

## **PESQUISA NACIONAL DE SAÚDE**

### **PLANO AMOSTRAL**

Marcos Paulo Soares de Freitas

**RIO DE JANEIRO**

**Setembro de 2014**

# Sumário

1. Introdução.....	3
2. Objetivos da Pesquisa.....	3
3. População Alvo.....	3
4. Plano Amostral.....	3
5. Tamanho da amostra.....	4
5.1. Tamanho da amostra para realização de exames.....	10
6 .Definição dos fatores de expansão.....	12
6.1. Peso das Unidades Primárias de Amostragem.....	12
6.2. Peso dos domicílios e de todos os seus moradores.....	12
6.3. Peso do morador selecionado.....	15
6.4. Peso para subamostra de exames.....	16
7. Bibliografia.....	16
8. Anexos.....	18

## **1. Introdução**

Este documento descreve o planejamento amostral da Pesquisa Nacional de Saúde, definido em conjunto com técnicos da Fundação Oswaldo Cruz (FIOCRUZ). A pesquisa é parte do Sistema Integrado de Pesquisas Domiciliares (SIPD) do IBGE e utilizou a infraestrutura amostral construída para o SIPD.

São descritos, nas seções a seguir, alguns objetivos da pesquisa que nortearam o planejamento da amostra, a população alvo, o plano amostral, o dimensionamento da amostra e a definição dos fatores de expansão.

## **2. Objetivos da Pesquisa**

A PNS produzirá dados em âmbito nacional sobre a situação de saúde e os estilos de vida da população brasileira, bem como 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.

## **3. População Alvo**

A população alvo é constituída pelos moradores em domicílios particulares permanentes pertencentes à área de abrangência geográfica da pesquisa.

Foi definido como a abrangência geográfica todo o território nacional, dividido nos setores censitários da Base Operacional Geográfica de 2010, excluídas áreas com características especiais e com pouca população. Estas áreas são 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.

## **4. Plano Amostral**

A principal finalidade do plano amostral para esta pesquisa foi permitir a obtenção de estimativas da proporção de pessoas em várias categorias relacionadas com alguns indicadores de interesse. A pesquisa foi domiciliar e o plano amostral empregado foi amostragem conglomerada em três estágios, com estratificação das unidades primárias de amostragem. Os setores censitários ou conjunto de setores formam as unidades primárias de amostragem (UPAs), os domicílios são as unidades de segundo estágio e os moradores com 18 anos ou mais de idade definem as unidades de terceiro estágio.

Como parte integrante do SIPD, as UPAs da pesquisa foram obtidas da Amostra Mestra, que é a estrutura amostral do sistema e é composta por UPAs selecionadas com um método probabilístico e da qual é possível selecionar uma subamostra para as pesquisas que integram o sistema. Portanto, a estratificação das UPAs nesta pesquisa foi a mesma adotada para a Amostra Mestra. Para mais detalhes ver Freitas et al.(2014). A seleção da subamostra de UPAs foi feita por amostragem aleatória simples.

No segundo estágio foi selecionado por amostragem aleatória simples um número fixo de domicílios particulares permanentes em cada UPA selecionada no primeiro estágio. A seleção foi feita do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) em sua mais recente atualização. Nestes domicílios selecionados foram investigadas características do domicílio, características de todos os moradores e características de todos os idosos.

E dentro de cada domicílio selecionado, um morador com 18 anos ou mais de idade foi selecionado para responder ao questionário específico, também por amostragem aleatória simples, da lista de moradores construída no momento da entrevista.

## **5. Tamanho da amostra**

O tamanho da amostra foi definido considerando o nível de precisão desejado para as estimativas de alguns indicadores de interesse, que são basicamente proporções de pessoas em determinadas categorias.

Como os indicadores referem-se a grupos populacionais diferentes, também foi preciso avaliar a proporção de domicílios que possuíam pessoas nestes grupos, com base nos dados do Censo 2010, para que fosse possível definir o tamanho de amostra necessário para estimar com a precisão requerida.

Por falta de informações para toda a população sobre as características envolvidas nos indicadores, os cálculos iniciais foram baseados em amostragem aleatória simples (AAS), considerando que os indicadores são proporções de pessoas com a característica de interesse, e foram efetuados a fim de definir o tamanho de amostra, fixando-se os níveis de precisão desejados para as estimativas dos indicadores e para os diversos níveis geográficos em que se pretende estimá-los.

O plano amostral que será adotado na pesquisa não será amostragem aleatória simples de pessoas, por isso foi feito um ajuste no tamanho de amostra obtido considerando os valores do efeito de plano amostral (EPA). Esta medida, EPA, indica o quanto o plano amostral por conglomerados é menos eficiente (maior variância) que a amostragem aleatória simples, por isso os tamanhos da amostra obtidos para AAS foram aumentados para que se alcance a mesma precisão. Para mais detalhes sobre a definição teórica do efeito de plano amostral ver Lila e Freitas (2006).

Os EPAs utilizados nos cálculos foram estimados com base nos dados do suplemento saúde da PNAD 2008 pelos técnicos da FIOCRUZ.

A tabela 1 mostra os indicadores, os grupos populacionais relacionados, os coeficientes de variação desejados, o valor esperado para cada um dos indicadores e os EPAs estimados.

As fórmulas utilizadas, dentro de cada domínio, foram as seguintes:

$$n_{AAS} = \frac{N}{N-1} \cdot P \cdot Q \cdot \frac{1}{CV^2 \cdot P^2 + \frac{P \cdot Q}{N-1}}$$

$$n_{AC} = n_{AAS} \cdot EPA$$

onde

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

$N$  é o número total de pessoas

$P$  é a proporção de interesse

$$Q = 1 - P$$

$CV$  é o coeficiente de variação desejado da estimativa de proporção

$n_{AC}$  é o tamanho de amostra sob amostragem conglomerada

$EPA$  é o efeito de plano amostral.

Os dados utilizados nos cálculos foram obtidos no Censo Demográfico 2010 e na PNAD 2008.

Os primeiros tamanhos de amostra foram calculados como os necessários para estimar os indicadores com o coeficiente de variação desejado para cada um dos níveis geográficos inicialmente pensados como domínios de divulgação. São eles: Brasil, Grande Região, Unidade da Federação (UF), Região Metropolitana (RM), Capital e restante da UF. A tabela 2 apresenta os tamanhos iniciais calculados para cada um dos indicadores.

Os tamanhos mostrados na tabela deveriam ser aplicados a cada um dos domínios, fazendo com que nas UFs sem RM que contém o município da capital, o tamanho total da amostra seria aproximadamente o dobro do indicado, pois possuem dois domínios de divulgação, capital e restante da UF. E nas UFs com RM, o tamanho seria um pouco maior que o dobro, pois a RM é também um domínio de divulgação.

Com os resultados encontrados e pelo que foi exposto anteriormente, alguns ajustes foram feitos nos tamanhos e nos níveis geográficos pensados como domínios de divulgação no planejamento.

Tabela 1 – Grupo populacional relacionado ao indicador, proporção de domicílios com pessoas no grupo populacional, valor esperado do indicador, coeficiente de variação desejado para a estimativa do indicador e EPA estimado, segundo o indicador de interesse

Indicador	Grupo populacional	Proporção de domicílios com pessoas no grupo populacional – Censo 2010	Valor esperado do indicador (%)	CV desejado (%)	EPA estimado
1. Prevalência de diabetes 35 anos e mais	Adultos 35 anos e mais selecionados	0,6	8	12,8	1,4
2. Prevalência de hipertensão 35 anos e mais	Adultos 35 anos e mais selecionados	0,6	30	8,5	1,6
3. Prevalência de depressão 35 anos mais	Adultos 35 anos e mais selecionados	0,6	8	12,8	1,9
4. Taxa de internação último ano	Todos os moradores do domicílio	1,0	7	14,6	2,9
5. Uso nas últimas 2 semanas	Todos os moradores do domicílio	1,0	14	7,3	4,6
6. Cobertura de plano de saúde privado	Todos os moradores do domicílio	1,0	26	5,9	10,4
7. Cobertura de exame preventivo de câncer de colo de útero entre mulheres 25 a 59 anos	Mulheres de 25-59 anos selecionadas	0,4	80	3,2	2,2
8. Cobertura de mamografia entre mulheres 50 e mais	Mulheres de 50-69 anos selecionadas	0,1	65	5,5	1,8
9. Prevalência de fumo	Adultos selecionados	1,0	17	9,0	1,8
10. Prevalência de sobrepeso/ obesidade	Adultos selecionados	1,0	12	12,8	1,8
11. Prevalência de sedentarismo	Todos os moradores do domicílio	1,0	15	10,2	4,1
12. Uso abusivo de álcool	Adultos selecionados	1,0	7	14,6	1,8
13. % que sofreu violência com lesões corporais	Adultos selecionados	1,0	2	25,5	2,8
14. % de idosos com limitações (problemas de funcionalidade)	Todos os idosos do domicílio	0,4	15	13,6	2,1
15. Cobertura de vacinação em crianças com menos de 2 anos	Todas as crianças com menos de 2 anos do domicílio	0,1	80	3,2	2,0

Tabela 2 – Tamanho de amostra de domicílios para estimar os indicadores com o cv estipulado em cada domínio de divulgação, segundo o indicador

Indicador	Tamanho de amostra de domicílios
1. Prevalência de diabetes 35 anos e mais	2.725
2. Prevalência de hipertensão 35 anos e mais	1.669
3. Prevalência de depressão 35 anos mais	2.778
4. Taxa de internação último ano	387
5. Uso nas últimas 2 semanas	1.303
6. Cobertura de plano de saúde privado	1.857
7. Cobertura de exame preventivo de câncer de colo de útero entre mulheres 25 a 59 anos	2.904
8. Cobertura de mamografia entre mulheres 50 e mais	2.645
9. Prevalência de fumo	1.872
10. Prevalência de sobrepeso/ obesidade	1.347
11. Prevalência de sedentarismo	525
12. Uso abusivo de álcool	1.833
13. % que sofreu violência com lesões corporais	3.787
14. % de idosos com limitações (problemas de funcionalidade)	1.467
15. Cobertura de vacinação em crianças com menos de 2 anos	4.167

Após várias avaliações, optou-se por determinar o tamanho mínimo da amostra de domicílios por UF em 1.800, o que permitiria estimar 9 dos indicadores com a precisão desejada inicialmente e os demais com precisão inferior a desejada, porém ainda dentro de faixas de cvs considerados aceitáveis, proporcionando estimativas ainda precisas.

Com esta definição, e como são esperados no mínimo dois domínios de divulgação nas UFs, já comentado neste texto, ficou determinado que o tamanho mínimo de amostra de domicílios por domínio seria de 900 domicílios.

Para a definição do tamanho da amostra de UPAs ( $m_d$ ), foi fixado o número de domicílios selecionados em cada UPA, e então dividido o tamanho da amostra de domicílios por este número, que foi de 10 ou de 14, dependendo do domínio,  $m_d = \frac{n_{ACd}}{10}$  ou  $m_d = \frac{n_{ACd}}{14}$ . Essa diferença se deveu a limitação no tamanho da amostra de UPAs pela Amostra Mestra. Em alguns domínios não seria possível atingir o tamanho total de domicílios, selecionando apenas 10, pois o número de UPAs seria maior que a Amostra Mestra, e, como dito na seção anterior, a amostra de

UPAs para essa pesquisa é uma subamostra da Amostra Mestra, portanto deveria ser no máximo igual.

Os tamanhos finais da amostra por UF são apresentados na tabela 3 e em anexo os tamanhos por domínio.

Tabela 3 – Tamanho planejado e selecionado da amostra para a Pesquisa Nacional de Saúde segundo os domínios de interesse

Domínio	Tamanho planejado da amostra de UPAs	Tamanho da amostra		
		Selecionado de UPAs	Esperado de Domicílios com entrevista realizada (23% de perda)	Selecionado de Domicílios
<b>Brasil</b>	6.081	6.069	62.658	81.357
<b>Norte</b>	1.169	1.161	13.578	17.553
<b>Rondônia</b>	130	129	1.806	2.322
<b>Acre</b>	180	180	1.800	2.340
<b>Amazonas</b>	246	245	2.594	3.365
<b>Roraima</b>	130	124	1.736	2.232
<b>Pará</b>	246	246	2.652	3.438
<b>Amapá</b>	108	107	1.498	1.926
<b>Tocantins</b>	129	130	1.492	1.930
<b>Nordeste</b>	1.916	1.916	19.160	24.908
<b>Maranhão</b>	180	181	1.810	2.353
<b>Piauí</b>	180	180	1.800	2.340
<b>Ceará</b>	290	290	2.900	3.770
<b>Rio Grande do Norte</b>	180	179	1.790	2.327
<b>Paraíba</b>	180	182	1.820	2.366
<b>Pernambuco</b>	276	279	2.790	3.627
<b>Alagoas</b>	180	180	1.800	2.340
<b>Sergipe</b>	180	179	1.790	2.327
<b>Bahia</b>	270	266	2.660	3.458
<b>Sudeste</b>	1.446	1.451	14.510	18.863
<b>Minas Gerais</b>	366	366	3.660	4.758
<b>Espírito Santo</b>	180	181	1.810	2.353
<b>Rio de Janeiro</b>	360	365	3.650	4.745
<b>São Paulo</b>	540	539	5.390	7.007
<b>Sul</b>	760	767	7.670	9.971
<b>Paraná</b>	290	287	2.870	3.731
<b>Santa Catarina</b>	180	186	1.860	2.418
<b>Rio Grande do Sul</b>	290	294	2.940	3.822
<b>Centro-oeste</b>	790	774	7.740	10.062
<b>Mato Grosso do Sul</b>	180	179	1.790	2.327
<b>Mato Grosso</b>	180	162	1.620	2.106
<b>Goiás</b>	250	253	2.530	3.289
<b>Distrito Federal</b>	180	180	1.800	2.340

Na tabela constam os tamanhos de amostra de UPAs planejados e os efetivamente obtidos após a alocação pelos estratos da pesquisa, o que ocasionou a redução em alguns casos. Constam, também, os tamanhos de amostra de domicílios selecionados e o tamanho esperado de domicílios



com entrevista realizada, considerando a perda total de cerca de 23% das entrevistas, percentual observado em outras pesquisas do IBGE, como a Especial sobre Tabagismo de 2008, em que também foi selecionada uma pessoa no domicílio.

Para avaliar o impacto na precisão das estimativas dos indicadores obtido pela decisão de fixar os tamanhos para permitir estimar no nível de UF, foram calculados os cvs esperados para cada um dos indicadores em todos dos domínios de divulgação e feita a avaliação dos resultados, classificando os cvs em faixas.

A tabela 4 contém, para cada indicador, o percentual de domínios por classes de cvs. As classes escolhidas para esta avaliação foram “até 15%”, “mais de 15% até 30%” e “mais de 30%”, que são classes adotadas em divulgações do IBGE. Os domínios considerados foram os já citados, Grande Região, Unidade da Federação, Região Metropolitana, Capital e Restante da UF.

Tabela 4 – Percentual de domínios de divulgação por classes de coeficientes de variação esperados, segundo o indicador de interesse

Indicador	Percentual de domínios de divulgação		
	Classes de coeficientes de variação esperados		
	Até 15%	Mais de 15% a 30%	Mais de 30%
1. Prevalência de diabetes 35 anos e mais	7,8	77,7	14,6
2. Prevalência de hipertensão 35 anos e mais	91,3	8,7	0,0
3. Prevalência de depressão 35 anos mais	4,9	45,6	49,5
4. Taxa de internação último ano	100,0	0,0	0,0
5. Uso nas últimas 2 semanas	100,0	0,0	0,0
6. Cobertura de plano de saúde privado	100,0	0,0	0,0
7. Cobertura de exame preventivo de câncer de colo de útero entre mulheres 25 a 59 anos	100,0	0,0	0,0
8. Cobertura de mamografia entre mulheres 50 e mais	4,9	46,6	48,5
9. Prevalência de fumo	100,0	0,0	0,0
10. Prevalência de sobrepeso/ obesidade	100,0	0,0	0,0
11. Prevalência de sedentarismo	100,0	0,0	0,0
12. Uso abusivo de álcool	99,0	1,0	0,0
13. % que sofreu violência com lesões corporais	23,3	75,7	1,0
14. % de idosos com limitações (problemas de funcionalidade)	12,6	85,4	1,9
15. Cobertura de vacinação em crianças com menos de 2 anos	41,7	57,3	1,0

Observando os dados da tabela 3, espera-se que para quase todos os indicadores a maioria dos domínios tenha estimativa com cv abaixo de 30%, que é considerada publicável pela classificação escolhida. Apenas os indicadores 3 e 8 apresentaram uma quantidade elevada de

domínios com cv esperado superior a 30%. Em todos os casos, em especial nestes dois indicadores, deve-se observar os resultados da pesquisa para decidir a divulgação ou não dos indicadores para domínios que sejam menores, podendo ser divulgados para níveis geográficos mais agregados.

### **5.1. Tamanho da amostra para realização de exames**

A Pesquisa Nacional de Saúde contou com uma etapa que constava da realização de exames laboratoriais para as pessoas selecionadas nos domicílios. Devido ao alto custo desta operação, ficou definido que os exames seriam realizados apenas em uma parte da amostra.

Em função desta decisão, foram selecionados cerca de 25% das UPAs em cada estrato para compor a subamostra para exames, e todas as pessoas selecionadas nos domicílios selecionados netas UPAs foram encaminhadas para a realização dos exames.

A seleção da subamostra foi feita com probabilidade proporcional ao inverso da distância do município onde se localiza a UPA e o município mais próximo com 80.000 habitantes ou mais. Isto se deu, pois estes municípios possuem melhor infraestrutura para a realização dos exames e, dando maior probabilidade às UPAs mais próximas destes municípios, pretendia-se reduzir os custos da operação.

Com esta redução na amostra total, não é esperado que seja possível divulgar estimativas das características investigadas pelos exames para os domínios de interesse já comentados neste texto. Provavelmente o nível geográfico mais detalhado que será permitido produzir estimativas com precisão desejada serão as Grandes Regiões. O que foi dito em relação aos indicadores de interesse utilizados para o planejamento da amostra também vale para os exames, será necessário fazer uma avaliação com os dados da pesquisa para decidir para quais domínios serão divulgados os resultados.

Os tamanhos da subamostra por UF são indicados na tabela 5.

Tabela 5 – Tamanho da subamostra de UPAs e pessoas para realização de exames, segundo os domínios de interesse

Domínio	Subamostra para exames	
	UPAs	Domicílios/Pessoas
<b>Brasil</b>	1.727	17.902
<b>Norte</b>	316	3.792
Rondônia	37	518
Acre	46	460
Amazonas	69	730
Roraima	35	490
Pará	66	712
Amapá	28	392
Tocantins	35	490
<b>Nordeste</b>	554	5.540
Maranhão	56	560
Piauí	52	520
Ceará	85	850
Rio Grande do Norte	48	480
Paraíba	54	540
Pernambuco	77	770
Alagoas	50	500
Sergipe	49	490
Bahia	83	830
<b>Sudeste</b>	413	4.130
Minas Gerais	112	1120
Espírito Santo	52	520
Rio de Janeiro	101	1010
São Paulo	148	1480
<b>Sul</b>	224	2.240
Paraná	84	840
Santa Catarina	56	560
Rio Grande do Sul	84	840
<b>Centro-oeste</b>	220	2.200
Mato Grosso do Sul	50	500
Mato Grosso	47	470
Goiás	75	750
Distrito Federal	48	480

## 6 .Definição dos fatores de expansão

Na PNS foi preciso definir fatores de expansão ou pesos amostrais para as UPAs, para os domicílios, todos os seus moradores e todos os idosos, e para o morador selecionado. Também foi preciso definir o peso para a subamostra para exames.

### 6.1. Peso das Unidades Primárias de Amostragem

Os pesos das UPAs foram calculados considerando a probabilidade de seleção da UPA para a Amostra Mestra e a probabilidade de seleção para a amostra da pesquisa. Foram definidos da seguinte maneira

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

onde

$w_{hi}$  é o peso básico da UPA  $i$  do estrato  $h$  na PNS

$m_h$  é o número de UPAs 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 do estrato  $h$ , dados atualizados do CNEFE (Cadastro Nacional de Endereços para Fins Estatísticos) no momento da seleção das UPAs para a Amostra Mestra

$N_h$  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 (Cadastro Nacional de Endereços para Fins Estatísticos) no momento da seleção das UPAs para a Amostra Mestra

$m_h^{PNS}$  é o número de UPAs selecionadas no estrato  $h$  para a PNS

### 6.2. Peso dos domicílios e de todos os seus moradores

Os pesos para os domicílios e todos os seus moradores, utilizados para a estimação das características investigadas para todos os moradores e para todos os idosos, foram definidos levando-se em conta o peso da UPA correspondente e ajustes para correção de não respostas e também para calibrar as estimativas com totais populacionais conhecidos de outras fontes.

Os domicílios foram selecionados com probabilidade igual em cada UPA, assim o peso dos domicílios dentro da UPA é dado por

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

onde

$w_{j|hi}$  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

$n_h$  é o número de domicílios selecionados na UPA  $i$  do estrato  $h$

E o peso básico do domicílio é obtido pela expressão

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

Após a determinação do peso básico, foi feito um ajuste para compensar as perdas de entrevistas por não resposta, ou seja, entrevistas não realizadas por recusa do informante, por não contato com o morador ou por outro motivo para perda em domicílios ocupados. Esses números de domicílios em cada uma das situações de coleta, além das taxas de resposta e de perdas são apresentados na tabela 6.

Foi utilizada a seguinte fórmula

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

onde

$n_h^*$  é o número de domicílios selecionados com morador na UPA  $i$  do estrato  $h$

$n_h^{**}$  é o número de domicílios selecionados com entrevista realizada na UPA  $i$  do estrato  $h$

Outro ajuste feito nos pesos foi para calibrar os resultados provenientes da pesquisa a totais conhecidos de outras fontes. Os pesos foram ajustados para 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.

A calibração foi feita para a população do dia 27 de julho de 2013, pela seguinte expressão

$$w_{hij}^{**} = w_{hij}^* * \frac{P_a^{tri}}{\hat{P}_a^{tri}}$$

onde

$P_a^{tri}$  é a estimativa populacional produzida pela COPIS para o nível geográfico  $a$  para do dia 27 de julho de 2013;

$\hat{P}_a^{tri}$  é a estimativa populacional obtida com os dados da pesquisa para o nível geográfico  $a$

Tabela 6 – Número de domicílios na amostra planejada, número de domicílios selecionados por situação de coleta (em UPAs não coletadas, fora do âmbito, no âmbito com entrevista realizada e sem entrevista realizada), taxas de perda total, de excesso de cobertura e de resposta, segundo Unidades da Federação - PNS 2013

Unidade da Federação	Número de domicílios na amostra planejada	Número de domicílios selecionados						Taxa de perda total (%) ((A+B+C3)/T)	Taxa de excesso de cobertura (%) (B/T)	Taxa de resposta (%) (C2/C1)
		Total (T)	Em UPAs com amostra não coletada (A)	Fora do âmbito (B)	No âmbito (ocupados)					
					Total (C1)	Com entrevista realizada (C2)	Sem entrevista realizada (C3)			
Brasil	81.357	81.254	67	11.193	69.994	64.348	5.646	20,8	13,8	91,9
Rondônia	2.322	2.322	0	361	1.961	1.849	112	20,4	15,5	94,3
Acre	2.340	2.340	0	305	2.035	1.892	143	19,1	13,0	93,0
Amazonas	3.365	3.352	18	388	2.946	2.795	151	16,6	11,6	94,4
Roraima	2.232	2.226	18	373	1.835	1.749	86	21,4	16,9	94,5
Pará	3.438	3.429	18	529	2.882	2.438	444	28,9	15,5	84,1
Amapá	1.926	1.918	0	281	1.637	1.522	115	20,6	14,7	93,0
Tocantins	1.930	1.930	0	192	1.738	1.601	137	17,0	9,9	92,1
Maranhão	2.353	2.353	0	281	2.072	1.882	190	20,0	11,9	90,8
Piauí	2.340	2.340	0	383	1.957	1.859	98	20,6	16,4	95,0
Ceará	3.770	3.770	0	546	3.224	2.793	431	25,9	14,5	86,6
Rio Grande do Norte	2.327	2.327	0	340	1.987	1.807	180	22,3	14,6	90,9
Paraíba	2.366	2.366	0	366	2.000	1.960	40	17,2	15,5	98,0
Pernambuco	3.627	3.598	0	555	3.043	2.719	324	24,4	15,4	89,4
Alagoas	2.340	2.340	0	342	1.998	1.901	97	18,8	14,6	95,1
Sergipe	2.327	2.327	0	372	1.955	1.734	221	25,5	16,0	88,7
Bahia	3.458	3.458	13	565	2.880	2.776	104	19,7	16,4	96,0
Minas Gerais	4.758	4.749	0	639	4.110	3.932	178	17,2	13,5	95,7
Espírito Santo	2.353	2.351	0	320	2.031	1.894	137	19,4	13,6	93,3
Rio de Janeiro	4.745	4.737	0	592	4.145	3.801	344	19,8	12,5	91,7
São Paulo	7.007	7.007	0	798	6.209	5.623	586	19,8	11,4	90,6
Paraná	3.731	3.723	0	466	3.257	3.122	135	16,1	12,5	95,9
Santa Catarina	2.418	2.415	0	440	1.975	1.721	254	28,7	18,2	87,1
Rio Grande do Sul	3.822	3.814	0	500	3.314	2.996	318	21,4	13,1	90,4
Mato Grosso do Sul	2.327	2.327	0	311	2.016	1.914	102	17,7	13,4	94,9
Mato Grosso	2.106	2.106	0	279	1.827	1.621	206	23,0	13,2	88,7
Goiás	3.289	3.289	0	458	2.831	2.548	283	22,5	13,9	90,0
Distrito Federal	2.340	2.340	0	211	2.129	1.899	230	18,8	9,0	89,2

Nota1: A diferença entre o número de domicílios na amostra planejada e o número de domicílios selecionados foi decorrente da menor quantidade de domicílios disponíveis para seleção no cadastro de algumas UPAs

Nota2: Os domicílios em UPAs com amostra não coletada foram considerados no cálculo das taxas de excesso de cobertura proporcionalmente à quantidade de domicílios fora do âmbito

Nota3: Os domicílios em UPAs com amostra não coletada foram considerados no cálculo das taxas de resposta proporcionalmente à quantidade de domicílios no âmbito

### 6.3. Peso do morador selecionado

Para o cálculo do peso do morador selecionado foi preciso definir a probabilidade de seleção do morador dentro do domicílio. A seleção foi feita por amostragem aleatória simples, deste modo o peso básico é dado por

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

E o peso com ajuste de não resposta do domicílio inteiro é dado por

$$w'_{hijk} = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}^*} * \frac{n_{hi}^*}{n_{hi}^{**}} * O_{hij} \quad \text{onde}$$

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

Como após a seleção do morador houve perda de entrevista, a correção dos pesos para compensar a não resposta foi diferente da realizada para os pesos dos domicílios. Em um primeiro momento o ajuste foi feito de maneira equivalente a já descrita, dentro da UPA, considerando os domicílios com entrevista do morador realizada. Após análise das estimativas de pessoas por sexo e das características dos moradores não respondentes, decidiu-se fazer o ajuste por sexo, pois percebeu-se que a perda foi maior entre os homens que entre as mulheres.

As quantidades de domicílios por situação de coleta do questionário específico do morador selecionado e as taxas de respostas e de perdas são mostradas na tabela 7.

As fórmulas usadas foram

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

$$w_{hijk}^F = \frac{1}{m_h} * \frac{N_h}{N_{hi}} * \frac{m_h}{m_h^{PNS}} * \frac{N_{hi}^*}{n_{hi}^*} * \frac{n_{hi}^*}{n_{hi}^{**}} * O_{hij} * \frac{\sum_j^{n_{hijk}} w'_{hijk} * \alpha_{hijk}^F}{\sum_i^{n_{hijk}} w'_{hijk} * \alpha_{hijk}^F} \quad \text{onde}$$

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

$\alpha_{hijk}^M$  indica se o morador selecionado no domicílio  $j$  da UPA  $i$  do estrato  $h$  é do sexo masculino

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

Tabela 7 – Número de domicílios com morador de 18 anos ou mais de idade na amostra planejada, número de domicílios selecionados por situação de coleta (em UPAs não coletadas, fora do âmbito, no âmbito com entrevista realizada e sem entrevista realizada), taxas de perda total, de excesso de cobertura e de resposta, segundo Unidades da Federação - PNS 2013

Unidade da Federação	Número domicílios com morador de 18 anos ou mais de idade na amostra planejada	Número de domicílios selecionados						Taxa de perda total (%) ((A+B+C3)/T)	Taxa de excesso de cobertura (%) (B/T)	Taxa de resposta (%) (C2/C1)
		Total (T)	Em UPAs com amostra não coletada (A)	Fora do âmbito (B)	No âmbito					
					Total (C1)	Com entrevista realizada (C2)	Sem entrevista realizada (C3)			
Brasil	81.357	81.254	67	11.233	69.954	60.202	9.752	25,9	13,8	86,0
Rondônia	2.322	2.322	0	363	1.959	1.694	265	27,0	15,6	86,5
Acre	2.340	2.340	0	307	2.033	1.814	219	22,5	13,1	89,2
Amazonas	3.365	3.352	18	389	2.945	2.586	359	22,9	11,7	87,3
Roraima	2.232	2.226	18	380	1.828	1.591	237	28,5	17,2	86,3
Pará	3.438	3.429	18	532	2.879	2.004	875	41,6	15,6	69,2
Amapá	1.926	1.918	0	281	1.637	1.332	305	30,6	14,7	81,4
Tocantins	1.930	1.930	0	192	1.738	1.515	223	21,5	9,9	87,2
Maranhão	2.353	2.353	0	282	2.071	1.774	297	24,6	12,0	85,7
Piauí	2.340	2.340	0	383	1.957	1.804	153	22,9	16,4	92,2
Ceará	3.770	3.770	0	550	3.220	2.560	660	32,1	14,6	79,5
Rio Grande do N	2.327	2.327	0	340	1.987	1.691	296	27,3	14,6	85,1
Paraíba	2.366	2.366	0	368	1.998	1.943	55	17,9	15,6	97,2
Pernambuco	3.627	3.598	0	555	3.043	2.591	452	28,0	15,4	85,1
Alagoas	2.340	2.340	0	345	1.995	1.748	247	25,3	14,7	87,6
Sergipe	2.327	2.327	0	372	1.955	1.553	402	33,3	16,0	79,4
Bahia	3.458	3.458	13	567	2.878	2.641	237	23,6	16,5	91,4
Minas Gerais	4.758	4.749	0	639	4.110	3.779	331	20,4	13,5	91,9
Espírito Santo	2.353	2.351	0	321	2.030	1.724	306	26,7	13,7	84,9
Rio de Janeiro	4.745	4.737	0	592	4.145	3.486	659	26,4	12,5	84,1
São Paulo	7.007	7.007	0	799	6.208	5.305	903	24,3	11,4	85,5
Paraná	3.731	3.723	0	469	3.254	3.012	242	19,1	12,6	92,6
Santa Catarina	2.418	2.415	0	440	1.975	1.623	352	32,8	18,2	82,2
Rio Grande do Su	3.822	3.814	0	500	3.314	2.913	401	23,6	13,1	87,9
Mato Grosso do	2.327	2.327	0	314	2.013	1.809	204	22,3	13,5	89,9
Mato Grosso	2.106	2.106	0	281	1.825	1.476	349	29,9	13,3	80,9
Goiás	3.289	3.289	0	461	2.828	2.423	405	26,3	14,0	85,7
Distrito Federal	2.340	2.340	0	211	2.129	1.811	318	22,6	9,0	85,1

Nota1: A diferença entre o número de domicílios na amostra planejada e o número de domicílios selecionados foi decorrente da menor quantidade de domicílios disponíveis para seleção no cadastro de algumas UPAs

Nota2: Os domicílios em UPAs com amostra não coletada foram considerados no cálculo das taxas de excesso de cobertura proporcionalmente à quantidade de domicílios fora do âmbito

Nota3: Os domicílios em UPAs com amostra não coletada foram considerados no cálculo das taxas de resposta proporcionalmente à quantidade de domicílios no âmbito



Como é feita uma amostra aleatória simples de um morador dentro do domicílio, é natural que, por conta da aleatoriedade de seleção, 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.

No entanto, os moradores dos domicílios formam uma amostra muito maior que os moradores selecionados e portanto, uma estimativa mais precisa destes totais populacionais. Para que estas estimativas fossem 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 de 18 a 24 anos, de 25 a 39 anos, de 40 a 59 anos e mais de 60 anos.

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

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

onde:

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

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

#### 6.4. Peso para subamostra de exames

Os fatores de expansão que serão associados aos moradores selecionados para a subamostra de exames serão calculados de maneira semelhante aos pesos já definidos anteriormente. Primeiramente, considerando o inverso da probabilidade de seleção para a subamostra, serão definidos os pesos básicos. Em seguida serão feitos os ajustes para compensar as perdas de entrevista por não resposta.

Como a coleta dos dados relacionados à subamostra de exames está em fase de finalização e o número de realizações tem sido, até agora, muito baixo, será preciso esperar o encerramento da coleta para que seja feita uma avaliação do aproveitamento da amostra, para só então serem definidos os pesos finais.

## 7. Bibliografia

ALBIERI, S. [1999]. *Apresentação da precisão de estimativas nas tabelas de pesquisas por amostragem do IBGE*. Instituto Brasileiro de Geografia e Estatística (IBGE) – Rio de Janeiro, Brasil. 16 de Junho 1999 (documento interno).

COCHRAN, W.G. *Sampling Techniques*. 3rd. ed. New York: John Wiley, 1977.

FREITAS, M.P.S.; ANTONACI, G.A. *Sistema Integrado de Pesquisas Domiciliares - Amostra Mestra 2010 e Amostra da PNAD Contínua*. Rio de Janeiro: IBGE, 2014. (Texto para discussão, nº 50)

FREITAS, M.P.S.; LILA, M.F.; AZEVEDO, R.V.; ANTONACI, G.A. *Amostra Mestra para o Sistema Integrado de Pesquisas Domiciliares*. Rio de Janeiro: IBGE, 2007. (Texto para discussão, nº 23)

KISH, L. *Survey Sampling*. Nova Iorque: John Wiley & Sons, 1965

LILA, M.F., FREITAS, M.P.S, Uma medida de homogeneidade da migração nos municípios brasileiros. Rio de Janeiro, IBGE, 2006.

## 8.Anexos

Tabela A1 – Tamanho planejado e selecionado da amostra para a Pesquisa Nacional de Saúde segundo os domínios de interesse (continua)

Domínio	Tamanho planejado da amostra de UPAs	Tamanho da amostra		
		Selecionado de UPAs	Esperado de Domicílios com entrevista realizada (23% de perda)	Selecionado de Domicílios
<b>Brasil</b>	6.081	6.069	62.986	81.357
<b>Norte</b>	1.169	1.161	13.578	17.553
<b>Rondônia</b>	130	129	1.806	2.322
Porto Velho	65	65	910	1.170
Restante	65	64	896	1.152
<b>Acre</b>	180	180	1.800	2.340
Rio Branco	90	90	900	1.170
Restante	90	90	900	1.170
<b>Amazonas</b>	246	245	2.594	3.365
RM	156	155	1.694	2.195
Manaus	120	119	1.190	1.547
Resto RM	36	36	504	648
Restante	90	90	900	1.170
<b>Roraima</b>	130	124	1.736	2.232
Boa Vista	70	70	980	1.260
Restante	60	54	756	972
<b>Pará</b>	246	246	2.652	3.438
RM	156	156	1.752	2.268
Belém	108	108	1.080	1.404
Resto RM	48	48	672	864
Restante	90	90	900	1.170
<b>Amapá</b>	108	107	1.498	1.926
RM	84	84	1.176	1.512
Macapá	60	60	840	1.080
Resto RM	24	24	336	432
Restante	24	23	322	414
<b>Tocantins</b>	129	130	1.492	1.930
Palmas	48	48	672	864
Restante	81	82	820	1.066

Tabela A1 – Tamanho planejado e selecionado da amostra para a Pesquisa Nacional de Saúde segundo os domínios de interesse (continua)

Domínio	Tamanho planejado da amostra de UPAs	Tamanho da amostra		
		Selecionado de UPAs	Esperado de Domicílios com entrevista realizada (23% de perda)	Selecionado de Domicílios
<b>Nordeste</b>	1.916	1.916	19.160	24.908
<b>Maranhão</b>	180	181	1.810	2.353
RM	110	110	1.100	1.430
São Luís	90	90	900	1.170
Resto RM	20	20	200	260
Restante	70	71	710	923
<b>Piauí</b>	180	180	1.800	2.340
Teresina	84	84	840	1.092
Restante	96	96	960	1.248
<b>Ceará</b>	290	290	2.900	3.770
RM	200	199	1.990	2.587
Fortaleza	110	109	1.090	1.417
Resto RM	90	90	900	1.170
Restante	90	91	910	1.183
<b>Rio Grande do Norte</b>	180	179	1.790	2.327
RM	104	104	1.040	1.352
Natal	84	84	840	1.092
Resto RM	20	20	200	260
Restante	76	75	750	975
<b>Paraíba</b>	180	182	1.820	2.366
RM	110	111	1.110	1.443
João Pessoa	90	91	910	1.183
Resto RM	20	20	200	260
Restante	70	71	710	923
<b>Pernambuco</b>	276	279	2.790	3.627
RM	186	185	1.850	2.405
Recife	96	96	960	1.248
Resto RM	90	89	890	1.157
Restante	90	94	940	1.222
<b>Alagoas</b>	180	180	1.800	2.340
RM	110	109	1.090	1.417
Maceió	90	89	890	1.157
Resto RM	20	20	200	260
Restante	70	71	710	923
<b>Sergipe</b>	180	179	1.790	2.327
RM	92	92	920	1.196
Aracaju	72	72	720	936
Resto RM	20	20	200	260
Restante	88	87	870	1.131
<b>Bahia</b>	270	266	2.660	3.458
RM	180	180	1.800	2.340
Salvador	120	120	1.200	1.560
Resto RM	60	60	600	780
Restante	90	86	860	1.118

Tabela A1 – Tamanho planejado e selecionado da amostra para a Pesquisa Nacional de Saúde segundo os domínios de interesse (continua)

Domínio	Tamanho planejado da amostra de UPAs	Tamanho da amostra		
		Selecionado de UPAs	Esperado de Domicílios com entrevista realizada (23% de perda)	Selecionado de Domicílios
<b>Sudeste</b>	1.446	1.451	14.510	18.863
<b>Minas Gerais</b>	366	366	3.660	4.758
RM	246	246	2.460	3.198
Belo Horizonte	156	156	1.560	2.028
Resto RM	90	90	900	1.170
RIDE DF	20	20	200	260
Restante	100	100	1.000	1.300
<b>Espírito Santo</b>	180	181	1.810	2.353
RM	110	112	1.120	1.456
Vitória	84	84	840	1.092
Resto RM	26	28	280	364
Restante	70	69	690	897
<b>Rio de Janeiro</b>	360	365	3.650	4.745
RM	270	275	2.750	3.575
Rio de Janeiro	180	181	1.810	2.353
Resto RM	90	94	940	1.222
Restante	90	90	900	1.170
<b>São Paulo</b>	540	539	5.390	7.007
RM	360	361	3.610	4.693
São Paulo	270	271	2.710	3.523
Resto RM	90	90	900	1.170
Restante	180	178	1.780	2.314
<b>Sul</b>	760	767	7.670	9.971
<b>Paraná</b>	290	287	2.870	3.731
RM	200	199	1.990	2.587
Curitiba	110	109	1.090	1.417
Resto RM	90	90	900	1.170
Restante	90	88	880	1.144
<b>Santa Catarina</b>	180	186	1.860	2.418
RM	110	111	1.110	1.443
Florianópolis	84	84	840	1.092
Resto RM	26	27	270	351
Restante	70	75	750	975
<b>Rio Grande do Sul</b>	290	294	2.940	3.822
RM	200	201	2.010	2.613
Porto Alegre	110	111	1.110	1.443
Resto RM	90	90	900	1.170
Restante	90	93	930	1.209

Tabela A1 – Tamanho planejado e selecionado da amostra para a Pesquisa Nacional de Saúde segundo os domínios de interesse (conclusão)

Domínio	Tamanho planejado da amostra de UPAs	Tamanho da amostra		
		Selecionado de UPAs	Esperado de Domicílios com entrevista realizada (23% de perda)	Selecionado de Domicílios
<b>Centro-oeste</b>	790	774	7.740	10.062
<b>Mato Grosso do Sul</b>	180	179	1.790	2.327
<b>Campo Grande</b>	90	90	900	1.170
<b>Restante</b>	90	89	890	1.157
<b>Mato Grosso</b>	180	162	1.620	2.106
<b>RM</b>	108	92	920	1.196
<b>Cuiabá</b>	72	72	720	936
<b>Resto RM</b>	20	20	200	260
<b>Restante</b>	70	70	700	910
<b>Goiás</b>	250	253	2.530	3.289
<b>RM</b>	110	111	1.110	1.443
<b>Goiânia</b>	90	90	900	1.170
<b>Resto RM</b>	20	21	210	273
<b>RIDE DF</b>	50	50	500	650
<b>Restante</b>	90	92	920	1.196
<b>Distrito Federal</b>	180	180	1.800	2.340