

Sampling plan of the *Brazilian National Survey on Child Nutrition (ENANI-2019)*: a population-based household survey

Plano amostral do *Estudo Nacional de Alimentação e Nutrição Infantil (ENANI-2019)*: inquérito domiciliar de base populacional

Plan de muestreo del *Estudio Nacional de Alimentación y Nutrición Infantil (ENANI-2019)*: encuesta domiciliar de base poblacional

Maurício Teixeira Leite de Vasconcellos ^{1,2}
Pedro Luis do Nascimento Silva ^{1,2}
Inês Rugani Ribeiro de Castro ³
Cristiano Siqueira Boccolini ⁴
Nadya Helena Alves-Santos ⁵
Gilberto Kac ⁵

doi: 10.1590/0102-311X00037221

Abstract

The article describes methodological aspects in defining the study population, sampling plan, and sample weighting and calibration of effective sample of the Brazilian National Survey on Child Nutrition (ENANI-2019). This population-based household survey assessed breastfeeding and dietary intake, anthropometric assessment of nutritional status, and micronutrient deficiencies by blood biomarkers in children under five years of age. The data were obtained with a probability sample, with stratification by the five geographic regions in the country and clustering by census enumeration areas (CEAs). The sample was calculated at 15,000 households distributed in 1,500 CEAs, with 300 allocated in each of Brazil's five major geographic regions and 10 eligible households per CEA, sampled using inverse sampling. The required population parameters were thus estimated to reach the study's objectives. The basic sampling design weights were calculated as the inverse probabilities of the households' inclusion in the study. Imputation was used to compensate for non-response to items in the target variables, except for data on the blood biomarkers. Finally, calibration used population totals of children in 60 post-strata, defined by cross-classification of the following variables: major geographic region, sex, and age. The final sample included 14,558 children residing in 12,524 households, distributed in 1,382 CEAs in the 26 states of Brazil and the Federal District. The data from the ENANI-2019 survey will support strategies for the promotion and implementation of public policies for children under five years of age.

Infant; Preschool Child; Statistical Models; Sampling Studies; Methods

¹ Escola Nacional de Ciências Estatísticas, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brasil.
² Science – Sociedade para o Desenvolvimento da Pesquisa Científica, Rio de Janeiro, Brasil.
³ Instituto de Nutrição, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.
⁴ Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.
⁵ Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

Correspondence

G. Kac
Departamento de Nutrição Social e Aplicada, Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro.
Av. Brigadeiro Trompowsky s/n, Bloco J, 2º andar, Rio de Janeiro, RJ 21941-590, Brasil.
gilberto.kac@gmail.com



Introduction

The Ministry of Health funded the *Brazilian National Survey on Child Nutrition* (ENANI-2019) (call for projects CNPq/MS/SCTIE/DECIT/SAS/DAB/CGAN n. 11/2017). ENANI-2019 is structured in three domains: assessment of breastfeeding and dietary intake; anthropometric assessment of nutritional status; and assessment of micronutrient deficiencies in children under five years of age, by major geographic region, sex, and age group.

The data were obtained with a probabilistic household sample survey, with geographic stratification and clustering by census enumeration areas (CEAs), conducted with sampling methods such as those adopted by the official statistical institutes in their human population surveys¹. This allowed the ENANI-2019 to reproducibly and scientifically estimate the required population parameters to reach its objectives. The basic idea when sampling human populations consists of sampling them through the households, grouped in turn in CEAs, which are grouped in turn according to the situation (urban versus rural) in subdistricts, districts, municipalities, and so on. The basis is the concept of household and resident (the latter to ensure that individuals with more than one residential address would not be more likely to enter the sample) and selection of areas.

ENANI-2019 provides a unique opportunity to elucidate the various aspects of the nutritional assessment of children and to support public health policies for this vulnerable age group. Thus, the manuscript aims to describe methodological aspects in defining of the study population, sampling plan, sample weighting, and effective sample of the ENANI-2019.

Study population

The study population for ENANI-2019 was defined as the set of children under five years of age residing in permanent private households throughout Brazil with at least one child under five years of age on the date of the survey interview. Therefore, the study population did not include: (1) children residing in collective households (hotels, boarding houses, orphanages, shelters, detention centers, barracks, hospitals, etc.), improvised private households, and permanent private households without children; (2) indigenous children living in villages; (3) foreign children living in households where Portuguese was not spoken; and (4) children with conditions that prevented them from undergoing anthropometric measurement.

Ethical aspects

The Institutional Review Board of the Clementino Fraga Filho University Hospital of the Federal University of Rio de Janeiro (UFRJ) approved the study under number CAAE 89798718.7.0000.5257. Data were collected after the child's parents or guardians signed two copies of the free and informed consent form. The methods used in the development of ENANI-2019 have been described in detail in specific publications^{2,3,4,5}.

Sampling plan

The sampling plan of ENANI-2019 used stratification and clustering, incorporating two or three selection stages. The population's stratification for sampling purposes was guided by the study's objectives and the definition of the five major geographic regions in Brazilian territory as target domains for publication of results.

The primary sampling units (PSUs) were the municipalities or the CEAs, and the elementary sampling units were always the households. In each selected household, all residents were enrolled, and the study's target data were recorded for all the resident children under five years of age.

Stratification

Strata were formed through the allocation of Brazilian municipalities (according to the territorial base used by the Brazilian Institute of Geography and Statistics – IBGE, in the population estimates for July 1, 2016) ⁶ in two blocks: (1) each of the state capitals plus the Federal District (27 strata) and each of the 20 municipalities with more than 500,000 inhabitants (20 strata) and (2) the other municipalities in each major geographic region (5 strata) (Table 1). Therefore, all the state capitals and municipalities with large populations (> 500,000 inhabitants) defined as strata in block 1 were included in the sample with certainty and are not primary sampling units but selection strata.

The data for the total population and the population of children under five years of age were estimated for July 1, 2016, for each of the 5,570 Brazilian municipalities using the linear trend method ⁷, the same used by the IBGE in the elaboration of the population estimates used by the Federal Accounts Court to determine their share of the participatory fund for municipalities ⁶. Table 1 shows the estimates obtained per selection stratum.

In the 47 strata formed by each of the municipalities included with certainty in the sample (block 1), the PSU was the CEA (IBGE), and the secondary sampling unit (SSU) was the eligible household (with children from the study population). In the other strata (block 2), the PSU was the municipality, the SSU was the CEA, and the tertiary sampling unit (TSU) was the eligible household.

Calculation of the sample size

Calculation of the sample size was guided by the project's budget parameters, the blood sample collection logistics, and the experience with similar surveys conducted by the Society for the Development of Scientific Research (Science).

Considering the target domain (major geographic region), the minimum proportion was specified as $P_{\min} = 2\%$. The estimated relative margin of error should be a maximum of $d_R = 35\%$, with a confidence coefficient of $(1-\alpha) = 95\%$. According to Cochran ⁸ and assuming simple random sampling without replacement (SRS), the necessary sample size to estimate proportions equal to or greater than P_{\min} with a relative error no greater than d_R with a level of confidence $1-\alpha$ is calculated by:

$$n_{\text{SRS}} = \frac{z_{\alpha/2}^2}{d_r^2} \times \frac{1-P_{\min}}{P_{\min}} \quad (1)$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Since the sample design is complex (stratified and clustered), it is necessary to consider the design effect on calculating the sample size. Pessoa & Silva ⁹ recommend multiplying the sample size obtained by the Expression 1 by an estimate of the design effect (deff) referring to the key survey variable. A deff of 1.95 was set for calculating the sample size since there were no data on deff from previous household surveys on the topic. However, selecting an arbitrary value for deff greater than one is preferable to the alternative of not making any adjustment to the sample size for the expected effects of clustering with the sampling design adopted. Data from the study showed that the deff for the estimates of the proportion of children that did not receive breastmilk on the eve of the interview by sex and age group varied from 2.3 to 5.7, and the proportions varied from 12.8% (girls under six months of age) to 97.4% (girls four years old). Similar ranges for deff were observed for estimates by sex and age concerning the children's average weight (2.9 to 7.3) and height (2.7 to 6.6). These results suggest that the value used in calculating the sample size was small. For future calculation of samples from the same population, the data from this study can be used to estimate deff values for other key survey variables.

The sample size of households to be interviewed for each major geographic region was thus calculated by the Expression 2:

$$n = \text{deff} \times n_{\text{SRS}} = \text{deff} \times \frac{z_{\alpha/2}^2}{d_r^2} \times \frac{1-P_{\min}}{P_{\min}} = 1,95 \times \frac{1.96^2}{0.35^2} \times \frac{1-0.02}{0.02} = 2,998 \quad (2)$$

Since there are five estimation domains, the total sample size was calculated at 14,990 (= 5 x 2,998) households.

It was also determined that ten eligible households would be interviewed for each selected CEA, which led to a sample of $m = 1,500$ CEAs, 300 in each major geographic region. This definition also resulted from the accumulated experience with samples from household surveys by the Science team and from the evidence of the effects of CEA sample size on the estimates' precision and data collection costs. The number ten could be considered small compared to that adopted in other household surveys, such as the *Brazilian Continuous National Household Sample Survey* (PNAD Contínua), which selects 14 households (eligible or not) per CEA¹⁰. However, in ENANI-2019, it would be difficult to reach 14 eligible households per CEA. Based on an average CEA size of 300 households, considering that the proportion of children under five years of age in 2016 was estimated at 7.2%, besides assuming that each household would have a maximum of one child under five years of age, there would be an expected number of 21.6 eligible households per CEA. Since the CEA sizes vary considerably (above and below the average number of 300 households), and since the above estimate is optimistic, dependent on the hypothesis of one eligible child per household, the target sample size of ten eligible households per CEA appeared reasonable and was adopted.

Allocation of the CEA sample in the selection strata

There are various ways of allocating the CEA sample size among the selection strata. At one extreme, there is proportional allocation, which ensures that the sample size in each stratum is proportional to its population, with the disadvantage of concentrating the sample in the strata with the largest population. The other extreme is equal allocation (as among the major geographic regions), which ensures that the margin of error (or sampling precision) is similar across strata, but only recommendable when the strata are estimation domains. Finally, between these extremes, there is power allocation, which ensures a certain proportionality between the sample size in the stratum and a power p ($0 < p < 1$) of its population. The larger the power p , the more closely power allocation approximates proportional allocation, and the smaller the power p , the closer it gets to equal allocation.

Expression 3 presents the form of power allocation¹¹ used to define the sample size of CEAs for each selection stratum h within each major geographic region:

$$m_h = 300 \times \frac{\text{pop}_h^p}{\sum_h \text{pop}_h^p} \quad (3)$$

where pop_h represents the population under five years of age in stratum h , estimated for July 1, 2016, as previously indicated in Table 1.

The Science experience in household sample surveys led to the use of a power allocation with $p = 1/3$, which displays a certain proportionality with the stratum's population, without allowing excessive concentration in the more highly populated strata.

For the strata of "other municipalities" in the five major geographic regions, the definition of the number of CEAs to select in each municipality determined the number of municipalities to select in each of these strata. The decision was to select five CEAs per municipality in all the major geographic regions, except in the North, where eight CEAs were selected per municipality. This larger number of CEAs per municipality in the stratum "other municipalities" in the North of Brazil allowed reducing the number of selected municipalities. The North of Brazil has huge difficulties involving access and traveling time from the municipalities to their respective state capitals. In most municipalities, the traveling time could prevent taking blood samples and increasing the study's costs. Table 2 shows the planned sample size for CEAs and households.

Table 1

Projection of the Brazilian population under five years of age according to major geographic regions and sample selection strata. *Brazilian National Survey on Child Nutrition* (ENANI-2019).

Major geographic regions and sample selection strata	Total population	Pop _h under five years of age	Cubic root of pop _h	Size of sample of CEAs
Brazil	206,081,432	14,948,464		
North	17,707,783	1,741,528	409.4	300
Porto Velho	511,219	42,184	34.8	25
Rio Branco	377,057	34,093	32.4	24
Manaus	2,094,391	188,889	57.4	42
Boa Vista	326,419	30,983	31.4	23
Belém	1,446,042	103,253	46.9	34
Macapá	465,495	45,847	35.8	26
Palmas	279,856	24,863	29.2	21
Ananindeua	510,834	40,450	34.3	25
Other municipalities in the North *	11,729,105	1,230,966	107.2	80
Northeast	56,915,936	4,539,260	623.7	300
São Luís	1,082,935	80,557	43.2	21
Teresina	847,430	61,522	39.5	19
Fortaleza	2,609,716	179,659	56.4	27
Natal	877,662	57,825	38.7	19
João Pessoa	801,718	55,731	38.2	18
Recife	1,625,583	102,381	46.8	22
Maceió	1,021,709	78,189	42.8	20
Aracaju	641,523	45,685	35.7	17
Salvador	2,938,092	181,478	56.6	27
Jaboatão dos Guararapes	691,125	49,855	36.8	18
Feira de Santana	622,639	45,821	35.8	17
Other municipalities in the Northeast **	43,155,804	3,600,557	153.3	75
Southeast	86,356,952	5,581,236	913.9	300
Belo Horizonte	2,513,451	140,968	52.0	17
Vitória	359,555	21,497	27.8	9
Rio de Janeiro	6,498,837	374,307	72.1	24
São Paulo	12,038,175	760,498	91.3	30
Contagem	653,800	42,289	34.8	12
Juiz de Fora	559,636	31,308	31.5	10
Uberlândia	669,672	43,468	35.2	12
Duque de Caxias	886,917	61,270	39.4	13
Nova Iguaçu	797,435	53,830	37.8	12
São Gonçalo	1,044,058	61,262	39.4	13
Campinas	1,173,370	69,386	41.1	14
Guarulhos	1,337,087	98,157	46.1	15
Osasco	696,382	47,133	36.1	12
Ribeirão Preto	674,405	39,935	34.2	11
Santo André	712,749	41,414	34.6	11
São Bernardo do Campo	822,242	52,860	37.5	12
São José dos Campos	695,992	46,612	36.0	12
Sorocaba	652,481	40,608	34.4	11
Other municipalities in the Southeast ***	53,570,708	3,554,434	152.6	50

(continues)

Table 1 (continued)

Major geographic regions and sample selection strata	Total population	Pop _h under five years of age	Cubic root of pop _h	Size of sample of CEAs
South	29,439,773	1,899,647	305.3	300
Curitiba	1,893,997	116,672	48.9	48
Florianópolis	477,798	25,905	29.6	29
Porto Alegre	1,481,019	82,625	43.6	43
Londrina	553,393	35,100	32.7	32
Joinville	569,645	37,905	33.6	33
Other municipalities in the South #	24,463,921	1,601,440	117.0	115
Central West	15,660,988	1,186,793	305.5	300
Campo Grande	863,982	62,549	39.7	39
Cuiabá	585,367	43,075	35.1	34
Goiânia	1,448,639	93,978	45.5	44
Brasília	2,977,216	219,027	60.3	59
Aparecida de Goiânia	532,135	43,638	35.2	34
Other municipalities in the Central West ##	9,253,649	724,526	89.8	90

CEAs: census enumeration areas; pop_h: population under five years of age in stratum h.

Source: *Brazilian National Survey on Child Nutrition* (ENANI-2019).

* Other municipalities in the North: Ariquemes and Ji-Paraná (Rondônia State); Itacoatiara and Manacapuru (Amazonas States); Abaetetuba, Bragança, Castanhal, and Marituba (Pará State); Araguaína (Tocantins State). In Barcarena (Pará State), 8 CEAs and 80 households were selected, but there was no data collection due to the interruption of collection;

** Other municipalities in the Northeast: Caxias and Imperatriz (Maranhão State); Parnaíba (Piauí State); Caucaia and Maracanaú (Ceará State); Parnamirim (Rio Grande do Norte State); Campina Grande (Paraíba State); Olinda and Paulista (Pernambuco State); Arapiraca (Alagoas State); Nossa Senhora do Socorro (Sergipe State); Camaçari, Juazeiro, Lauro de Freitas, and Simões Filho (Bahia State);

*** Other municipalities in the Southeast: Betim, Montes Claros, and Ribeirão das Neves (Minas Gerais State), Serra (Espírito Santo State); Niterói (Rio de Janeiro); Araraquara, Diadema, Itaquaquecetuba, Pindamonhangaba, and São Vicente (São Paulo State);

Other municipalities in the South: Araucária, Colombo, Foz do Iguaçu, Guarapuava, Maringá, Pinhais, Ponta Grossa, and São José dos Pinhais (Paraná State); Blumenau, Camboriú, Criciúma, Jaraguá do Sul, Palhoça, and São José (Santa Catarina State); Alvorada, Canoas, Caxias do Sul, Gravataí, Novo Hamburgo, Rio Grande, São Leopoldo, Sapucaia do Sul, and Viamão (Rio Grande do Sul State);

Other municipalities in the Central West: Corumbá, Dourados, Nova Andradina, and Três Lagoas (Mato Grosso do Sul State); Cáceres, Lucas do Rio Verde, Primavera do Leste, Rondonópolis, and Várzea Grande (Mato Grosso State); Águas Lindas de Goiás, Anápolis, Cidade Ocidental, Formosa, Jataí (substituído por Luziânia), Novo Gama, Planaltina, Santo Antônio do Descoberto, and Valparaíso de Goiás (Goiás State).

Sample selection methods in the various stages

When the municipality was the PSU (block 2, strata “other municipalities” of the major geographic regions), its selection was performed with systematic sampling with probabilities proportional to size (PPS), used as a measure of the size of the population under five years of age in the municipality, estimated for July 1, 2016.

Since lower-income CEAs were expected to have more eligible children than the higher-income CEAs, care was taken for the sample to cover the range of the population’s income in the selected municipalities, guaranteeing different children’s feeding patterns in the study population. Thus, before the CEAs’ selection, an additional stratification was performed, based on quartiles of the distribution of the average head-of-household’s income in each CEA, according to the *2010 Population Census*. Next, the numbers of CEAs to be selected in each income stratum were allocated. Finally, within each municipality and income stratum, CEAs were selected by Pareto’s PPS sampling^{12,13}. The size measure for CEA sampling was the number of children under five years of age in the CEA, based on the *2010 Population Census*, the most recent source of information available per CEA at the time of the survey.

Table 2

Size of sample of census enumeration areas and households for Brazil and according to major geographic regions, selection strata, and municipalities.
Brazilian National Survey on Child Nutrition (ENANI-2019).

Major geographic regions, selection strata, and municipalities in the sample	Sample of CEAs			Sample of households		
	Selected (n)	Effective (n)	Effective (%)	Selected (n)	Effective (n)	Effective (%)
Brazil	1,500	1,382	92.1	15,000	12,524	83.6
North	300	233	77.7	3,000	2,264	75.5
Porto Velho	25	25	100.0	250	250	100.0
Rio Branco	24	24	100.0	240	240	100.0
Manaus	42	42	100.0	420	422	100.0
Boa Vista	23	23	100.0	230	216	93.9
Belém	34	16	47.1	340	164	48.2
Macapá	26	26	100.0	260	260	100.0
Palmas	21	21	100.0	210	199	94.8
Ananindeua	25	12	48.0	250	97	38.8
10 other municipalities in the North *	80	44	55.0	800	416	52.0
Northeast	300	259	86.3	3,000	2,430	81.0
São Luís	21	17	81.0	210	150	71.4
Teresina	19	15	78.9	190	136	71.6
Fortaleza	27	14	51.9	270	116	43.0
Natal	19	15	78.9	190	144	75.8
João Pessoa	18	18	100.0	180	177	98.3
Recife	22	19	86.4	220	177	80.5
Maceió	20	19	95.0	200	203	100.0
Aracaju	17	17	100.0	170	155	91.2
Salvador	27	27	100.0	270	266	98.5
Jaboatão dos Guararapes	18	10	55.6	180	99	55.0
Feira de Santana	17	17	100.0	170	170	100.0
15 other municipalities in the Northeast **	75	71	94.7	750	637	84.9
Southeast	300	299	99.7	3,000	2,629	87.6
Belo Horizonte	17	17	100.0	170	178	100.0
Vitória	9	9	100.0	90	79	87.8
Rio de Janeiro	24	24	100.0	240	223	92.9
São Paulo	30	30	100.0	300	251	83.7
Contagem	12	12	100.0	120	121	100.0
Juiz de Fora	10	10	100.0	100	101	100.0
Uberlândia	12	13	108.3	120	129	100.0
Duque de Caxias	13	13	100.0	130	127	97.7
Nova Iguaçu	12	12	100.0	120	107	89.2
São Gonçalo	13	13	100.0	130	112	86.2
Campinas	14	14	100.0	140	107	76.4
Guarulhos	15	14	93.3	150	115	76.7
Osasco	12	12	100.0	120	86	71.7
Ribeirão Preto	11	11	100.0	110	91	82.7
Santo André	11	9	81.8	110	69	62.7
São Bernardo do Campo	12	12	100.0	120	82	68.3
São José dos Campos	12	13	108.3	120	91	75.8
Sorocaba	11	11	100.0	110	94	85.5
10 other municipalities in the Southeast ***	50	50	100.0	500	466	93.2

(continues)

Table 2 (continued)

Major geographic regions, selection strata, and municipalities in the sample	Sample of CEAs			Sample of households		
	Selected (n)	Effective (n)	Effective (%)	Selected (n)	Effective (n)	Effective (%)
South	300	300	100.0	3,000	2,518	83.9
Curitiba	48	48	100.0	480	296	61.7
Florianópolis	29	29	100.0	290	232	80.0
Porto Alegre	43	43	100.0	430	355	82.6
Londrina	32	32	100.0	320	282	88.1
Joinville	33	33	100.0	330	315	95.5
23 other municipalities in the South #	115	115	100.0	1,150	1,038	90.3
Central West	300	291	97.0	3,000	2,683	89.4
Campo Grande	39	39	100.0	390	375	96.2
Cuiabá	34	34	100.0	340	288	84.7
Goiânia	44	38	86.4	440	336	76.4
Brasília	59	59	100.0	590	554	93.9
Aparecida de Goiânia	34	31	91.2	340	278	81.8
18 other municipalities in the Central West ##	90	90	100.0	900	852	94.7

CEAs: census enumeration areas.

Source: *Brazilian National Survey on Child Nutrition* (ENANI-2019).

* Other municipalities in the North: Ariquemes and Ji-Paraná (Rondônia State); Itacoatiara and Manacapuru (Amazonas States); Abaetetuba, Bragança, Castanhal, and Marituba (Pará State); Araguaína (Tocantins State). In Barcarena (Pará State), 8 CEAs and 80 households were selected, but there was no data collection due to the interruption of collection;

** Other municipalities in the Northeast: Caxias and Imperatriz (Maranhão State); Parnaíba (Piauí State); Caucaia and Maracanaú (Ceará State); Parnamirim (Rio Grande do Norte State); Campina Grande (Paraíba State); Olinda and Paulista (Pernambuco State); Arapiraca (Alagoas State); Nossa Senhora do Socorro (Sergipe State); Camaçari, Juazeiro, Lauro de Freitas, and Simões Filho (Bahia State);

*** Other municipalities in the Southeast: Betim, Montes Claros, and Ribeirão das Neves (Minas Gerais State), Serra (Espírito Santo State); Niterói (Rio de Janeiro); Araraquara, Diadema, Itaquaquecetuba, Pindamonhangaba, and São Vicente (São Paulo State);

Other municipalities in the South: Araucária, Colombo, Foz do Iguaçu, Guarapuava, Maringá, Pinhais, Ponta Grossa, and São José dos Pinhais (Paraná State); Blumenau, Camboriú, Criciúma, Jaraguá do Sul, Palhoça, and São José (Santa Catarina State); Alvorada, Canoas, Caxias do Sul, Gravataí, Novo Hamburgo, Rio Grande, São Leopoldo, Sapucaia do Sul, and Viamão (Rio Grande do Sul State);

Other municipalities in the Central West: Corumbá, Dourados, Nova Andradina, and Três Lagoas (Mato Grosso do Sul State); Cáceres, Lucas do Rio Verde, Primavera do Leste, Rondonópolis, and Várzea Grande (Mato Grosso State); Águas Lindas de Goiás, Anápolis, Cidade Ocidental, Formosa, Jataí (substituído por Luziânia), Novo Gama, Planaltina, Santo Antônio do Descoberto, and Valparaíso de Goiás (Goiás State).

The adopted selection scheme prioritized the CEAs' stratification by income and did not consider stratification by the urban-versus-rural situation. In this sense, the participation of rural CEAs in the sample would be approximately proportional to that observed in the municipalities. However, due to the logistic difficulty of household blood sample collection and the samples' transportation to the local laboratory for processing, 46 rural CEAs which were more than two hours' travel time from the municipal center (time interval greater than allowed by the study's protocol for collection and transportation of blood samples) were replaced by closer CEAs. Later, as a function of the blood sample logistics, another 11 rural CEAs were also replaced during data collection. The implication of these operational restrictions of the blood sample collection and processing was the small presence of rural CEAs in the sample (only 32 rural CEAs among the 1,392 CEAs with data collected), resulting in estimates with a low level of precision for this setting.

The selection of eligible households within each selected CEA used inverse sampling^{14,15,16} during the data collection operation.

The collection began with the identification of selected CEAs (maps, descriptions, limits, and areas of exclusion, and the list of addresses in the National Registry of Addresses for Statistical Purposes – CNEFE, all available on the IBGE website). This was followed by updating the registry of addresses per CEA via the Census Tract Address Updating System (SAES), an app developed by Science and

operated via the mobile data collection device (MDC). At this time, the interviewers canvassed each selected CEA, conducting the confirmations, corrections, inclusions, and exclusions of addresses for the buildings found along the way. Each identified building was classified as either a household (private or collective) or an establishment.

In each selected CEA, having concluded the update of the address registry, the SAES numbered the addresses classified as private households (PH) sequentially, starting with one, according to the order of the path taken by the interviewer in the CEA. Then, selection tables were used to generate a random permutation of the PH by blocks of ten for the CEA's addresses (in each block, the ten PH were placed in the order of the path to facilitate the interviewer's movement). The interviewer's MDC displayed the first 20 addresses (in random order) to be visited to define the household's eligibility and obtain (if eligible) the family's consent to conduct the interview.

For each selected household in which the visit and contact did not result in an interview (ineligible PH, vacant PH, refusal, etc.), the data control app installed in the MDC added a new address to the list of PH addresses to be visited. This procedure ended when ten complete interviews had been obtained in the CEA or when all PH in the CEA had been visited. Thus, in each eligible interviewed household, information was collected on all the resident children under five years of age.

Probabilistic sampling scheme

The probability of inclusion in the sample of municipality i in stratum h , represented by $P(M_{hi})$, depends on it being included with certainty in the sample (making it a selection stratum) or on it having been a PSU in one of the "other municipalities" strata, as indicated in Expression 4:

$$P(M_{hi}) = \begin{cases} t_h \times \frac{\text{pop}_{hi}}{\sum_{i=1}^{T_h} \text{pop}_{hi}} & \text{if the municipality is a PSU} \\ 1 & \text{if it was included with certainty in the sample} \end{cases} \quad (4)$$

where pop_{hi} represents the population under five years of age in municipality M_{hi} , estimated for July 1, 2016, by the linear trend method ⁷; T_h represents the total number of municipalities in stratum h ; and t_h represents the size of the sample of municipalities in stratum h .

The conditional probability of inclusion in the sample of CEA j in municipality i in stratum h , conditioned by the selection or inclusion of municipality M_{hi} , represented by $P(S_{hij}|M_{hi})$, is indicated by the Expression 5:

$$P(S_{hij}|M_{hi}) = t_{hi} \times \frac{\text{dom}_{hij}}{\sum_{j=1}^{T_{hi}} \text{dom}_{hij}} \quad (5)$$

where dom_{hij} represents the number of households in CEA S_{hij} according to the *2010 Population Census*; T_{hi} represents the total number of CEAs in income stratum g , to which CEA j of municipality M_{hi} belongs; and t_{hi} represents the sample size of CEAs in income stratum g , to which CEA j of municipality M_{hi} belongs.

The sum of households in the CEAs was calculated in the set of CEAs belonging to each income stratum g in the municipality.

Thus, the probability of inclusion in the sample of CEA S_{hij} is expressed by:

$$P(S_{hij}) = P(M_{hi}) \times P(S_{hij}|M_{hi}) \quad (6)$$

In CEA S_{hij} , the conditional probability of interviewing household D_{hijk} is expressed by:

$$P(D_{hijk}|S_{hij}) = \frac{v_{hij}}{\text{dom}_{hij}^*} \times \frac{e_{hij} - 1}{v_{hij} - 1} \quad (7)$$

where dom_{hij}^* represents the number of private households in CEA S_{hij} obtained after updating the CEA's address registry, performed at the time of the study; v_{hij} is the total number of eligible private households selected and visited in CEA S_{hij} ; and e_{hij} represents the total number of households interviewed in CEA S_{hij} .

Thus, the probability of inclusion in the sample of household D_{hijk} is expressed by:

$$P(D_{hijk}) = P(M_{hi}) \times P(S_{hij}|M_{hi}) \times P(D_{hijk}|S_{hij}) \quad (8)$$

Sample weighting

The objective of this stage was to calculate and assign sampling weights to the children to allow estimating target parameters in the study population as a whole and for specific target analyses. Good sampling weights allow unbiased estimation for the target population parameters, compensating for non-response effects (of units) and estimating with efficiency (small margin of error). The guide proposed by Valliant & Dever¹⁷ was followed in the elaboration of the study's final sampling weights.

Since the study sample was stratified and clustered with unequal selection probabilities, it was necessary to calculate and use sampling weights for each of the households interviewed to allow unbiased estimation of target parameters in the population. The sampling weights were calculated in three or four stages, depending on the set of target information. The sampling weights were all calibrated to known population totals, seeking to correct typical biases in household samples and biases resulting from potential differential non-response or due to other difficulties faced while conducting the study.

Basic sampling weights were obtained in the first stage, corresponding to the inverse probabilities of inclusion of interviewed households. The basic weights for the households were calculated with the Expression (9):

$$d_{hijk} = 1/P(D_{hijk}) = 1/[P(M_{hi}) \times P(S_{hij}|M_{hi}) \times P(D_{hijk}|S_{hij})] \quad (9)$$

To better control the estimates' variability, the basic weights received upper truncation at 10,000 (that is, weights greater than 10,000 were trimmed to this value). This type of treatment is frequently used when the basic weights vary widely¹⁸.

The household's basic weight is applied to all the data obtained since no selection is made among the resident children. Therefore, the basic weight for all the children was set equal to their household's basic weight. The basic weight calculated with Expression 9 underwent two or three adjustment stages, depending on the set of target variables for the analysis.

The study's data collection was interrupted on March 17, 2020, due to the adoption of social distancing measures in response to the COVID-19 pandemic. Due to the interruption of data collection, the sample of CEAs was not collected in its entirety. The collection was concluded in most strata and PSUs, but in some, it did not occur (Table 2). In these strata and PSU, the weights calculated in Expression 9 were adjusted via multiplication by a given factor as indicated in the Expression 10:

$$w_{hijk} = d_{hijk} \times \left[\sum_{j \in A_{hi}} 1/P(S_{hij}) \right] / \left[\sum_{j \in C_{hi}} 1/P(S_{hij}) \right] \quad (10)$$

where A_{hi} is the set of CEAs sampled in PSU i in stratum h ; and C_{hi} is the set of CEAs collected in PSU i in stratum h .

To facilitate the presentation of the following stages in the weights' adjustment, we will change the notation, omitting the stratum, municipality, and CEA indices, which are unnecessary to facilitate understanding the expressions and calculations of the adjustment factors in the subsequent stages.

In the absence of non-response, the population total for a study variable y , denoted $Y = \sum_{k \in U} y_k$, could be estimated without bias using the Horvitz-Thompson estimator¹⁹, as given by the Expression 11:

$$\hat{Y}_{HT} = \sum_{k \in s} w_k y_k \quad (11)$$

where w_k is the adjusted basic weight of unit k , obtained by the Expression 10 at the end of stage 1, and s is the set of units in the sample.

Likewise, the population average $\bar{y} = \left(\sum_{k \in U} y_k \right) / N$, where N is the population's size, would be estimated using the Hájek estimator ²⁰, as shown in Expression 12:

$$\bar{y}_H = \sum_{k \in s} w_k y_k / \sum_{k \in s} w_k \quad (12)$$

As in any study, the ENANI-2019 sample presented both unit and item non-response that need to be compensated for in the analyses. Therefore, imputation was used to compensate for item non-response for most of the variables.

Laboratory analyses of blood samples showed unit non-response (lack of measures for all the biomarkers) and item non-response (lack of measures for some subset of biomarkers), as observed in Castro et al. ⁵. Considering the nature of these measurements, we decided that it was not possible to compensate for non-response in this set of variables using imputation. To compensate for this non-response, adjustments were made to the children's basic weights via the following steps.

Step 1: 25 groups of children were created with available responses for different subsets of variables in the data of blood biomarkers (Table 3). These groups were identified by dummy variables g_{kr} taking a value of 1 for available responses in group r for child k , and 0 otherwise, with r varying from 1 to 25.

Step 2: for each dummy variable with an available response in group r , a logistic regression model was fitted for the probability of response, defined in the Expression 13:

$$\Pr(g_{kr}) = 1 | x_k; \theta = \exp(x_k' \theta) / [1 + \exp(x_k' \theta)] \quad (13)$$

where x_k is a vector with selected predictive variables for explaining the propensity to respond, and θ is a vector of parameters to be estimated.

The fitted model was used to obtain estimates of response probabilities in group r , as shown in the Expression 14:

$$\hat{\Pr}(g_{kr}) = 1 | x_k; \theta = \exp(x_k' \theta) / [1 + \exp(x_k' \theta)] \quad (14)$$

The predictive variables considered in the fitted models in all the response groups were the same and are listed in Box 1. The selection of variables for inclusion in these models was based on a set of potentially relevant predictors for explaining the pattern of responses to groups of blood biomarkers, including characteristics of the region, households, and children. Next, initial models were fitted to the data, followed by step-by-step inclusion of new predictors until reaching the set of variables with significant and relevant main effects. No models were tested for interactions between predictors.

Step 3: for each group of records with an available response, the inverse estimated probability of response in the group was used as a factor to correct the child's basic weight, obtaining adjusted weights according to the Expression 15:

$$w_{kr} = w_k / \hat{\Pr}(g_{kr} = 1 | x_k; \hat{\theta}) \quad (15)$$

Since 25 groups of children were formed with different sets of available variables in the section on blood biomarkers, there are 25 sets of weights adjusted for non-response. In addition to the basic weight recommended for all the other analyses, each child will have a specific weight for each of these 25 sets. For each set of variables in which the child presents a complete response in all the variables, the corresponding weight is positive. It is null in case of non-response in at least one variable in the set of target variables. The data analysts will be responsible for selecting the adequate weights for the analyses that include blood biomarkers.

The final stage in the adjustment of the basic weights was calibration. The basic idea of calibration is to estimate factors f_k (called calibration factors) that multiply basic weights to generate the calibrated weights. These factors have the property of eliminating differences between estimates obtained with the calibrated weights and the corresponding population totals (known from other

Table 3

Number of children 6 to 59 months of age with results of blood biomarkers, according to age group. *Brazilian National Survey on Child Nutrition* (ENANI-2019).

Groups of blood biomarkers		Brazil		6-23 months		24-59 months	
Code	Description	N	%	N	%	N	%
1a	Hemoglobin + CRP + ferritin *	7,473	59.2	2,312	52.9	5,161	62.6
1b	Hemoglobin + CRP + ferritin **	7,293	57.8	2,254	51.6	5,039	61.1
1c	CRP + ferritin *	7,931	62.9	2,473	56.6	5,458	66.2
2	Hemoglobin + vitamin A	7,716	61.2	2,435	55.7	5,281	64.0
3	Hemoglobin + vitamin A + D + zinc	7,037	55.8	2,153	49.3	4,884	59.2
4	CRP + zinc	7,659	60.7	2,369	54.2	5,290	64.1
5	CRP + vitamin A	7,944	63.0	2,509	57.4	5,435	65.9
6	CRP + selenium	7,660	60.7	2,369	54.2	5,291	64.1
7	Folic acid + vitamin B12	7,419	58.8	2,267	51.9	5,151	62.4
8	CRP + zinc + selenium + vitamins A and E	7,506	59.5	2,322	53.1	5,184	62.8
9a	Folic acid, hemoglobin, ferritin, CRP, vitamins B12, D, A, B1, and B6, zinc and selenium *	6,023	47.7	1,779	40.7	4,244	51.4
9b	Folic acid, hemoglobin, ferritin, CRP, vitamins B12, D, A, B1 e B6, zinc, and selenium ***	6,532	51.8	1,956	44.8	4,576	55.5
10	Hemoglobin, folic acid, vitamin B12, ferritin, and CRP *	6,526	51.7	1,950	44.6	4,576	55.5
11	Hemoglobin, folic acid, vitamin B12, ferritin, and CRP #	6,556	52.0	1,965	45.0	4,591	55.7
12	Hemoglobin	8,187	64.9	2,626	60.1	5,561	67.4
13	Folic acid	7,436	58.9	2,273	52.0	5,162	62.6
14	Ferritin	8,156	64.6	2,570	58.8	5,586	67.7
15	Vitamin B12	7,441	59.0	2,277	52.1	5,164	62.6
16	Vitamin D	8,217	65.1	2,608	59.7	5,609	68.0
17	Vitamin A	8,393	66.5	2,686	61.5	5,707	69.2
18	Vitamin B1	8,328	66.0	2,665	61.0	5,663	68.7
19	Vitamin B6	8,329	66.0	2,666	61.0	5,663	68.7
20	Vitamin E	8,388	66.5	2,686	61.5	5,702	69.1
21	Zinc	8,052	63.8	2,513	57.5	5,539	67.1
22	Selenium	8,054	63.8	2,513	57.5	5,541	67.2

CRP: C-reactive protein.

Source: *Brazilian National Survey on Child Nutrition* (ENANI-2019).

Note: number of eligible children per age group (Brazil: 12,618; 6-23 months: 4,369; 24-59 months: 8,249).

* 1a, 1c, 9a, and 10: parameters collected on the same date;

** 1b: ferritin and CRP on the same date + ferritin and hemoglobin on the same date or with a 30-day interval;

*** 9b: collected on different dates;

11: collected on the same date or with hemoglobin at a 30-day interval.

sources) for a set of ancillary calibration or post-strata variables ^{21,22}. Calibration helps compensate for children's total non-response, seeking to mitigate the effects of differential non-response that can affect estimates derived from the sample.

Calibration in ENANI-2019 employed total populations of children for 60 post-strata defined by cross-classifying the following variables: major geographic region (5 classes), sex (2 classes), and age (6 classes – 0 to 5 months; 6 to 11 months; 1 year; 2 years; 3 years; 4 years).

The subdivision of children under one year in two age classes for calibration purposes was necessary given the rules for applying part of the questionnaire and collecting blood samples: only children six months of age or older had blood samples drawn. Therefore, to avoid the need to use different population totals for calibration of the principal weights and the weights for groups of blood bio-

Box 1

Predictive variables used to model the probability of response for each group of blood biomarkers. *Brazilian National Survey on Child Nutrition* (ENANI-2019).

DESCRIPTION OF VARIABLE	CATEGORIES OF RESPONSES
Major geographic regions	North Northeast Southeast South Central West
Child weighed	Yes No
Child's age in complete years	0 (6 months or more), 1, 2, 3, 4
Child enrolled in daycare center or school	Yes, daycare center or public school Yes, daycare center or private school; no; doesn't know or doesn't want to answer No, never attended, not applicable
Child with fever, diarrhea, or vomiting on the three days prior to collection	Yes, fever; yes, fever, diarrhea, and vomiting; doesn't know or doesn't want to answer No, yes, diarrhea; yes, vomiting, yes fever and diarrhea; yes, fever and vomiting; yes, diarrhea and vomiting NA = not applicable
Item from the <i>Brazilian Food Insecurity Scale</i> (EBIA)	Yes No; doesn't know or doesn't want to answer
Child's race/color	Brown White; Black, Asian-descendent; Indigenous; doesn't know or doesn't want to answer
Receives social benefits	Yes No; doesn't know or doesn't want to answer
Use of vitamin or mineral supplements in the last 6 months	Yes No; doesn't know or doesn't want to answer
Type of dwelling	Rented Owned, fully paid; owned, still making payments; on loan from the employer; on some other loan; another arrangement; doesn't know or doesn't want to answer

Source: *Brazilian National Survey on Child Nutrition* (ENANI-2019).

marker variables, the calibrations of all the weights to the two age classes for children under one year were considered separately. The totals used for the calibration are population projections by IBGE for January 1, 2020, disaggregated by major geographic region, sex, and five groups of individual ages in years. To obtain the totals for the two age groups under one year, the IBGE projections for children under one year were divided by two.

In calibration of weights, the objective is to minimize the distance (Expression 16):

$$D = \sum_{k \in C} (f_k w_k - w_k)^2 / w_k \quad (16)$$

between the calibrated weights ($f_k w_k$) and the weights one wishes to calibrate (w_k), simultaneously complying with two sets of restrictions:

$$\hat{X}_C = \sum_{k \in C} f_k w_k x_k = \sum_{k \in U} x_k$$

and

$$f_k w_k = f_j w_j \quad \forall k \neq j \in H, \quad \forall H$$

where U represents the set of children in the study population; C represents the set of children in the available effective sample; H represents a household with two or more children interviewed; x_k is the vector with values for the variables that identify the post-stratum to which the children belong (indicators of cells in the table obtained by cross-classifying major geographic region x sex x age group); \bar{X}_C is the estimated total with calibrated weights $f_k w_k$ for the post-strata, and $\sum_{k \in U} x_k$ are the population totals for the post-strata according to the respective population projections.

The estimator using calibrated weights for totals is expressed as:

$$\hat{Y}_C = \sum_{k \in C} f_k w_k y_k \quad (17)$$

and the corresponding estimator for population means is expressed as:

$$\bar{y}_C = \sum_{k \in C} f_k w_k y_k / \sum_{k \in C} f_k w_k \quad (18)$$

The calibrated weights should be used in all the analyses, not only with the children's data but also with other data, such as those of the households, the children's parents or guardians. Calibration of weights in the way described here is called "integrated household weighting", ensuring that all the units (children, etc.) from the same household have equal weights²³.

This statement applies to basic weights but not to weights for the groups of blood biomarker variables. In this case, if the household has children under six months and children over six months of age, the former will have null weights since they did not participate in this part of the survey. As mentioned, for groups of blood biomarker variables, it will always be up to the data analyst to select the adequate weight for each analysis.

To estimate variances, it is recommended to use a combination of the ultimate cluster and linearization methods⁹, as implemented, for example, in the survey package²⁴ of R software (<http://www.r-project.org>).

Effective sample

Even samples with optimal planning may undergo adjustments during data collection for various reasons. Although potential sources of bias, in practice, such adjustments are unavoidable. In the specific case of the ENANI-2019 sample, the main problem during data collection was the interruption on March 17, 2020, due to the COVID-19 pandemic. Before the interruption, there was a need to make substitutions and inclusions of CEAs in the sample, as described below. Even one entire municipality, Jataí (Goiás State), had to be replaced by the municipality of Luziânia (Goiás State), since it was not possible to find a clinical laboratory that could perform the blood sample collection in Jataí (Table 2).

As presented in Table 2, from a total of 1,500 selected CEAs, data collection was performed in 1,382, for a total loss of 7.9%. Losses resulting from the data collection's premature interruption were the highest in the North and Northeast regions (22.3% and 13.7% of the 300 selected CEAs in these regions, respectively). Conversely, there were no losses in the South, and the losses were smaller in the Central West and Southeast (3% and 0.3% of the CEAs, respectively; Table 2).

In the total sample of CEAs, 37 had to be replaced due to difficulties during data collection, because of distance to the municipal center (preventing the blood sample collection), or difficulties in access to the CEA, resulting from civil unrest (drug trafficking, land disputes, etc.), representing 2.5% of the total planned sample of CEAs. Besides these replacements, CEAs had to be added to the study sample to solve 18 cases. The data collection did not produce interviews with eligible households, having exhausted all the PH addresses. Eighteen CEAs were added to the sample to compensate for these cases.

As for the sample of households, 12,524 (83.5%) households were obtained, compared to the expected total of 15,000 eligible households. In addition, data were obtained from 14,558 children, with a loss of only 3% in relation to the expected total of 15,000.

One technique used in the survey was the inverse sampling of households, which functions as “sample screening”. In a sampling process that seeks to locate households with members of a specific population, as in the current survey (children under five years of age), a standard alternative procedure would be to use complete or “census screening”, visiting all the households in the selected CEAs and attempting to determine whether they contained members of the target population. This alternative would involve a higher cost in updating the registry of addresses in the CEAs. In addition, it would represent a stage of creation of a complete registry of eligible households in each CEA. This stage would involve an increase not only in costs but also in time.

By adopting inverse sampling, the screening for eligible households was carried out by sampling and allowed the sample selection and approach for interviewing to occur during the same process of visiting the CEA to locate and visit the selected households. The necessary cost and time for data collection were thus much smaller. An effect of this approach is the selection of large numbers of addresses that lack an eligible household. Still, this cost is much lower than with the alternative approach of registering all the households in each CEA with a visit to verify eligibility.

Overall, it was necessary to select and visit 193,212 addresses in the selected CEAs, resulting in an average of 140 households visited per CEA, with an average of 9 eligible households interviewed per selected CEA. Of all the selected addresses, 75.1% were ultimately ineligible for various reasons, mostly households without children under five years of age (Table 4).

In addition to the ineligible households, 14.2% of all the selected and visited households were classified as closed at the end of the operation (Table 4). The classification of households as closed was for those that had residents at the time of the survey, but where it was not possible to contact residents to apply the protocol to attempt to conduct the data collection for the survey (at least four visits on different days and at different times).

ENANI-2019 experienced a refusal rate of 35.8%, considering only selected eligible households that were contacted successfully (Table 4). Considering all the selected and visited households as the denominator, the refusal rate was 3.7%. The largest loss among the selected and visited households was due to initial refusal to participate in the survey (33.2% of eligible households, as shown in Table 4), not surprising in a survey with the kind of demand that ENANI-2019 exerts on families (obtaining data on their children, including the collection of blood samples).

Table 4

Number of addresses visited, according to the result of visit. *Brazilian National Survey on Child Nutrition* (ENANI-2019).

Result of visit to address	Frequency	% of total	% of eligible households
Total addresses visited	193,212	100.0	
Ineligible households	173,672	89.9	
Household without children under five years of age	139,695	72.3	
Household vacant or with occasional use	4,452	2.3	
Non-residential	922	0.5	
Household closed (after 4 visits)	27,397	14.2	
Address registered but not located	795	0.4	
Other reasons	411	0.2	
Eligible households	19,540	10.1	100.0
Complete interview performed	12,524	6.5	64.1
Refusal by the household	6,479	3.4	33.2
Interview refused after begun	518	0.3	2.7
Interview begun but interrupted by COVID-19	19	0.0	0.1

Source: *Brazilian National Survey on Child Nutrition* (ENANI-2019).

Final remarks

ENANI-2019 is the first nationwide household survey in Brazil that jointly investigated breastfeeding and complementary feeding practices, individual dietary intake, anthropometric nutritional status, and micronutrient deficiencies in children under five years of age. Determination of the sample size and the methodology used in allocating CEAs in the selection strata allowed the representation of the target population in each major geographic region. The study's fieldwork presented good results compared to other household surveys using the highest sampling standards in Brazil. The results will allow comparisons with previous studies and support strategic decisions on implementing public policies for under-five children.

Contributors

M. T. L. Vasconcellos and P. L. N. Silva contributed to the sampling of the study and to the article's writing and review. I. R. R. Castro, C. S. Boccolini, N. H. Alves-Santos, and G. Kac contributed to the study conception and design, and to the article's writing and review. All authors approved the final version of the article.

Acknowledgments

To the field staff and participating families who made this study possible. To the Brazilian Ministry of Health/Brazilian Nacional Research Council (CNPq) – process: 440890/2017-9.

References

1. Instituto Brasileiro de Geografia e Estatística. Projeções da população. Pesquisa Nacional por Amostra de Domicílios 2012. Notas metodológicas. Pesquisa básica. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2013.
2. Alves-Santos NH, Castro IRR, Anjos LA, Lacerda EMA, Normando P, Freitas MB, et al. General methodological aspects in the *Brazilian National Survey on Child Nutrition* (ENANI-2019): a population-based household survey. *Cad Saúde Pública* 2021; 37:e00300020.
3. Lacerda EMA, Boccolini CS, Alves-Santos NH, Castro IRR, Anjos LA, Crispim SP, et al. Methodological aspects of the assessment of dietary intake in the *Brazilian National Survey on Child Nutrition* (ENANI-2019): a population-based household survey. *Cad Saúde Pública* 2021; 37:e00301420.
4. Anjos LA, Ferreira HS, Alves-Santos NH, Freitas MB, Boccolini CS, Lacerda EMA, et al. Methodological aspects of the anthropometric assessment in the *Brazilian National Survey on Child Nutrition* (ENANI-2019): a population-based household survey. *Cad Saúde Pública* 2021; 37:e00293320.
5. Castro IRR, Normando P, Alves-Santos NH, Bezerra FF, Citelli M, Pedrosa LFC, et al. Methodological aspects of the micronutrient assessment in the *Brazilian National Survey on Child Nutrition* (ENANI-2019): a population-based household survey. *Cad Saúde Pública* 2021; 37:e00301120.

6. Instituto Brasileiro de Geografia e Estatística. Estimativas da população residente nos municípios e para as Unidades da Federação brasileiros com data de referência em 1º de julho de 2016 (notas metodológicas). Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2016.
7. Madeira JL, Simões CCS. Estimativas preliminares da população urbana e rural segundo as unidades da federação, de 1960/1980 por uma nova metodologia. *Revista Brasileira de Estatística* 1972; 33:3-11.
8. Cochran WG. *Sampling techniques*. 3rd Ed. New York: John Wiley & Sons; 1977.
9. Pessoa DG, Silva PLN. Análise de dados amostrais complexos. São Paulo: Associação Brasileira de Estatística; 1998.
10. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios Contínua. Notas metodológicas, versão 1.8. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2020.
11. Bankier MD. Power allocations: determining sample sizes for subnational areas. *Am Stat* 1988; 42:174-7.
12. Rosén BA. *A user's guide to Pareto tops sampling*. Stockholm: Statistiska Centralbyrån; 2000.
13. Freitas MPS, Antonaci GA. Sistema Integrado de Pesquisas Domiciliares: amostra mestra 2010 e amostra da PNAD Contínua. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2014.
14. Haldane JBS. On a method of estimating frequencies. *Biometrika* 1945; 33:222-5.
15. Vasconcellos MTL, Silva PLN, Szwarcwald CL. Sampling design for the *World Health Survey* in Brazil. *Cad Saúde Pública* 2005; 21 Supl:S89-99.
16. Vasconcellos MTL, Silva PLN, Anjos LA. Sample design for the Nutrition, Physical Activity and Health Survey (PNAFS), Niterói, Rio de Janeiro, Brazil. *Estatística* 2013; 65:47-61.
17. Valliant R, Dever JA. *Survey weights: a step-by-step guide to calculation*. College Station: Stata Press; 2018.
18. Potter FJ. Survey of procedures to control extreme sampling weights. In: *Proceedings of the Survey Research Methods Section*. Alexandria: American Statistical Association; 1988. p. 453-8.
19. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 1952; 47:663-85.
20. Hájek J. Comment on a paper by D Basu. In: Godambe VP, Sprott DA, editors. *Foundations of statistical inference*. Toronto: Holt, Rinehart and Winston; 1971. p. 236.
21. Silva PLN. Calibration estimation: when and why, how much and how. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2004. (Textos para Discussão da Diretoria de Pesquisas, 14).
22. Deville JC, Särndal CE. Calibration estimators in survey sampling. *J Am Stat Assoc* 1992; 87:376-82.
23. Lemaitre GE, Dufour J. An integrated method for weighting persons and families. *Surv Methodol* 1987; 13:199-207.
24. Lumley T. *Complex surveys: a guide to analysis using R*. Hoboken: John Wiley & Sons; 2010. (Wiley Series in Survey Methodology).

Resumo

O objetivo deste artigo é descrever aspectos metodológicos referentes à definição da população da pesquisa, plano amostral, ponderação e calibração da amostra efetiva do Estudo Nacional de Alimentação e Nutrição Infantil (ENANI-2019). Trata-se de um inquérito populacional de base domiciliar que realizou avaliação do aleitamento materno e de consumo alimentar, avaliação antropométrica do estado nutricional, e avaliação das deficiências de micronutrientes mediante análise de biomarcadores sanguíneos em crianças menores de 5 anos de idade. Seus dados foram obtidos por meio de uma amostra probabilística domiciliar, com estratificação geográfica por macrorregião e conglomeração por setores censitários. A amostra foi dimensionada em 15.000 domicílios, distribuídos em 1.500 setores censitários, sendo 300 em cada macrorregião e 10 domicílios elegíveis por setor, através de amostragem inversa. Assim, estimaram-se os parâmetros populacionais requeridos para atingir os objetivos do estudo. Os pesos amostrais básicos do desenho foram calculados como inversos das probabilidades de inclusão dos domicílios na pesquisa. Para compensar a não resposta de itens das variáveis pesquisadas foi usada imputação, com exceção para os dados de biomarcadores sanguíneos. A calibração empregou totais populacionais de crianças para 60 pós-estratos definidos por cruzamento das variáveis macrorregião, sexo e idade. A amostra final compreendeu 14.558 crianças, residentes em 12.524 domicílios, distribuídos em 1.382 setores censitários nas 27 Unidades da Federação. Os dados do ENANI-2019 poderão subsidiar estratégias de promoção e implementação de políticas públicas para crianças menores de 5 anos.

Lactente; Pré-escolar; Modelos Estatísticos; Amostragem; Métodos

Resumen

El objetivo de este artículo es describir aspectos metodológicos referentes a definición de la población de la investigación, plan de muestreo, ponderación de la muestra y muestra efectiva del Estudio Nacional de Alimentación y Nutrición Infantil (ENANI-2019). Se trata de una encuesta poblacional de base domiciliaria, que realizó una evaluación de la lactancia materna y de consumo alimentario, así como una evaluación antropométrica del estado nutricional y de las deficiencias de micronutrientes, mediante análisis de biomarcadores sanguíneos en niños menores de cinco años de edad. Sus datos se obtuvieron mediante una muestra probabilística domiciliaria, con estratificación geográfica por macrorregión y conglomerados por sectores censitarios. La muestra se circunscribió a 15.000 domicilios, distribuidos en 1.500 sectores censitarios, encontrándose 300 en cada macrorregión, junto 10 domicilios elegibles por sector, a través de un muestreo inverso. De esta forma, se estimaron los parámetros poblacionales requeridos para alcanzar los objetivos del estudio. Los pesos básicos de las muestras del diseño se calcularon como inversos a las probabilidades de inclusión de los domicilios en la investigación. Para compensar la no respuesta de ítems de las variables investigadas se usó la imputación, con excepción de los datos con biomarcadores sanguíneos. La calibración empleó totales poblacionales de niños para los 60 post estratos, definidos mediante el cruce de las variables macrorregión, sexo y edad. La muestra final comprendió 14.558 niños, residentes en 12.524 domicilios, distribuidos en 1.382 sectores censitarios dentro de las 27 Unidades de la Federación. Los datos del ENANI-2019 podrán apoyar estrategias de promoción e implementación de políticas públicas para niños menores de cinco años.

Lactante; Preescolar; Modelos Estadísticos; Muestreo; Métodos

Submitted on 13/Feb/2021

Final version resubmitted on 29/Apr/2021

Approved on 18/May/2021