

Sampling Design for the National Health Survey, Brazil 2013

Paulo Roberto Borges de Souza-Júnior

Fundação Instituto Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Rio de Janeiro-RJ, Brasil

Marcos Paulo Soares de Freitas

Instituto Brasileiro de Geografia e Estatística, Coordenação de métodos e Qualidade, Rio de Janeiro-RJ, Brasil

Giuseppe de Abreu Antonaci

Instituto Brasileiro de Geografia e Estatística, Diretoria de Pesquisas, Rio de Janeiro-RJ, Brasil

Célia Landmann Szwarcwald

Fundação Instituto Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Rio de Janeiro-RJ, Brasil

Resumo

Este artigo descreve o plano amostral usado na Pesquisa Nacional de Saúde 2013. A população-alvo constitui-se de pessoas residentes em domicílios particulares permanentes de todo o território nacional; a pesquisa foi domiciliar, com amostragem estratificada e três estágios de conglomeração; os setores censitários formaram as unidades primárias de amostragem, os domicílios foram as unidades de segundo estágio e os moradores adultos (18 anos ou mais) definiram as unidades de terceiro estágio; o tamanho da amostra considerou o nível de precisão desejado para as estimativas de alguns indicadores em diferentes níveis de desagregação e grupos populacionais; o peso final foi um produto do inverso das probabilidades de seleção em cada estágio do plano amostral, incluindo processos de correção de não respostas e calibrações de ajustes para os totais populacionais conhecidos. Por se tratar de uma amostra complexa, alguns cuidados devem ser levados em consideração no momento do processamento dos dados.

Palavras-chave: Inquéritos Populacionais; Amostragem por Conglomerados; Tamanho de Amostra.

Abstract

This paper describes the sample design used in the Brazilian 2013 National Health Survey. The target population is composed by people resident in permanent private households throughout the country; the survey was household-based with stratified sampling and three clustering stages; census tracts form the primary sampling units, households are the units of second stage and adults (aged 18 years or older) define the third-stage units; the sample size considered the desired level of precision for the estimates of some indicators at different levels of disaggregation and different population groups; the final weighting was a product of inverse selection probabilities at each stage of the sampling plan, including non-response correction procedures and adjustment calibrations for the known population totals. Since this is a complex sample, appropriate procedures must be used during data processing.

Key words: Population Surveys; Cluster Sampling; Sample Size.

Endereço para correspondência:

Paulo Roberto Borges de Souza-Júnior – Fundação Instituto Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde/Ministério da Saúde, Av. Brasil, nº 4365, Pavilhão Haity Moussatché, Manguinhos, Rio de Janeiro-RJ, Brasil. CEP: 21040-360. E-mail: paulo.borges@icict.fiocruz.br.

Introdução

A Pesquisa Nacional de Saúde (PNS), realizada entre 2013 e 2014, foi desenvolvida em uma parceria entre a Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS), a Fundação Oswaldo Cruz (Fiocruz) e o Instituto Brasileiro de Geografia e Estatística (IBGE). Trata-se de um inquérito populacional, de abrangência nacional, que tem como principal objetivo produzir (i) dados em âmbito nacional, sobre a situação de saúde e os estilos de vida da população brasileira, e (ii) dados sobre a atenção à saúde, no que se refere ao acesso e uso dos serviços de saúde, às ações preventivas, à continuidade dos cuidados e ao financiamento da assistência de saúde.¹

A PNS, como parte do Sistema Integrado de Pesquisas Domiciliares (SIPD) do IBGE, utilizou a infraestrutura amostral construída para esse sistema.² Essa etapa de planejamento amostral da pesquisa também foi realizada pelo IBGE, em parceria com a Fiocruz.

Este artigo descreve o plano amostral usado na PNS, incluindo a população-alvo, os estágios de seleção, o cálculo do tamanho da amostra e a definição dos fatores de expansão.

População-alvo

A população-alvo da Pesquisa Nacional de Saúde – PNS – é composta pelas pessoas residentes em domicílios particulares permanentes (DPP), pertencentes à área de abrangência geográfica da pesquisa. Os DPP são aqueles domicílios construídos para servir exclusivamente como habitação, com a finalidade de servir de moradia a uma ou mais pessoas.³

Foi definido como 'abrangência geográfica' da pesquisa todo o território nacional, dividido nos setores censitários da Base Operacional Geográfica de 2010. Foram excluídas da pesquisa as áreas com características especiais e com pouca população, classificadas pelo IBGE na base de setores, como aldeias indígenas, quartéis, bases militares, alojamentos, acampamentos, embarcações, penitenciárias, colônias penais, presídios, cadeias, asilos, orfanatos, conventos e hospitais. Também foram excluídos os setores censitários localizados em terras indígenas.

Questionário

O questionário da PNS é constituído de três partes:
1ª parte – Domiciliar

Questões sobre informações do domicílio e visitas domiciliares realizadas pela equipe de Saúde da Família e agentes de endemias.

Esta primeira parte do questionário é respondida pela pessoa responsável pelo domicílio ou pela pessoa que detenha essas informações no momento da entrevista.

2ª parte – Moradores do domicílio

Questões relativas às características gerais de todos os moradores do domicílio, incluindo nível de educação, trabalho, rendimento, deficiências, cobertura de plano de saúde, utilização dos serviços de saúde, saúde do idoso, cobertura de mamografia e características de crianças menores de 2 anos de idade.

Esta segunda parte é respondida por todos os moradores do domicílio. Em caso de ausência ou impossibilidade do morador ser entrevistado, a pessoa responsável pode responder ao questionário (ou parte dele) em substituição ao morador ausente, como é feito na Pesquisa Nacional por Amostra de Domicílios (PNAD). Quando isso ocorre, essa informação fica registrada no questionário, que identifica o número de ordem do respondente.

3ª parte – Individual

Questões dirigidas a um morador adulto (18 anos ou mais de idade), selecionado aleatoriamente, sobre outras características de trabalho e apoio social, percepção do estado de saúde, acidentes e violências, estilos de vida, doenças crônicas, saúde da mulher, atendimento pré-natal, saúde bucal e atendimento médico.

Nesta terceira parte do questionário, é selecionado um adulto (≥ 18 anos) aleatoriamente, entre os moradores elegíveis. Essa parte do questionário só pode ser respondida pelo morador selecionado, não sendo permitido que outro morador do domicílio responda por ele.

Plano amostral

Como parte integrante do SIPD, a amostra da PNS é uma subamostra da Amostra Mestra. Antes de apresentar o plano amostral da PNS, faz-se necessário uma breve descrição da Amostra Mestra do SIPD. Para mais informações sobre a Amostra Mestra ou sobre o SIPD, recomenda-se consultar Freitas e colaboradores⁴ e Freitas e Antonaci.⁵

Amostra Mestra

A Amostra Mestra é um conjunto de unidades de áreas selecionadas para atender a diversas pesquisas.

Essas unidades são consideradas unidades primárias de amostragem (UPA), dentro do planejamento amostral de cada uma das pesquisas que utilizam a Amostra Mestra – incluída a PNS. As UPA são setores censitários ou conjuntos de setores censitários (quando esses setores contam poucos domicílios).

Para a seleção da Amostra Mestra, as UPA foram estratificadas segundo quatro diferentes critérios (Figura 1):

1. Administrativo

Estratifica as UPA por Unidades da Federação (UF); e dentro de cada UF, subdivide as UPA em capital, resto da Região Metropolitana (RM) ou Região Integrada de Desenvolvimento Econômico (RIDE)

(nas UF que dispõem dessas divisões administrativas) e resto da UF.

2. Geográfico

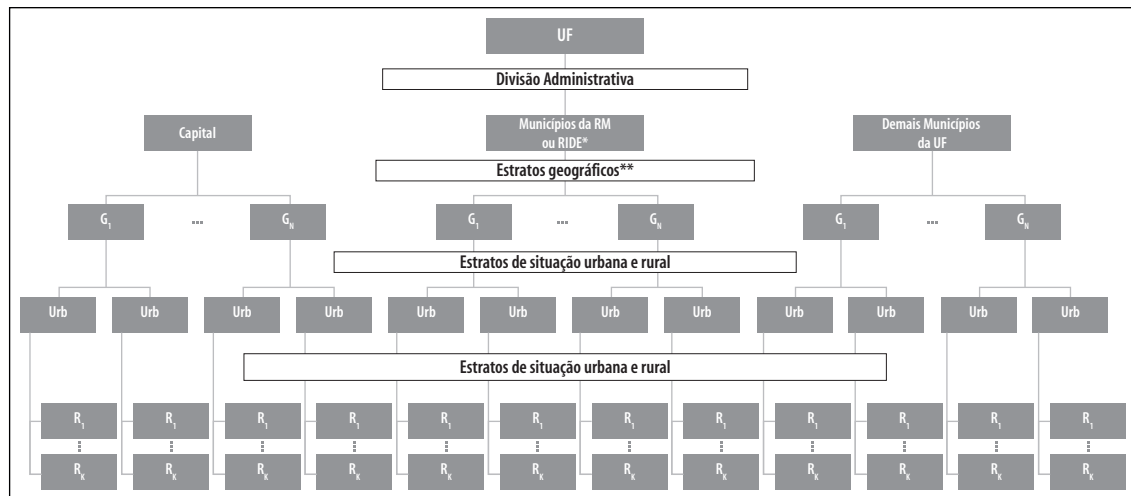
Subdivide as capitais e outros municípios de grande porte em mais estratos, a exemplo de distrito, sub-distrito e bairros.

3. De situação

Subdivide os estratos, descritos em epígrafe, como urbano e rural.

4. Estatístico

Subdivide os estratos em epígrafe em estratos homogêneos, segundo as informações de rendimento total dos domicílios e total de DPP, com o objetivo de melhorar a precisão das estimativas.



* RM, Região Metropolitana; RIDE, Região Integrada de Desenvolvimento Econômico.

** Subdivide as capitais e outros municípios de grande porte em mais estratos, como distrito, sub-distrito e bairros, por exemplo.

Nota: Figura adaptada de Freitas e colaboradores⁴

Figura 1 – Esquema de estratificação da Amostra Mestra do Sistema Integrado de Pesquisas Domiciliares, utilizado na Pesquisa Nacional de Saúde. Brasil, 2013

A seleção das UPA dentro de cada estrato foi feita por amostragem probabilística proporcional ao tamanho (PPT), sendo considerado como medida de tamanho das UPA o número de DPP.

Desenho amostral da PNS

A PNS é uma pesquisa domiciliar, cujo plano amostral empregado foi o de amostragem por conglomerado em três estágios de seleção, com estratificação das UPA.

Como parte do SIPD, no primeiro estágio, a seleção das UPA foi obtida por amostragem aleatória simples (AAS) entre aquelas previamente selecionadas para a

Amostra Mestra, mantendo-se a estratificação das UPA utilizada na Amostra Mestra e descrita anteriormente.

No segundo estágio, foi selecionado – por AAS – um número fixo de DPP em cada UPA selecionada no primeiro estágio. A seleção dos domicílios foi feita a partir do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) em sua última atualização antes da conclusão dessa etapa do plano amostral. É importante ressaltar que para as informações contidas na primeira e na segunda parte do questionário, ou seja, as partes referentes às características do domicílio e ao conjunto de todos os moradores do domicílio, o plano amostral da PNS possui apenas esses dois estágios de seleção.

No terceiro estágio, dentro de cada domicílio da amostra, um morador com 18 anos ou mais de idade foi selecionado – também por AAS – para responder à 3ª parte (individual) do questionário. Essa seleção foi feita a partir de uma lista de moradores elegíveis, construída no momento da entrevista.

Cálculo do tamanho da amostra

O tamanho da amostra foi definido considerando-se o nível de precisão desejado, em intervalos de 95% de confiança ($IC_{95\%}$), para as estimativas de alguns indicadores (parâmetros) de interesse em diferentes níveis de desagregação geográfica e grupos populacionais específicos. Para que fosse possível definir o tamanho

da amostra nesses grupos populacionais específicos, foi preciso avaliar a proporção de DPP que possuíam pessoas nesses grupos, com base nos dados do Censo Demográfico 2010. O nível de precisão desejado para cada indicador foi baseado na amplitude do intervalo de 95% de confiança que se esperava obter. A partir daí, calculou-se o erro-padrão e o coeficiente de variação (CV) desejado para cada indicador, este último utilizado como medida de precisão no cálculo do tamanho da amostra.

Os cálculos iniciais do tamanho da amostra foram baseados em amostragem aleatória simples, sendo necessário considerar o efeito do plano amostral (EPA), ou efeito de desenho, que indica o quanto a amostragem por conglomerados (AC) – caso da

Tabela 1 – Grupo populacional, proporção de domicílios com pessoas no grupo populacional, valor esperado, coeficiente de variação, efeito do plano amostral estimado e tamanho inicial da amostra de domicílios segundo o indicador de interesse – Pesquisa Nacional de Saúde. Brasil, 2013

Indicador	Grupo populacional	Proporção de domicílios com pessoas no grupo populacional Censo 2010	Valor esperado do indicador (%)	CV ^a desejado (%)	EPA ^b estimado	Tamanho da amostra de domicílios
1. Prevalência de diabetes em adultos de 35 anos e mais	Adultos de 35 anos e mais anos selecionados	60,0	8,0	12,8	1,4	1641
2. Prevalência de hipertensão em adultos de 35 anos e mais	Adultos de 35 anos e mais anos selecionados	60,0	30,0	8,5	1,6	856
3. Prevalência de depressão em adultos de 35 anos e mais	Adultos de 35 anos e mais anos selecionados	60,0	8,0	12,8	1,9	1425
4. Taxa de internação no último ano	Todos os moradores do domicílio	100,0	7,0	14,6	2,9	546
5. Utilização de serviços de saúde nas últimas 2 semanas	Todos os moradores do domicílio	100,0	14,0	7,3	4,6	1602
6. Cobertura de plano de saúde privado	Todos os moradores do domicílio	100,0	26,0	5,9	10,4	1447
7. Cobertura de exame preventivo de câncer de colo de útero em mulheres de 25 a 59 anos	Mulheres de 25-59 anos selecionadas	40,0	80,0	3,2	2,2	1490
8. Cobertura de mamografia entre mulheres de 50 anos e mais	Mulheres de 50-69 anos selecionadas	10,0	65,0	5,5	1,8	1671
9. Prevalência de fumo	Adultos selecionados	100,0	17,0	9	1,8	1084
10. Prevalência de sobrepeso/obesidade	Adultos selecionados	100,0	12,0	12,8	1,8	811
11. Prevalência de sedentarismo	Todos os moradores do domicílio	100,0	15,0	10,2	4,1	672
12. Uso abusivo de álcool	Adultos selecionados	100,0	7,0	14,6	1,8	1125
13. Proporção de quem sofreu violência com lesões corporais	Adultos selecionados	100,0	2,0	25,5	2,8	1205
14. Proporção de idosos com limitações (problemas de funcionalidade)	Todos os idosos do domicílio	40,0	15,0	13,6	2,1	1786

a) CV: coeficiente de variação

b) EPA: efeito do plano amostral

Tabela 2 – Tamanho selecionado da amostra segundo domínios de interesse – Pesquisa Nacional de Saúde. Brasil, 2013

Domínio	Número de UPA ^a selecionadas na amostra	Número esperado de domicílios com entrevista realizada	Número de domicílios selecionados na amostra ^b
Norte	1.161	13.578	17.553
Rondônia	129	1.806	2.322
Acre	180	1.800	2.340
Amazonas	245	2.594	3.365
Roraima	124	1.736	2.232
Pará	246	2.652	3.438
Amapá	107	1.498	1.926
Tocantins	130	1.492	1.930
Nordeste	1.916	19.160	24.908
Maranhão	181	1.810	2.353
Piauí	180	1.800	2.340
Ceará	290	2.900	3.770
Rio Grande do Norte	179	1.790	2.327
Paraíba	182	1.820	2.366
Pernambuco	279	2.790	3.627
Alagoas	180	1.800	2.340
Sergipe	179	1.790	2.327
Bahia	266	2.660	3.458
Sudeste	1.451	14.510	18.863
Minas Gerais	366	3.660	4.758
Espírito Santo	181	1.810	2.353
Rio de Janeiro	365	3.650	4.745
São Paulo	539	5.390	7.007
Sul	767	7.670	9.971
Paraná	287	2.870	3.731
Santa Catarina	186	1.860	2.418
Rio Grande do Sul	294	2.940	3.822
Centro-Oeste	774	7.740	10.062
Mato Grosso do Sul	179	1.790	2.327
Mato Grosso	162	1.620	2.106
Goiás	253	2.530	3.289
Distrito Federal	180	1.800	2.340
Brasil	6.069	62.658	81.357

a) UPA: unidade primária de amostragem

b) Incluindo, aproximadamente, 23% de perdas esperadas.

PNS – é menos eficiente que a AAS. Por essa razão, os tamanhos de amostra iniciais foram aumentados para que se alcançasse a mesma precisão.⁶ Os EPA utilizados nos cálculos de tamanho de amostra da PNS foram estimados com base nos dados da Pesquisa Nacional por Amostra de Domicílios 2008 (PNAD 2008).

Os indicadores utilizados para o dimensionamento da amostra, seus respectivos grupos populacionais, os valores esperados, coeficientes de variação desejados, os EPA estimados e os tamanhos inicialmente calculados para a amostra de domicílios são apresentados na Tabela 1.

A seguir, são apresentadas as formulas utilizadas para cálculo do tamanho da amostra em cada domínio:

$$n_{AAS} = \frac{N}{N-I} * P * Q * \frac{I}{CV^2 * P^2 + \frac{P * Q}{N-I}}$$

$$n_{AC} = n_{AAS} * EPA$$

Onde

n_{AAS} é o tamanho de amostra de pessoas sob amostragem aleatória simples;

N é o número total de pessoas residentes em DPP no domínio;

P é a proporção de interesse;

$Q = I - P$;

CV é o coeficiente de variação desejado da estimativa

de proporção;

n_{AC} é o tamanho de amostra sob amostragem por conglomerados; e

EPA é o efeito de plano amostral.

Os dados utilizados nos cálculos foram obtidos do Censo Demográfico 2010 e da PNAD 2008.

Inicialmente, foram calculados os tamanhos mínimos de amostra necessários para estimar os indicadores apresentados na Tabela 1, com o coeficiente de variação (CV) desejado, para cada um dos domínios de divulgação, ou seja, áreas onde a pesquisa é capaz de gerar estimativas com precisão suficiente e representativa da população. Os domínios são: Brasil; grande região nacional; unidade da Federação (UF); Região Metropolitana (RM); capital e restante da UF. A Tabela 1 também apresenta os tamanhos mínimos de amostra calculados inicialmente, para cada um dos indicadores.

Como as RM e as capitais são domínios de divulgação, as UF que não possuem RM contam com dois domínios de divulgação – capital e restante da UF –, fazendo com que o tamanho mínimo de amostra nessas UF, necessário para estimar cada indicador, seja aproximadamente o dobro do apresentado na Tabela 1. Nos casos das UF que possuem RM, o tamanho necessário seria um pouco maior que o dobro, haja vista a RM também ser um domínio de divulgação.

Com os resultados encontrados e pelo que foi exposto anteriormente, foram necessários alguns ajustes nos tamanhos mínimos de amostra e nos níveis geográficos inicialmente pensados como domínios.

Após diversas avaliações, optou-se por determinar o tamanho mínimo da amostra de domicílios por UF em 1.800 e, como são esperados no mínimo dois domínios de divulgação por UF, determinou-se que o tamanho mínimo da amostra fosse de 900 domicílios por domínio.

O número de domicílios selecionados em cada UPA foi fixado em 10. Porém, observou-se que em alguns domínios não havia UPA suficientes na Amostra Mestra para completar a amostra da PNS. Nesses domínios, foi necessário selecionar 14 domicílios por UPA. Ou seja, o número de domicílios por UPA foi de 10 ou 14, a depender do domínio de divulgação. Os tamanhos da amostra por UF são apresentados na Tabela 2.

O número de domicílios selecionados para a pesquisa foi, aproximadamente, 23% maior que a

amostra mínima descrita em epígrafe, para levar em consideração a perda total, que inclui a taxa de não resposta e a taxa de má classificação das unidades no cadastro de seleção. Esse percentual foi estimado com base em outras pesquisas do IBGE – por exemplo, a Pesquisa Especial do Tabagismo (PETaB), incorporada à PNAD 2008, porém, entrevistou apenas uma pessoa no domicílio.

Expansão da amostra

Por ter um desenho complexo de amostragem e com probabilidades desiguais de seleção, para a análise dos dados da PNS, é preciso definir os fatores de expansão ou pesos amostrais dos domicílios e todos seus moradores, assim como do morador selecionado para responder a terceira parte do questionário. O peso final é um produto do inverso das probabilidades de seleção em cada estágio do plano amostral, incluídos os processos de correção de não respostas e as calibrações para ajustes dos totais populacionais conhecidos.

Peso das unidades primárias de amostragem (UPA)

Os pesos das UPA foram calculados como o produto entre o inverso da probabilidade de seleção da UPA para a Amostra Mestra e o inverso da probabilidade de seleção da UPA para a PNS:

$$W_{hi} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}}$$

onde

h é o índice do estrato;

i é o índice da UPA;

W_{hi} é o peso básico da UPA i do estrato h na PNS;

m_h é o número de UPA selecionadas no estrato h para a Amostra Mestra;

N_{hi} é o número de domicílios particulares permanentes ocupados, ocupados sem entrevistas realizadas (equivalentes aos domicílios fechados) e vagos na UPA i do estrato h [dados atualizados do Cadastro Nacional de Endereços para Fins Estatísticos –CNEFE – no momento da seleção das UPA para a Amostra Mestra];

N_b é o número de domicílios particulares permanentes ocupados, ocupados sem entrevistas realizadas

(equivalentes aos domicílios fechados) e vagos no estrato h [dados atualizados do CNEFE no momento da seleção das UPA para a Amostra Mestra]; e

m_h^{PNS} é o número de UPA selecionadas no estrato h para a PNS.

Peso do domicílio e de todos seus moradores

Os pesos dos domicílios e de todos seus moradores foram calculados pelo produto do peso da UPA correspondente e o inverso da probabilidade de seleção do domicílio dentro da UPA. Os pesos foram ajustados para correção de não respostas e para calibrar as estimativas com os totais populacionais conhecidos de outras fontes. Esses pesos são utilizados na análise das questões referentes às duas primeiras partes do questionário, com informações sobre características dos domicílios e todos seus moradores.

Os domicílios foram selecionados com probabilidade igual em cada UPA; assim, o peso do domicílio dentro da UPA é dado por:

$$W_{j|hi} = \frac{N_{hi}^*}{n_{hi}}$$

onde

h é o índice do estrato;

i é o índice da UPA;

j é o índice do domicílio;

$W_{j|hi}$ é o peso de seleção do domicílio j na UPA i do estrato h ;

N_{hi}^* é o número de domicílios particulares permanentes ocupados e fechados na UPA i do estrato h [dados da última atualização do CNEFE no momento da seleção dos domicílios]; e

n_{hi} é o número de domicílios selecionados na UPA i do estrato h .

E o peso básico do domicílio é obtido pela expressão:

$$W_{hij} = W_{hi}^* \cdot W_{j|hi} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}}$$

Após a determinação do peso básico, fez-se um ajuste para compensar as perdas de entrevistas por não resposta, ou seja, entrevistas não realizadas por recusa do informante, impossibilidade de contato com o morador ou outro motivo para perda em domicílios ocupados.

O fator de ajuste para as perdas foi calculado pela razão entre o número de domicílios selecionados

e ocupados (com morador) na UPA e o número de domicílios selecionados e ocupados com entrevista realizada na UPA. A fórmula utilizada para esse cálculo é apresentada a seguir:

$$W_{hij}^* = W_{hij} * \frac{n_{hi}^*}{n_{hi}^{**}}$$

onde

n_{hi}^* é o número de domicílios selecionados e ocupados (com morador) na UPA i do estrato h ; e

n_{hi}^{**} é o número de domicílios selecionados e ocupados (com morador) com entrevista realizada na UPA i do estrato h .

Após a correção dos pesos para as não respostas, foi feito um novo ajuste nos pesos dos domicílios, denominado de calibração ou pós-estratificação. O objetivo principal da calibração dos pesos é obter estimativas coerentes com as projeções da população divulgadas pelo IBGE,^{7,8} de modo que, ao estimar o total populacional de certos níveis geográficos, a estimativa obtida coincida com a estimativa populacional produzida pela Coordenação de População e Indicadores Sociais (COPIS) da Diretoria de Pesquisas do IBGE.

A calibração foi feita para a população estimada no dia 27 de julho de 2013, utilizando-se a seguinte expressão:

$$W_{hij}^{**} = W_{hij}^* * \frac{P_d^{tri}}{\hat{P}_d^{tri}}$$

onde

P_d^{tri} é a estimativa populacional produzida pela COPIS para o nível geográfico d no dia 27 de julho de 2013; e \hat{P}_d^{tri} é a estimativa populacional obtida com os dados da pesquisa para o nível geográfico d .

Portanto, o peso final do domicílio, com correção para as perdas e calibração para os totais populacionais, é dado por:

$$W_{hij}^{**} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}} * \frac{n_{hi}^*}{n_{hi}^{**}} * \frac{P_d^{tri}}{\hat{P}_d^{tri}}$$

Peso do morador selecionado

A seleção do morador que respondeu à entrevista individual foi feita por amostragem aleatória simples; assim, o peso do morador selecionado foi calculado pelo produto do peso do domicílio pelo número de moradores elegíveis no domicílio (equivalente ao inverso da probabilidade de seleção). Desse modo, o

peso básico é dado por:

$$W_{hijk} = W_{hij} * O_{hij} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}} * O_{hij}$$

E o peso do morador selecionado, incluído o ajuste de não resposta do domicílio, é dado por:

$$W'_{hijk} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}} * \frac{n_{hi}^*}{n_h^{**}} * O_{hij}$$

onde

k é o índice do morador selecionado; e

O_{hij} é o número de moradores com 18 anos ou mais de idade no domicílio j na UPA i do estrato h .

Como houve perda de entrevista após a seleção do morador, também houve necessidade de correção de não respostas nesse estágio. Em um primeiro momento, o ajuste foi feito de maneira equivalente à já descrita, dentro da UPA, considerando-se os domicílios com entrevista do morador realizada. Após a análise das estimativas de pessoas por sexo e das características dos moradores não respondentes, decidiu-se pelo ajuste por sexo ao se perceber perda maior entre os homens, frente às mulheres. Os pesos, segundo sexo, ficaram da seguinte forma:

$$W_{hijk}^M = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}} * \frac{n_{hi}^*}{n_h^{**}} * O_{hij} * \frac{\sum_j n_{hijk}^{**} W'_{hijk} * \alpha_{hijk}^M}{\sum_i n_{hijk}^{***} W'_{hijk} * \alpha_{hijk}^M}$$

$$W_{hijk}^F = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}} * \frac{n_{hi}^*}{n_h^{**}} * O_{hij} * \frac{\sum_j n_{hijk}^{**} W'_{hijk} * \alpha_{hijk}^F}{\sum_i n_{hijk}^{***} W'_{hijk} * \alpha_{hijk}^F}$$

onde

n_{hij}^{**} é o número de domicílios selecionados com entrevista realizada na UPA i do estrato h ;

n_{hi}^{***} é o número de moradores selecionados com entrevista realizada na UPA i do estrato h ;

α_{hijk}^M indica se o morador selecionado no domicílio j da UPA i do estrato h é do sexo masculino; e

α_{hijk}^F indica se o morador selecionado no domicílio j da UPA i do estrato h é do sexo feminino.

Como é feita uma amostra aleatória simples de um morador dentro do domicílio, é natural que, por conta da aleatoriedade na seleção desse morador, os totais populacionais obtidos com os fatores de expansão do morador selecionado não sejam exatamente iguais aos totais populacionais obtidos com os fatores de expansão de domicílio.

Entretanto, os moradores dos domicílios formam uma amostra muito maior que os moradores selecionados e, portanto, uma estimativa mais precisa desses totais populacionais. Para que essas estimativas ficassem iguais, optou-se por calibrar o peso do morador selecionado de maneira que os totais populacionais por sexo e classes de idade correspondessem aos totais obtidos com o peso do domicílio. As quatro classes de idade utilizadas foram: 18 a 24; 25 a 39; 40 a 59; e 60 anos e mais.

A calibração do peso do morador selecionado foi feita pela seguinte fórmula, análoga para o peso do morador selecionado do sexo feminino:

$$W_{hijk}^{M*} = W_{hijk}^M * \frac{\hat{P}_{d,M,c}^M}{\hat{P}_{d,M,c}^S}$$

onde

$\hat{P}_{d,M,c}^M$ é a estimativa populacional obtida com os dados dos moradores dos domicílios da pesquisa para o nível geográfico d sexo M e classe de idade c ; e

$\hat{P}_{d,M,c}^S$ é a estimativa populacional obtida com os dados dos moradores selecionados da pesquisa para o nível geográfico d sexo M e classe de idade c .

As taxas de não respostas para os domicílios e para o morador selecionado são apresentadas na Tabela 3.

Considerações finais

Por se tratar de dados provenientes de uma amostra complexa que gera estimativas a partir de dados coletados para o domicílio e todos seus moradores (segundo estágio de seleção da amostra), assim como para os moradores de 18 anos ou mais de idade selecionados aleatoriamente (terceiro estágio), incluído um processo de calibração dos pesos em ambos os estágios, alguns cuidados devem ser levados em consideração no momento do processamento dos dados.

Toda amostra oriunda de pesquisas por conglomerados deve ser analisada em programas estatísticos com algoritmos para análise de dados complexos, capazes de corrigir o efeito que a conglomeração das unidades primárias de amostragem causa nas estimativas: o efeito do plano amostral – EPA. A maioria dos pacotes estatísticos possuem módulos ou bibliotecas capazes de corrigir o EPA, a exemplo da biblioteca Survey do pacote R, do módulo *Complex Sample* do SPSS, do SUDAAN, entre outros.

Tabela 3 – Número de domicílios na amostra e taxas de resposta segundo a unidade da Federação – Pesquisa Nacional de Saúde. Brasil, 2013

Unidade da Federação	Total de domicílios ocupados ^a				Total de domicílios ocupados com morador selecionado para a entrevista ^a			
	Total (C1)	Com entrevista realizada (C2)	Sem entrevista realizada (C3)	Taxa de resposta dos domicílios (%) (C2/C1)	Total (C1s)	Com entrevista realizada (C2s)	Sem entrevista realizada (C3s)	Taxa de resposta dos moradores selecionados (%) (C2s/C1s)
Rondônia	1.961	1.849	112	94,3	1.959	1.694	265	86,5
Acre	2.035	1.892	143	93,0	2.033	1.814	219	89,2
Amazonas	2.946	2.795	151	94,4	2.945	2.586	359	87,3
Roraima	1.835	1.749	86	94,5	1.828	1.591	237	86,3
Pará	2.882	2.438	444	84,1	2.879	2.004	875	69,2
Amapá	1.637	1.522	115	93,0	1.637	1.332	305	81,4
Tocantins	1.738	1.601	137	92,1	1.738	1.515	223	87,2
Maranhão	2.072	1.882	190	90,8	2.071	1.774	297	85,7
Piauí	1.957	1.859	98	95,0	1.957	1.804	153	92,2
Ceará	3.224	2.793	431	86,6	3.220	2.560	660	79,5
Rio Grande do Norte	1.987	1.807	180	90,9	1.987	1.691	296	85,1
Paraíba	2.000	1.960	40	98,0	1.998	1.943	55	97,2
Pernambuco	3.043	2.719	324	89,4	3.043	2.591	452	85,1
Alagoas	1.998	1.901	97	95,1	1.995	1.748	247	87,6
Sergipe	1.955	1.734	221	88,7	1.955	1.553	402	79,4
Bahia	2.880	2.776	104	96,0	2.878	2.641	237	91,4
Minas Gerais	4.110	3.932	178	95,7	4.110	3.779	331	91,9
Espírito Santo	2.031	1.894	137	93,3	2.030	1.724	306	84,9
Rio de Janeiro	4.145	3.801	344	91,7	4.145	3.486	659	84,1
São Paulo	6.209	5.623	586	90,6	6.208	5.305	903	85,5
Paraná	3.257	3.122	135	95,9	3.254	3.012	242	92,6
Santa Catarina	1.975	1.721	254	87,1	1.975	1.623	352	82,2
Rio Grande do Sul	3.314	2.996	318	90,4	3.314	2.913	401	87,9
Mato Grosso do Sul	2.016	1.914	102	94,9	2.013	1.809	204	89,9
Mato Grosso	1.827	1.621	206	88,7	1.825	1.476	349	80,9
Goiás	2.831	2.548	283	90,0	2.828	2.423	405	85,7
Distrito Federal	2.129	1.899	230	89,2	2.129	1.811	318	85,1
Brasil	69.994	64.348	5.646	91,9	69.954	60.202	9.752	86,0

a) Excluídos os domicílios vagos

O processo de calibração – ou pós-estratificação – também merece cuidados especiais na hora das análises. Geralmente, quando não há calibração dos pesos, basta informar ao aplicativo estatístico, no momento da análise, as variáveis que definem os estratos, as UPA, o tipo de desenho e os pesos corrigidos pelas não respostas. Não obstante, quando os pesos passam por um processo de pós-estratificação, deve-se informar, também, as projeções populacionais e as categorias utilizadas para os ajustes dos totais populacionais no processo de calibração dos pesos. Estas informações são disponibilizadas pelo IBGE, juntamente com

o banco de dados; porém, nem todo pacote estatístico capacitado para análise de dados amostrais complexos apresenta essa opção.

No caso da PNS, há um conjunto de informações referente ao domicílio e todos seus moradores – incluídos os pesos com correção de não resposta – sem a calibração e com a calibração, as projeções populacionais e as categorias das projeções populacionais. O mesmo ocorre para o morador selecionado. O importante é saber quando utilizar cada um.

Os pesos finais com o ajuste da calibração devem ser utilizados sempre que não for possível incorporar o processo de pós-estratificação na análise (informando

as projeções populacionais e as categorias das projeções populacionais), pois as estimativas pontuais (médias, totais, proporções) geradas com essa ponderação são corretas.

Para profissionais com pouca experiência em análise de dados provenientes de amostras complexas, encontra-se disponível, no sítio eletrônico da Pesquisa Nacional de Saúde – PNS –,

<http://www.pns.iciet.fiocruz.br/>

o banco de dados e informações necessárias sobre as variáveis que definem o plano amostral e devem ser utilizadas para a análise dos dados =complexos – pelos pacotes estatísticos.

Referências

1. Szwarcwald CL, Malta DC, Pereira CA, Vieira MLFP, Conde WL, Souza Júnior PRB, et al. National Health Survey in Brazil: design and methodology of application. *Cien Saude Coletiva*. 2014 Feb;19(2):333-42.
2. Instituto Brasileiro de Geografia e Estatística. Diretoria de Pesquisas. Coordenação de Trabalho e Rendimento. Sistema Integrado de Pesquisas Domiciliares – SIPD [Internet]. Rio de Janeiro: IBGE; 2007 [citado 2015 fev 2]. (Texto para discussão; 24). Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/sipd/texto_discussao_24.pdf
3. Instituto Brasileiro de Geografia e Estatística. Coordenação de Comunicação Social. Guia do Censo 2010 para jornalistas. Rio de Janeiro: IBGE; 2010. 40 p.
4. Freitas MPS, Lila MF, Azevedo RV, Antonaci GA. Amostra mestra para o Sistema Integrado de Pesquisas Domiciliares [Internet]. Rio de Janeiro: IBGE; 2007 [citado 2015 fev 2]. (Texto para discussão; 23). Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/sipd/texto_discussao_23.pdf
5. Freitas MPS, Antonaci GA. Sistema Integrado de Pesquisas Domiciliares: amostra mestra 2010 e amostra da PNAD contínua. Rio de Janeiro: IBGE; 2014 [citado 2015 mar 28]. (Texto para discussão; 50). Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv86747.pdf>
6. Cochran WG. Sampling techniques. 3th. New York: John Wiley & Sons; 1977.
7. Vasconcellos MTL, Silva PLN, Szwarcwald CL. Aspectos de amostragem da Pesquisa Mundial de Saúde no Brasil. *Cad Saude Publica*. 2005;21 supl 1:589-99.
8. Ruiz CMM, Silva PLN. Explorando alternativas para a calibração dos pesos amostrais da Pesquisa Nacional por Amostra de Domicílios. In: 6º Congreso de la Asociación Latinoamericana de Población [Internet]; 2014 ago 12-15; Lima. Peru: Asociación Latinoamericana de Población; 2014 [citado 2015 fev 2]. Disponível em: http://www.alapop.org/Congreso2014/DOCSFINAIS_PDF/ALAP_2014_FINAL948.pdf.

Contribuição dos autores

Souza-Júnior PRB, Freitas MPS e Antonaci GA participaram da concepção e delineamento do estudo, redação, revisão crítica do conteúdo e aprovação da versão final do manuscrito.

Szwarcwald CL participou da concepção e delineamento do estudo, revisão crítica do conteúdo e aprovação da versão final do manuscrito.

Todos os autores aprovaram a versão final do manuscrito e declaram serem responsáveis por todos os aspectos do trabalho, garantindo sua precisão e integridade.

Recebido em 21/02/2015
Aprovado em 02/04/2015