techniques suggest that a two-phase design could have been employed in the 2013 PNS subsample to construct both basic and calibrated weighting systems. However, challenges in collecting the second-phase sample led to the non-disclosure of the design weights alongside the microdata of the Laboratory Tests subsample. In the absence of basic sampling weights, the managers of the subsample data provided calibrated weights using the post-stratification method.

This study proposed two alternative calibration methods based on post-stratification and raking. The weighting

Table 2. Performance measures of post-stratification and raking calibration methods – Laboratory Tests subsample, 2013 PNS.

| Evaluation Measure | $W_k^{(PSAAS)}$ | $W_k^{(rakeAAS)}$ |
|--|-----------------|-------------------|
| M1: Mean of absolute relative calibration error | 2.16 | 0.00 |
| M2: Mean of the coefficient of variation (CV) of the totals of auxiliary variables | 0.00% | 0.00% |
| M3: Proportion of extreme weights relative to 1/3 | 19.02% | 18.81% |
| M4: Proportion of extreme weights relative to 3 | 4.39% | 3.24% |
| M5: Coefficient of variation of the weights g | 91.37% | 83.94% |
| M6: Distance χ^2 between w and d | 13,440.07 | 11,447.09 |
| M7: Average efficiency of calibration estimates | 1.005 | 1.018 |

Table 3. Estimated totals and prevalences (%)[‡], and coefficient of variation (CV%) of the characteristics of interest obtained from the three calibrated weight systems — Laboratory Tests subsample, 2013 PNS.

| Characteristics of interest | $W_k^{(lab)}$ (N = 8,952) | | $W_k^{(PSAAS)}$ (N = 144,908,922) | | $W_k^{(rakeAAS)}$ (N = 145,572,210) | |
|---------------------------------|------------------------------|-------|--------------------------------------|-------|--|-------|
| | Estimate (%) | CV% | Estimate (%) | CV% | Estimate (%) | CV% |
| Diabetes | 573 (7.69) | 5.04 | 9,275,930 (7.7) | 4.76 | 9,266,527 (7.65) | 4.85 |
| Hypertension | 1,930 (25.91) | 2.53 | 31,236,815 (25.92) | 2.18 | 31,171,025 (25.75) | 2.25 |
| High cholesterol | 1,160 (15.57) | 3.42 | 18,773,491 (15.58) | 3.2 | 18,842,610 (15.56) | 3.25 |
| Heart disease | 370 (4.96) | 6.15 | 5,971,338 (4.95) | 5.83 | 6,092,130 (5.03) | 5.98 |
| Stroke | 146 (1.96) | 11.57 | 2,367,716 (1.96) | 10.47 | 2,266,318 (1.87) | 10.3 |
| Cancer | 158 (2.11) | 9.87 | 2,540,009 (2.11) | 9.45 | 2,549,079 (2.11) | 9.57 |
| CKD | 111 (1.48) | 12.04 | 1,785,325 (1.48) | 11.78 | 1,815,305 (1.5) | 11.73 |
| Dengue | 1,406 (18.88) | 3.06 | 22,735,194 (18.86) | 2.9 | 22,732,548 (18.78) | 2.91 |
| Creatinine (Mean, mg/dL) | 0.881 | 0.44 | 0.881 | 0.38 | 0.882 | 0.44 |
| Total cholesterol (Mean, mg/dL) | 185.031 | 0.31 | 185.01 | 0.29 | 184.815 | 0.29 |
| HbA1c (Mean, %) | 5.513 | 0.23 | 5.513 | 0.22 | 5.51 | 0.23 |

[‡]For the characteristics Creatinine, Total Cholesterol, and HbA1c, both the mean and CV were estimated.

Table 4. Prevalence of diabetes (%) and 95% confidence interval (95%CI) by population subgroups obtained from the three calibrated weighting systems – Subsample of Laboratory Tests, PNS 2013.

| Characteristics | $W_k^{(lab)}$ | | $W_k^{(PSAAS)}$ | | $W_k^{(rakeAAS)}$ | |
|--|---------------|-----------|-----------------|-----------|-------------------|-----------|
| | % | 95%CI | % | 95%CI | % | 95%CI |
| Age | | | | | | |
| 18 to 29 years | 1.4 | 0.5–2.4 | 1.4 | 0.6–2.3 | 1.3 | 0.5–2.0 |
| 30 to 44 years | 2.4 | 1.6–3.3 | 2.4 | 1.6–3.2 | 2.4 | 1.6–3.1 |
| 45 to 59 years | 10.8 | 9.1–12.5 | 10.8 | 9.2–12.5 | 11.1 | 9.4–12.8 |
| ≥60 years | 18.2 | 15.9–20.5 | 18.2 | 16.0–20.5 | 17.9 | 15.6–20.2 |
| Race/color | | | | | | |
| White | 7.3 | 6.2–8.4 | 7.3 | 6.3–8.3 | 7.3 | 6.3–8.4 |
| Black | 9.6 | 6.8–12.5 | 9.7 | 7.1–12.2 | 9.7 | 6.9–12.4 |
| Brown | 7.2 | 6.2–8.3 | 7.2 | 6.2–8.2 | 7.1 | 6.1–8.2 |
| Other | 6.0 | 1.1–10.8 | 45.5 | 2.0–8.9 | 4.6 | 1.1–8.0 |
| Body Mass Index – BMI (kg/m ²) | | | | | | |
| Underweight/normal (BMI<25) | 4.2 | 3.4–4.1 | 4.2 | 3.4–5.0 | 4.3 | 3.4–5.2 |
| Overweight (25≤BMI<30) | 7.5 | 6.2–8.7 | 7.5 | 6.2–8.7 | 7.2 | 6.0–8.4 |
| Obesity (BMI≥30) | 13.5 | 11.4–15.5 | 13.5 | 11.5–15.4 | 13.5 | 11.6–15.5 |
| Health Insurance | | | | | | |
| Yes | 7.1 | 5.8–8.4 | 7.1 | 5.8–8.3 | 7.0 | 5.8–8.3 |
| No | 7.6 | 6.8–8.5 | 7.7 | 6.8–8.5 | 7.6 | 6.7–8.4 |

systems obtained from these methods demonstrated performance comparable to the weighting system available with the Laboratory Test data. Notably, it is important to highlight that the estimates derived from the proposed weighting systems are generalizable to the target population of the 2013 PNS: the Brazilian adult population living in private households.

Another aspect to emphasize is that the two proposed calibration methods demonstrated greater precision for the estimates considered in this study. A possible explanation for this behavior is that the methods incorporated important aspects, albeit presumed, of the sampling survey design, which contributed to the accurate calculation of the variance estimates for the parameters of interest.

When using measures to assess the performance of the calibration methods, the two suggested weighting systems performed well, showing an advantage over the raking-based weights. Some previous studies have reached similar conclusions¹⁹⁻²². However, it is important to note that measures comparing the calibrated weights to the “pre-calibrated” weights reveal a substantial difference between the two sets of weights. This behavior is likely due to the assumption of a simple random sampling design for the first-phase sample of the 2013 PNS data from the Laboratory Tests subsample, while the post-strata were constructed based on projections of the Brazilian population.

As observed, the Laboratory Tests subsample of the 2013 PNS could be considered the result of a two-phase sampling design¹⁶. However, the selection probabilities (or, equivalently, the basic sampling weights) were not made available in the Laboratory Tests microdata file, limiting the potential benefits of a two-phase design. The availability of the basic design weights for the Laboratory Tests subsample would enable statisticians and epidemiologists to construct calibrated weighting systems, thereby improving the performance of calibration estimators.

Regarding the estimation of population totals (*e.g.*, the total number of individuals with diabetes), the two methods proposed in this article provide estimates for the target population of the 2013 PNS, allowing the results to be generalized to the Brazilian adult population. The weighting system provided with the data from the Laboratory Tests subsample produced estimates specific to the 2013 PNS sample. While it is possible to obtain estimates for the target population of the PNS indirectly by multiplying the sample proportion by the population size, it is preferable for the methods, along with the appropriate software, to provide these estimates directly to minimize errors in the interpretation of the results.

The approach adopted in this article involved weighting through calibration methods. Some limitations of this approach are inherent to the method itself, such as the application of weights to more complex estimates, like regression coefficients; the difficulty in assessing the standard errors of weighted estimates, and the decisions involved in constructing weighting systems²³. However, it is recommended that weighted estimates be used in preference to unweighted estimates, particularly when working with data from population surveys that employ complex sampling designs. Another limitation of this work was the assumption of certain aspects of the design used to collect data from the 2013 PNS Laboratory Tests subsample. As emphasized in the literature, analyses should account for the relevant aspects of the sampling plan¹⁴. Making such details available for the PNS Laboratory Tests subsample would enhance researchers' ability to estimate population quantities with greater precision.

We conclude with two suggestions for survey designers and managers of the Brazilian National Health Surveys that will be used by other researchers. First, all weights related to the design of the PNS should be thoroughly documented and included in the public data files. This would enable users to construct estimates based on either the basic or calibrated weights, utilizing their own auxiliary variables. Accurate precision of the estimates could then be achieved through the proper use of data analysis software designed for complex sampling plans.

Our second recommendation pertains to the keying of information obtained from the sample and subsample (of Laboratory Tests) of the 2013 PNS. Properly documenting this information would facilitate the application of methods such as data imputation to obtain more accurate estimates. Given the significant relevance of these data for research in epidemiology and public health, it is essential that the most appropriate methods be employed in their analysis.

REFERENCES

1. Malta DC, Stopa SR, Szwarcwald CL, Gomes NL, Silva Júnior JB, Reis AAC. Surveillance and monitoring of major chronic diseases in Brazil - National Health Survey, 2013. *Rev Bras Epidemiol.* 2015; 18(Supl. 2): 3-16. <https://doi.org/10.1590/1980-5497201500060002>
2. IBGE. Pesquisa nacional de saúde: 2013: percepção do estado de saúde, estilos de vida e doenças crônicas: Brasil, grandes regiões e unidades da federação [Internet]. Rio de Janeiro: IBGE; 2014 [cited on Sep 21, 2022]. Available at: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?id=291110&view=detalhes>
3. Machado IE, Malta DC, Bacal NS, Rosenfeld LGM. Prevalence of anemia in Brazilian adults and elderly. *Rev Bras Epidemiol.* 2019; 22(Supl. 2): E190008.Supl.2. <https://doi.org/10.1590/1980-549720190008.supl.2>
4. Malta DC, Szwarcwald CL, Machado IE, Pereira CA, Figueiredo AW, Sá ACMGN, et al. Prevalence of altered total cholesterol and fractions in the Brazilian adult population: National Health Survey. *Rev Bras Epidemiol.* 2019; 22(Supl. 2): E190005.SUPL.2. <https://doi.org/10.1590/1980-549720190005.supl.2>
5. Malta DC, Machado IE, Pereira CA, Figueiredo AW, Aguiar LK de, Almeida WS, et al. Evaluation of renal function in the Brazilian adult population, according to laboratory criteria from the National Health Survey. *Rev Bras Epidemiol.* 2019; 22(Supl. 2): E190010.SUPL.2. <https://doi.org/10.1590/1980-549720190010.supl.2>
6. Malta DC, Duncan BB, Schmidt MI, Machado IE, Silva AG da, Bernal RTI, et al. Prevalence of diabetes mellitus as determined by glycated hemoglobin in the Brazilian adult population, National Health Survey. *Rev Bras Epidemiol.* 2019; 22(Supl. 2): E190006.SUPL.2. <https://doi.org/10.1590/1980-549720190006.supl.2>

7. dos Reis RCP, Duncan BB, Szwarzwald CL, Malta DC, Schmidt MI. Control of glucose, blood pressure, and cholesterol among adults with diabetes: the Brazilian National Health Survey. *J Clin Med*. 2021; 10(15): 3428. <https://doi.org/10.3390/jcm10153428>
8. Malta DC, Szwarzwald CL, Silva JBD. First results of laboratory analysis in the National Health Survey. *Rev Bras Epidemiol*. 2019; 22(Supl. 02): E190001.SUPL.2. <https://doi.org/10.1590/1980-549720190001.supl.2>
9. Szwarzwald CL, Malta DC, Souza PRB de, Almeida WS de, 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(Supl. 2): E190004.SUPL.2. <https://doi.org/10.1590/1980-549720190004.supl.2>
10. Deville JC, Sarndal CE, Sautory O. Generalized raking procedures in survey sampling. *J Am Stat Assoc*. 1993; 88(423): 1013-20. <https://doi.org/10.2307/2290793>
11. Amorim G, Tao R, Lotspeich S, Shaw PA, Lumley T, Shepherd BE. Two-phase sampling designs for data validation in settings with covariate measurement error and continuous outcome. *J R Stat Soc Ser A Stat Soc*. 2021; 184(4): 1368-89. <https://doi.org/10.1111/rssa.12689>
12. Neyman J. Contribution to the theory of sampling human populations. *J Am Stat Assoc*. 1938; 33(201): 101-16. <https://doi.org/10.2307/2279117>
13. Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sampling design. *Am J Public Health*. 1991; 81(9): 1166-73. <https://doi.org/10.2105/AJPH.81.9.1166>
14. Silva, PLN, Pessoa DGC, Lila MF. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. *Ciênc Saúde Coletiva*. 2002; 7(4): 659-70. <https://doi.org/10.1590/S1413-81232002000400005>
15. Silva PLN. Calibration estimation: when and why, how much and how [Internet]. Rio de Janeiro: IBGE; 2004 [cited on Sep 21, 2022]. Available at: <https://biblioteca.ibge.gov.br/biblioteca-catalogo?id=281040&view=detalhes>
16. Haziza D, Beaumont JF. Construction of weights in surveys: a review. *Statist Sci*. 2017; 32(2): 206-26. <https://doi.org/10.1214/16-STS608>
17. R Core Team. R: A language and environment for statistical computing [Internet]. Viena: R Foundation for Statistical Computing; 2021 [cited on May 16, 2021]. Available at: <https://www.r-project.org/>
18. Lumley T. Analysis of complex survey samples. *J Stat Softw*. 2004; 9(8): 1-19. <https://doi.org/10.18637/jss.v009.i08>
19. Djerf K. Effects of post-stratification on the estimates of the finnish labour force survey. *J Off Stat*. 1997; 13(1): 29-39.
20. Ruiz CMM, Silva PLN. Explorando alternativas para a calibração dos pesos amostrais da Pesquisa Nacional por Amostra de Domicílios. In: Proceedings of the Conference Name. Lima, Peru; 2014.
21. Tu SH. A comparison of propensity score sub-classification and other calibration methods based on a telephone sample to estimate internet usage. *Taiwanese J Sociol*. 2015; 56: 115-50. [https://doi.org/10.6786/TJS.201506_\(56\).0003](https://doi.org/10.6786/TJS.201506_(56).0003)
22. 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. <https://doi.org/10.5123/S1679-49742017000400003>
23. Gelman A. Struggles with survey weighting and regression modeling. *Statist Sci*. 2007; 22(2): 153-64. <https://doi.org/10.1214/088342306000000691>

RESUMO

Objetivo: Este estudo visa comparar métodos de calibração de pesos na subamostra de Exames Laboratoriais da Pesquisa Nacional de Saúde (PNS) 2013, buscando avaliar sua representatividade e precisão. **Métodos:** Duas propostas alternativas de construção de pesos calibrados foram realizadas com base nos métodos de pós-estratificação e *raking*. Uma comparação entre os pesos fornecidos para a subamostra de Exames Laboratoriais e os dois pesos sugeridos foi realizada por meio de estimativas de parâmetros de interesse utilizando os dados da subamostra da PNS 2013. Além disso, foram utilizadas sete medidas para avaliar o desempenho dos sistemas de pesos propostos. **Resultados:** Os pesos de pós-estratificação e *raking* alternativos produziram estimativas generalizáveis para a população-alvo da PNS 2013, enquanto os pesos originais não. Os métodos alternativos apresentaram desempenho similar ao método original, com pequena vantagem para o *raking* em algumas medidas de avaliação. **Conclusão:** Recomenda-se que os pesos básicos de delineamento sejam documentados e incluídos nos arquivos de dados de uso público da PNS. Além disso, sugere-se o chaveamento das informações entre a amostra e a subamostra da PNS 2013 para possibilitar a exploração de métodos como a imputação de dados, visando obter estimativas mais precisas e representativas. Essas melhorias são essenciais para garantir a qualidade e utilidade dos dados da PNS em estudos epidemiológicos e de saúde coletiva.

Palavras-chave: Estimativas populacionais. Estudos de amostragem. Estudos epidemiológicos. Estudos populacionais em saúde pública. Métodos e procedimentos estatísticos.

AUTHORS' CONTRIBUTIONS: Souza, J. S.: Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft, Writing – Review & Editing. Barbian, M. H.: Conceptualization, Supervision, Writing – Original Draft, Writing – Review & Editing. Reis, R. C. P.: Conceptualization, Methodology, Supervision, Writing – Original Draft, Writing – Review & Editing.

FUNDING: none.