Universidade de São Paulo
Faculdade de Saúde Pública

**Inquéritos por Telefone: inferências válidas em
regiões com baixa taxa de cobertura de linhas
residenciais**

Regina Tomie Ivata Bernal

Tese apresentada ao Programa de Pós-Graduação em
Saúde Pública da Faculdade de Saúde Pública da
Universidade de São Paulo para obtenção do título de
Doutor em Ciências, Programa de Saúde Pública

Área de concentração: Epidemiologia

Orientadora: Profa. Dra. Nilza Nunes da Silva

São Paulo

2011

**Inquéritos por Telefone: inferências válidas em
regiões com baixa taxa de cobertura de linhas
residenciais**

Regina Tomie Ivata Bernal

Tese apresentada ao Programa de Pós-Graduação em
Saúde Pública da Faculdade de Saúde Pública da
Universidade de São Paulo para obtenção do título de
Doutor em Ciências, Programa de Saúde Pública

Área de concentração: Epidemiologia

Orientadora: Profa. Dra. Nilza Nunes da Silva

São Paulo

2011

Agradecimentos

Aos meus filhos, Volnys Kenzo e Nicolas, que alegram minha vida dando-me motivos e estímulos para progredir.

Ao meu esposo Volnys, pela compreensão e apoio.

Aos meus pais Kasuo Ivata (*in memoriam*) e Aiaco Karashima Ivata, pelo belo exemplo de dignidade e caráter, alicerces da minha formação.

A Prof^a Dr^a Nilza Nunes da Silva, pela orientação, formação acadêmica, amizade e confiança depositada.

Ao Prof. Dr. Pascoal Torres Muniz e Thiago Santos Araújo da Universidade Federal do Acre, pela disponibilidade do banco de dados para análise.

Ao Prof. Dr. Nelson Tanaka, pela disponibilidade em discutir métodos estatísticos.

A Sueli Daffre Carvalho, pela força, apoio e atenção recebida em todos os momentos.

A Prof^a Dr^a Deborah Carvalho Malta, membro da banca, pela colaboração, apoio e incentivo desde o início deste trabalho.

A Dr^a Maria Cecilia Goi Porto Alves, membro da banca, pela disponibilidade, apoio e a contribuição dada neste trabalho.

A todos que, de forma direta e indireta, contribuíram para a realização deste trabalho.

Resumo

Inquéritos por Telefone: inferências válidas em regiões com baixa taxa de cobertura de linhas residenciais

Introdução: O inquérito por telefone, quando comparado ao inquérito domiciliar possui vários atrativos, em especial baixo custo operacional e rapidez do processo de divulgação de resultados. No entanto, a exclusão de domicílios sem telefone fixo, pode representar séria questão de validade nas estimativas obtidas. **Objetivo:** Avaliar vícios potenciais nos resultados divulgados no Sistema de Vigilância de Fatores de Risco para Doenças Crônicas por Inquérito Telefônico (VIGITEL) em município de baixa cobertura de domicílios com telefone fixo. **Métodos:** A partir de resultados levantados pelo Inquérito Domiciliar realizado no município de Rio Branco-AC, com cobertura de 41% dos domicílios com telefone fixo, tentou-se localizar vícios introduzidos nos resultados do Vigitel. Foi usado método alternativo de ponderação para diminuir o vício da estimativa do Vigitel. **Resultados:** O Vigitel subestima a maioria das prevalências estimadas. Os pesos de pós-estratificação eliminam parcialmente o vício, cuja origem é proveniente de baixa taxa de cobertura de domicílios com telefone fixo. Por outro lado, o uso desses pesos, quando não necessário, potencializou o vício das variáveis não associadas à posse de telefone fixo. **Conclusões:** Em municípios de baixa taxa de cobertura de domicílios com telefone fixo, torna-se necessária a implementação de novo método de ponderação e estratégia de seleção de variáveis externas para construção dos pesos de pós estratificação, que minimizem o vício nas estimativas das variáveis levantadas.

Descritores: Inquérito domiciliar; Inquérito por telefone; Vigitel; Pesos de pós-estratificação; Vícios; Erro quadrático médio; Cobertura; Ausência de resposta;

Abstract

Telephone survey: valid inferences in regions with low coverage rate of residential lines

Introduction: The telephone survey, when compared to the household survey has many attractions, especially low operating costs and speed the process of dissemination of results. However, the exclusion of households without a landline, can pose serious question the validity at the estimates. **Objective:** Evaluating potential bias in the results published in the Vigilância de Fatores de Risco para Doenças Crônicas por Inquérito Telefônico (Vigitel) in the city of low coverage of households with landline. **Methods:** From the results collected by the Household Survey conducted in Rio Branco-AC, with coverage of 41% of households with landline, attempted to locate to bias in the results VIGITEL. Alternative method was used to reduce of the bias Vigitel. **Results:** Vigitel underestimates the majority of prevalence estimates. The poststratification to partially eliminate the bias, which originates from low coverage rate of households with a landline. Moreover, the use of these weights, if not necessary, increased the bias of the variables not associated with ownership of fixed telephone. **Conclusions:** In cities with low coverage rate of households with a landline, it is necessary to implement a new method of weighting and selection strategy of external variables for the construction of poststratification that minimize the bias in estimates of the variables studied.

Key words: Household survey, Telephone survey; Vigitel; Poststratification, Bias; Mean square error; Noncoverage; Nonresponse;

Índice

1.	Introdução	1
1.1	Procedimentos de ponderação e avaliação de vícios	12
1.2	Inquérito por telefone	22
2.	Objetivos	27
3.	Métodos.....	28
3.1	Bases de dados	28
3.1.1	Vigitel.....	28
3.1.2	Inquérito Domiciliar	30
3.2	Seleção das variáveis	33
3.3	Análises Estatísticas.....	35
3.3.1	Comparação das populações usuárias e não usuárias de linhas telefônicas residenciais	35
3.3.2	Perfil do não usuário de linhas telefônicas residenciais	35
3.3.3	Estimação de vício de cobertura.....	37
3.3.4	Redução de vício de cobertura	39
	Método de ponderação alternativo	44
	Seleção das variáveis utilizadas na construção dos pesos de pós estratificação	45
	Avaliação do método alternativo de ponderação.....	47
4.	Resultados.....	48
5.	Discussão.....	65
6.	Conclusão.....	69

7.	Referências Bibliográficas.....	70
8.	Anexo 1	77
8.1	Resultados dos modelos de regressão logística pelo método “ <i>forward</i> ” de seleção de variáveis	77
8.1.1	Consumo de feijão em cinco ou mais dias por semana	77
8.1.2	Consumo de feijão em cinco ou mais dias por semana	77
8.1.3	Consumo de frutas, verduras e hortaliças regularmente.....	78
8.1.4	Consumo de frutas, verduras e hortaliças recomendado	78
8.1.5	Consumo de carnes com gordura visível.....	79
8.1.6	Consumo de leite com teor integral de gordura	79
8.1.7	Consumo de refrigerante em cinco ou mais dias por semana	80
8.1.8	Prática de atividade física no tempo livre	81
8.1.9	Comportamento sedentário.....	82
8.1.10	Consumo de bebida alcoólica de forma abusiva	83
8.1.11	Fumante	84
8.1.12	Ex-fumante	85
8.1.13	Estado de saúde considerada ruim.....	86
8.1.14	Diagnóstico de hipertensão arterial	87
8.1.15	Diagnóstico de diabetes	87
8.1.16	Diagnóstico de infarto, derrame ou acidente vascular cerebral.....	88
8.1.17	Diagnóstico de colesterol ou triglicérides elevado.....	89
8.1.18	Diagnóstico de asma, bronquite asmática, bronquite crônica ou enfisema....	90
8.1.19	Diagnóstico de osteoporose	91
9.	Anexo 2	92
9.1	Saídas do programa “rake and trimm”	92
9.1.1	Pesos de pós estratificação segundo sexo e idade	92

9.1.2	Peso de pós estratificação segundo sexo, idade e escolaridade.....	98
-------	--	----

Lista de tabelas

Tabela 1 – Proporção de não respondentes e estimativa do vício igual a 0.5	14
Tabela 2 – Classificação do modelo de regressão logística.....	46
Tabela 3 – Diferença de prevalências das variáveis de interesse segundo posse de telefone fixo levantadas no Inquérito Domiciliar. Município de Rio Branco-AC, 2007	50
Tabela 4 – Distribuição bivariada da posse de telefone fixo segundo características demográficas do Inquérito Domiciliar. Município de Rio Branco-AC, 2007	51
Tabela 6 – Vício e erro quadrático médio (EQM) das prevalências das variáveis de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007	54
Tabela 7 – Diferença relativa entre as prevalências ponderadas e erros padrão segundo variável de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007.....	57
Tabela 8 – Variáveis explicativas por ordem de entrada e acerto médio do modelo de regressão logística, segundo variável de resposta levantadas no Vigitel. Município de Rio Branco-AC, 2007.....	60
Tabela 9 – Prevalências estimadas para diferentes pesos de pós-estratificação das variáveis de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007.....	62
Tabela 10 – Razão entre REQMR da prevalência obtida para diferentes pesos de pós estratificação pelo método “ <i>rake and trimm</i> ” e Vigitel. Município de Rio Branco-AC, 2007	64

Lista de figuras

Figura 1- Distribuição da média amostral	3
Figura 2- Classificação da origem de vícios de um inquérito	4
Figura 3 – Distribuição de frequência bivariada da população e da amostra. Método de ponderação por célula.....	17
Figura 4 – Tabela de contingência da população e da amostra	18
Figura 5 – Saída do algoritmo SAS macro Raking	19
Figura 6 – Pesos de pós-estratificação obtidos pelo método “ <i>raking</i> ”	20
Figura 7- Distribuição do prefixo das linhas telefônicas residenciais segundo réplicas	29
Figura 8 – Fluxo operacional de campo do Vigitel. Município de Rio Branco-AC, 2007.....	30
Figura 9 – Box plot da variável pesos amostrais (w1)	32
Figura 10- Fluxo operacional de campo do Inquérito Domiciliar. Município de Rio Branco-AC, 2007.....	33
Figura 11 – Ilustração do uso da probabilidade de ocorrência do sucesso para classificação do modelo de regressão logística.....	46
Figura 12 - Pirâmide etária. Município de Rio Branco-AC, 2007	55
Figura 13 - Anos de estudo na amostra e na população. Município de Rio Branco-AC, 2007	55
Figura 14 - Vício absoluto das prevalências não ponderadas e ponderadas pelo peso de pós-estratificação do Vigitel. Município de Rio Branco-AC, 2007	58

Lista de quadros

Quadro 1 – Métodos de ponderação para compensar ausência de resposta e cobertura de cadastro.....	15
Quadro 2 – Relação das variáveis no estudo utilizadas no estudo de comparação	34
Quadro 3 – Descrição das variáveis utilizadas no modelo de regressão logística múltipla, para identificar o perfil do não usuário de telefone fixo	36
Quadro 5 – Variáveis utilizadas no modelo de regressão logística múltipla.....	45

Lista de expressões

Expressão 1 – Modelo de regressão logística (PAULA, 2004).....	36
Expressão 2 – Vício de cobertura	37
Expressão 3 – Estimador da proporção ($\hat{P}$).....	37
Expressão 4 – Erro quadrático médio.....	38
Expressão 5 – REQM Relativo.....	38
Expressão 6 – Razão REQMR	47

1. Introdução

No Brasil, em 2006, a Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS) implantou o Sistema de Vigilância de Fatores de Risco para Doenças Crônicas (DCNT) por Inquérito Telefônico – VIGITEL, nas 26 capitais brasileiras e no Distrito Federal. Esse sistema coleta informações sobre fatores de risco, como o hábito de fumar, o consumo excessivo de alimentos com fonte de gordura saturada, o excesso de peso, o sedentarismo e o consumo excessivo de bebida. Como fatores de proteção, foram coletadas as informações sobre as práticas de atividade física no lazer, o consumo de hortaliças e verduras, e a prevenção de câncer. As DCNT ameaçam a qualidade de vida de milhões de pessoas, em especial os de baixa e média renda, e ainda geram grandes impactos econômicos, representando um dos desafios de saúde para o desenvolvimento global nas próximas décadas. Frente a esse cenário, a Organização Mundial de Saúde (OMS) propôs aos países membros compromissos para a redução das taxas de morbimortalidade por DCNT (BRASIL, 2007).

A população em estudo do VIGITEL é composta por adultos com 18 anos ou mais de idade, residente em domicílios com pelo menos uma linha telefônica residencial fixa. Esse sistema utiliza um questionário eletrônico para coletar informações acerca do padrão de alimentação e atividade física, atividade física realizada, características da composição corporal, frequência de consumo de cigarros e bebidas alcoólicas, e auto-avaliação do estado de saúde. Essa pesquisa é atualizada anualmente, o que permite acompanhar a evolução dos hábitos e dos comportamentos da população brasileira.

Os resultados apresentados por MONTEIRO et al. (2005) mostram que o custo da realização de entrevista por telefone é de quatro a oito vezes inferior ao custo de um

inquérito domiciliar realizado no município de São Paulo. A taxa de respostas alcançada foi de 84,7% e a duração das entrevistas variou de 6,5 a 13,3 minutos. A porcentagem de recusa foi de 11,7%, e 3,6% das linhas telefônicas não foram contatadas, devido a problemas como: telefone inexistente, fora de serviço, comercial, entre outros.

Assim, o VIGITEL tem como objetivo disponibilizar instrumentos eficazes para o monitoramento da frequência e da distribuição dos principais fatores determinantes das DCNT, de forma a subsidiar o delineamento e a avaliação de políticas públicas, visando à promoção da saúde da população adulta residente nas 26 capitais brasileiras e no Distrito Federal. Entretanto, a principal limitação do VIGITEL está associada à exclusão de indivíduos que residem em domicílios sem telefone fixo.

Embora a taxa de cobertura de residências atendidas por uma linha fixa tenha crescido a partir de 2000, os resultados da Pesquisa de Amostras por Domicílios (PNAD) do IBGE, realizada em 2007, apontam estimativas entre 33% (Palmas) e 78% (São Paulo) de domicílios localizados na área urbana das capitais servidos por pelo menos uma linha telefônica residencial. As regiões Norte e Nordeste apresentam as menores taxas, 40% e 48% respectivamente. Em Rio Branco 40% dos domicílios têm acesso a uma linha telefônica residencial. Na região Centro-Oeste, essa taxa encontra-se em 57%, enquanto as maiores taxas estão nas regiões Sudeste e Sul, com 76% e 71%, respectivamente.

Devido à exclusão do segmento da população sem telefone fixo na amostra, a inferência estatística é corrigida por métodos de ponderação que estendem os resultados observados da população com telefone fixo, para o segmento sem telefone fixo. Esses procedimentos visam à correção do vício potencial introduzido pela exclusão dos indivíduos moradores em domicílios sem telefone fixo.

O vício de uma estimativa é calculado a partir da diferença entre $[E(\bar{y}) - \bar{Y}]$ a esperança da distribuição de médias amostrais e o valor populacional, e quanto menor a sua magnitude, maior a validade da estimativa (KISH, 1965). A esperança da distribuição amostral é a medida central da distribuição das médias obtidas por meio de sorteio de todas as amostras possíveis com reposição. Nos procedimentos usuais de amostragem, espera-se que essa diferença seja igual a zero. No entanto, há situações onde isso não ocorre como ilustra a Figura 1. A origem do vício pode ser proveniente de diversas fontes amostrais e não amostrais ilustrados na figura 2 (KISH, 1965).

Figura 1- Distribuição da média amostral

Figura 2- Classificação da origem de vícios de um inquérito

Os vícios de amostragem são provenientes do cadastro usado para sortear a amostra e dos estimadores utilizados para obtenção dos resultados. Para cada modalidade de sorteio (sistemática, casual simples, estratificada ou por conglomerados) existe um procedimento estatístico adequado para obtenção de estimativas não viciadas (médias, proporções e variâncias). Esses procedimentos estão disponíveis em programas como SAS, STATA e SPSS. No entanto, ao desprezar a estrutura do plano de amostragem nas análises estatísticas, o pesquisador poderá introduzir vícios nos resultados. No processo de amostragem, o erro padrão indica a precisão, e mede o erro de natureza aleatória do procedimento de amostragem. Quanto menor o erro padrão, maior a precisão das estimativas obtidas na amostra (KISH, 1965).

Os vícios não amostrais ocorrem quando: parte da população de estudo é excluída da amostra, devido ao cadastro incompleto utilizado para fins de sorteio; recusa do indivíduo sorteado em participar da pesquisa. Taxas de cobertura ou respostas inferiores a 70% podem introduzir vícios nas variáveis de interesse, e valores acima de 70% os vícios são desprezíveis (BERNAL e SILVA, 2006; GROVES, 2006). SILVA (2001)

alerta que as atividades de coleta, digitação e processamento de dados devem ser realizadas conjunta e progressivamente, durante o trabalho de campo, para garantir a qualidade dos dados, uma vez que o trabalho de campo pode gerar diversas fontes de erros de observação, como o entrevistado esquecer uma informação e responder erroneamente ou distorcer a resposta, ou ainda, o próprio entrevistador responder o questionário, cometendo fraude. A qualidade dos dados de um inquérito é de extrema importância para acurácia, reprodutibilidade e obtenção de resultados válidos. A equipe que realiza a pesquisa deve implantar processo de supervisão para garantir a qualidade e para prevenir de práticas não aceitáveis e minimizar a chance de vícios não amostrais (ÜSTUN et al., 2005). Questões relacionadas à qualidade dos dados em inquéritos tem sido foco de atenção de estudiosos em amostragem nos congressos internacionais. Essa discussão se restringe a países desenvolvidos, com pouca participação de países em desenvolvimento. Essa é a principal conclusão, depois de avaliar publicações de várias conferências e iniciativas. Recentemente, em alguns países em desenvolvimento foram publicados artigos que abordam essa questão (SILVA, 2005; VASCONCELOS, 2009; FREITAS, 2009).

BERNAL e SILVA (2006) estudaram o efeito da taxa de cobertura de linhas telefônicas residenciais na introdução de vícios potenciais nas estimativas de prevalências de doenças crônicas no Brasil. O principal resultado foi a detecção de vícios relativos, considerados não desprezíveis, nas estimativas de prevalências de doenças crônicas, para indivíduos com 40 anos e mais. Nas regiões Norte, não metropolitanas, do Centro-Oeste, do Nordeste e do Sudeste, com taxas de cobertura de LTR abaixo de 60%, apresentaram vícios relativos acima de 1,5, o que reduz o nível de

confiança pré-fixado (95%) para 68%, enquanto na região metropolitana do Nordeste registrou-se alteração para aproximadamente 90%. Nas regiões metropolitanas do Sudeste, Sul e do Centro-Oeste, com taxas de cobertura entre 70 e 90%, os vícios associados podem ser considerados desprezíveis. Vale ressaltar que a distribuição da cobertura de domicílios com linhas telefônicas residenciais (LTR) não ocorreu de forma uniforme no Brasil. Em 2003, a taxa média de domicílios brasileiros atendidos por LTR era de 50,3%. E, ao se estratificar essa taxa por região, registrou-se uma variação de 28% a 62%, sendo a região Nordeste a de menor taxa de cobertura (28%), seguida da região Norte urbana (38%), Centro Oeste (51%), Sul (59%) e Sudeste (62%). Além disso, a proporção de adultos residentes em domicílios sem linha telefônica é maior nas áreas rurais e nas regiões não metropolitanas. Também os adultos declarados como não brancos apresentam maior proporção de “sem telefone residencial” do que os brancos; e a proporção de adultos sem telefone é inversa aos anos de escolaridade. Aplicando os pesos de pós-estratificação espera-se reduzir os vícios encontrados, devido à exclusão da população sem acesso à linha telefônica. Na região Norte Urbana, a escolaridade, cor da pele e posse de plano de saúde do chefe do domicílio são as características associadas à posse de linha telefônica, e na região Nordeste as características são: escolaridade, cor da pele, área urbana e posse de plano de saúde do chefe do domicílio. Nas regiões Sudeste, Sul e Centro-Oeste, as características são: escolaridade, cor da pele, posse de plano de saúde do chefe do domicílio, região metropolitana e área urbana.

O seminário sobre Inquéritos Popacionais promovido pela Comissão de Epidemiologia da ABRASCO (WALDMAN et al., 2008), em parceria com a Faculdade de Saúde Pública da USP e o Departamento de Medicina Social da Faculdade de

Ciências Médicas da Santa Casa, no período de 17 a 19 de abril de 2007, abriu uma discussão importante para o meio acadêmico sobre questões metodológicas, operacionais e éticas dos inquéritos atualmente utilizados para obtenção de informações sobre a saúde da população brasileira. No total, foram apresentados 39 “pôsteres” de pesquisadores das regiões Nordeste, Sudeste e Centro-Oeste, predominando SP e RJ. Dos 39 trabalhos apresentados, 10 estudos discutem questões relacionadas ao desenho das amostras, os cadastros utilizados no sorteio e a cobertura desses cadastros. O Sistema de Vigilância de doenças crônicas por entrevista telefônica – VIGITEL — implantado em 2006 pelo Ministério da Saúde, foi amplamente discutido. Em especial, foram apresentadas duas questões relevantes sobre o uso do inquérito por telefone:

“Interroga-se até que ponto os excluídos (sem linha telefônica) diferem dos incluídos em aspectos relevantes para a investigação. Com a expansão da metodologia do inquérito telefônico para as 26 capitais existe o risco de se construir um sistema nacional cujas informações não sejam comparáveis, pois em cada uma delas, pode-se estar trabalhando com estratos diferentes da população. A existência de mais de uma linha telefônica no domicílio poderia aumentar a chance de o mesmo ser sorteado, introduzindo outro fator de desequilíbrio na amostra. Por outro lado, em áreas de menor poder aquisitivo existem muitos telefones móveis (celulares) e poucos telefones fixos” (WALDMAN et al., 2011).

“Estas questões, para serem respondidas com bases em evidências empíricas e não apenas em considerações teóricas, necessitam de outras pesquisas que comparem várias características sociodemográficas da população que possui e não linhas de telefonia fixa e avalie as

características de subgrupos populacionais como os moradores de favelas com ou sem acesso a telefones fixos ou celulares” (WALDMAN et al., 2011).

Na 56th Session of the International Statistical Institute ocorrida em Lisboa, em agosto de 2007, foram apresentados quatro trabalhos (EUA, Suécia, Reino Unido e Brasil) na linha de pesquisa “Telephone Survey”. O trabalho apresentado por LESSER e NEWTON (2007) discutem a questão do aumento do número de não respostas e de não contatos ocorridos em vários inquéritos realizados pelo governo, no período de 1980 a 1997. Além disso, alertam sobre as mudanças ocorridas no setor de telefonia e no perfil do usuário, pois, nos seis primeiros meses de 2006, um em cada oito domicílios americanos possuíam telefone fixo. Os autores esperam um aumento do número de domicílios servidos somente por telefonia móvel, em geral, por usuários serem mais jovens em relação aos de telefonia fixa. Como consequência, os inquéritos a serem realizados por telefonia fixa apresentarão resultados com estimativas viciadas devido à baixa cobertura de telefones fixos. Entretanto, o aumento da cobertura de domicílios com acesso à Internet (75% em 2004) passa a ser atrativo para realização de inquéritos. Nesse contexto, o estudo comparou três resultados de inquéritos realizados por meio de Internet, Mala Direta e Telefone fixo. Apesar dos resultados das taxas de respostas não mostrarem diferenças entre os inquéritos, o trabalho mostra a preocupação do pesquisador em relação às mudanças de comportamentos da população. Enquanto o trabalho apresentado por THOMAS (2007) descreve como o “UK Office for National Statistics”, utiliza, desde 1955, o telefone como meio de coleta de dados de empresas. Também descreve os benefícios alcançados como: qualidade, rapidez e baixo custo.

ISAKSSON et al. (2007) discutem a questão do número de chamadas a serem realizadas no inquérito, levando em consideração desenho da amostra, taxa de resposta, erro padrão e o custo de coleta de dados. Os autores apresentam um modelo de custo e um estimador para o número ótimo de chamadas. Dessa forma, o número ótimo de chamadas é o produto da variância da variável de interesse e dos custos relacionados para obtenção dos dados, como custo da ligação, infraestrutura do local e do número de ligações planejadas inicialmente.

O trabalho apresentado por SILVA (2007) discute a questão do efeito da taxa de cobertura de domicílios com telefone fixo nas estimativas de doenças crônicas e descreve o perfil do usuário de telefone fixo. Os resultados apresentados mostram que o perfil do usuário está relacionado com a condição sócio-demográfica e a localização geográfica do domicílio. Além disso, as regiões com cobertura acima de 70% dos domicílios com telefone fixo apresentaram estimativas não viciadas.

KERCKHOVE et al. (2008) avaliam a estimativa do vício introduzido pelas baixas taxas de resposta e de cobertura de domicílios com telefone fixo ocorridas no inquérito “*National Household Education Surveys Program*” (NHES), em 2007, denominada NHES 2007. Esse inquérito utiliza o programa desenvolvido pelo “*National Center for Education Statistics*” (NCES) do “*Institute of Education Sciences*” pertencente ao “*U.S. Departamento of Education*”, que sorteia as linhas telefônicas residenciais por meio do método “*random digital dialing*” (RDD). O vício das estimativas foi avaliado por meio de comparações de resultados obtidos nos inquéritos por telefone e domiciliar. A amostra de domicílios, denominada “*NHES 2007 Bias Study*”, é composta por 7.500 endereços sorteados por meio de amostragem por conglomerados em três estágios de seleção, independentemente do inquérito NHES 2007. Essa amostra foi cruzada com

dois cadastros de endereços com telefones fixos para obter o número do telefone dos endereços sorteados. No total, foram encontrados os telefones de 5.433 (72%) domicílios. Destes, 2.396 foram contatados por telefone e 3.037 telefones que não atenderam foram entrevistados pessoalmente. No final, foram realizadas 4.894 entrevistas completas, sendo 2.550 pessoalmente e as demais por telefone, para encontrar a população de estudo composta por crianças em idade escolar (“School Readiness” (SR)), pais e familiares responsáveis na educação dessas crianças (“Parente and Family Involvement in Education” (PIF)), e Adult Education for Work-Related” (AEWR). No estudo de comparação das estimativas dos inquéritos por telefone e domiciliar, foram controlados os erros amostrais e não amostrais, de forma a minimizar o impacto desses componentes; no inquérito domiciliar, as estimativas do erro padrão são não viciadas, pois foram obtidas por meio de técnicas de simulação, sendo usado o estimador de variância “*Jackknife*”; os resultados do “NHES 2007 *Bias Study*” foram comparados com dados externos (“*Current Population Survey*” (CPS)), para garantir que fossem não viciados; para detectar diferenças entre os resultados, foi utilizado um teste *t-student* adequado para amostras correlatas com nível de significância de 5%. Em geral, as comparações das estimativas obtidas no “NHES 2007 *Bias Study*” e CPS indicam que os resultados do inquérito domiciliar podem ser usados como padrão de referência para detectar a presença de vício no inquérito por telefone. Dessa maneira, o vício é calculado pela diferença entre a estimativa obtida no inquérito NHES 2007 e NHES 2007 *Bias Study*, e o vício relativo é a razão entre o vício e a estimativa do NHES 2007 *Bias Study* multiplicado por 100. Os resultados mostram que a maioria das estimativas avaliadas apresentam vícios desprezíveis. No entanto, devido à tendência de queda no número de domicílios com linha telefônica residencial e nas taxas de

respostas, no futuro, certamente, haverá problemas de vício introduzidos pela baixas taxas de cobertura e de respostas.

BATTAGLIA et al. (2008) avaliam o vício de não resposta introduzido nas estimativas das variáveis que compõem os fatores de risco do “*Behavioral Risk Factor Surveillance System*” (BRFSS) do “*Center of Disease Control and Prevention*” (CDC, 2009). Esse sistema utiliza o inquérito por telefone como meio de coleta de informações, relacionadas aos fatores de risco, da população adulta residente em domicílios com telefone fixo, em todos os estados americanos. Apenas 15,8% dos domicílios não possuem telefone fixo. Entretanto, a taxa de resposta está abaixo de 50%. Para suavizar os vícios nas estimativas, introduzidas pela baixa taxa de resposta, o BRFSS utiliza pesos de pós-estratificação para ajustar a distribuição das variáveis idade, sexo e cor da pele, obtidas na amostra para a população, baseadas em fontes externas. Para melhorar o desempenho do método de ponderação do sistema, na redução do vício, foi avaliado novo procedimento em seis estados americanos. A proposta é baseada na seleção de variáveis demográficas associadas às variáveis de interesse, na construção dos pesos, e pelo método de ponderação “*rake*”, para obtenção dos pesos de pós-estratificação. Os resultados da avaliação do vício, expressa pelo erro quadrático médio do estimador, mostram redução significativa dos vícios nas estimativas das prevalências de fatores de risco. Devido à melhora na redução do vício, a proposta foi avaliada para os demais estados americanos.

Como pode ser visto, a ausência de respostas e baixa taxa de cobertura podem introduzir vícios não desprezíveis nas inferências produzidas.

1.1 Procedimentos de ponderação e avaliação de vícios

A ausência de resposta e a taxa de cobertura do cadastro utilizado para fins de sorteio são as principais preocupações de pesquisadores que utilizam técnicas de amostragem probabilísticas para inferência de resultados da amostra para a população. Dependendo do percentual de perda de informação da amostra, as estimativas dos parâmetros populacionais podem apresentar magnitudes de vícios não desprezíveis.

KISH (1965) define duas fontes de ausência de informação: cadastro incompleto e não resposta. O indivíduo que não consta no cadastro apresenta probabilidade zero de pertencer à amostra sorteada, enquanto a não resposta é resultado da operação de campo, cujo indivíduo sorteado não foi encontrado ou recusa. O efeito da ausência de informação é denotado por $\bar{Y} = \frac{Y}{N} = w_1\bar{Y}_1 + w_2\bar{Y}_2$, sendo que w_1 e w_2 representam a proporção de indivíduos que responderam e não responderam; $\bar{Y}_1$ a média observada no grupo que respondeu; $\bar{Y}_2$ a média observada no grupo que não respondeu; $\bar{Y}$ a média da população. Na exclusão de parte da população, o valor ($\bar{Y}_1$) observado no grupo que respondeu representa o valor da população ($\bar{Y}$). No entanto, se a diferença ($\bar{Y}_1 - \bar{Y}$) for diferente de zero, isso indica presença de vício. O vício relativo ($VR(\bar{Y}_1) = \frac{(\bar{Y}_1 - \bar{Y})}{\bar{Y}} = \frac{(\bar{Y}_1 - w_1\bar{Y}_1 - w_2\bar{Y}_2)}{\bar{Y}} = w_2 \frac{(\bar{Y}_1 - \bar{Y}_2)}{\bar{Y}}$) poderá ser desprezível se a diferença entre o resultado de quem respondeu com quem não respondeu for pequena.

KALTON (1983) define três tipos de ausência de informação: exclusão de parte da população de estudo, indivíduo não encontrado ou recusa e questões sem respostas. O

primeiro tipo está relacionado ao cadastro utilizado para fins de sorteio, no qual o indivíduo que não está no cadastro tem probabilidade zero de pertencer à amostra sorteada. O segundo está associado à recusa da população elegível em participar da pesquisa e, por último, o entrevistado pode se recusar a responder a algumas questões do questionário. O procedimento usual para compensar a ausência de informação, devido à exclusão de parte da população elegível, é ajustar a distribuição da amostra para a da população, mediante seleção de variáveis disponíveis em fontes externas. Esse procedimento é conhecido por pesos de pós-estratificação. A ausência do indivíduo ou questões sem respostas são compensadas por procedimentos de imputação de valores ou construções de pesos baseados nas características do grupo que respondeu e do grupo que não respondeu. Ao fazer isso, espera-se reduzir o vício ($B(\bar{Y}_r) = E(\bar{y}_r) - \bar{Y}$). O efeito da estimativa do vício de não resposta ($B(\bar{y}_r) = \frac{M}{R+M}(\bar{Y}_r - \bar{Y}_m)$), apresentada por KALTON (1983), parte de uma amostra aleatória simples de tamanho n de uma população N , cujo objetivo é conhecer o parâmetro ($\bar{Y}$) da variável y . Com o pressuposto da existência do grupo de indivíduos que sempre respondem (R) e aqueles que nunca respondem (M), a amostra foi dividida em dois estratos: o que responde é denotado por r e o que não responde por m . Quanto menor a proporção de não respondentes, mais próxima estará a estimativa do parâmetro dos respondentes ($\bar{Y}_r$) e não respondentes ($\bar{Y}_m$), como mostra a tabela 1. O parâmetro populacional é desconhecido, assim como a magnitude do vício. Por isso, é importante controlar a taxa de não respondentes, de forma a minimizar o vício.

Tabela 1 – Proporção de não respondentes e estimativa do vício igual a 0.5

$\frac{m}{r+m}$	$B(\bar{y}_r)$
0.10	0.050
0.20	0.100
0.30	0.150
0.40	0.200
0.50	0.250
0.60	0.300
0.70	0.350
0.80	0.400
0.90	0.450
1.00	0.500

KALTON (1983) recomenda o controle da taxa de resposta, pois o vício será pequeno se a proporção de não respondentes for pequena. Em situações onde há ausência de informação, o uso de procedimentos para compensar ausência de resposta é comumente utilizado para diminuir o vício (KALTON, 1983; BRICK et al., 1996; KALTON e FLORES-CERVANTES, 2003; GROVES, 2006; BATTAGLIA et al., 2008). Esses procedimentos pressupõem que os resultados dos respondentes sejam iguais aos que não responderam, ou a ausência de informação seja aleatória. Os procedimentos são divididos em dois grupos: métodos de ponderação e técnicas de imputação (KALTON, 1983).

KALTON e FLORES-CERVANTES (2003) apresentam seis métodos de ponderação para compensar ausência de resposta e baixa cobertura de cadastro (quadro 1). Esses métodos utilizam variáveis disponíveis na amostra e fontes externas, para ajustar a distribuição conjunta ou marginal das variáveis da amostra para a da população. A escolha de variáveis correlacionadas com ausência de resposta ou baixa

cobertura de cadastro é fundamental para obter melhor eficiência na redução do vício (quadro 1). O fator $F=1+CV(w_i)^2$, onde $CV(w_i)$ é o coeficiente de variação dos pesos (w_i), indica perda de precisão da estimativa ponderada. Esse fator é usado como critério na escolha do método.

Os procedimentos de ponderação podem gerar pesos com magnitudes diferentes. Na presença de pesos com valores extremos, o fator F é inflacionado pelo coeficiente de variação. Técnicas que aparam pesos extremos ou que juntam células são utilizadas para diminuir a variabilidade dos pesos. Métodos alternativos para tratar pesos com valores extremos foram propostos por POTTER (1988) e BATTAGLIA et al. (2004).

Quadro 1 – Métodos de ponderação para compensar ausência de resposta e cobertura de cadastro

Procedimento de ponderação	Característica	Vantagens	Desvantagens	Recomendação de uso
Célula	Ajusta a distribuição conjunta das variáveis da amostra para a da população	Fácil de calcular os pesos	Pode apresentar alta variabilidade na distribuição dos pesos. Isso implica aumento na variância da estimativa do parâmetro.	Ausência de resposta e baixa taxa de cobertura de cadastro. Quando o número de células é relativamente pequeno e a quantidade de observações (n) em cada célula é grande.
Raking	Ajusta a distribuição marginal da variável na amostra para a da população	Utiliza a distribuição marginal de cada variável auxiliar, obtida em fonte externa		Ausência de resposta e baixa taxa de cobertura de cadastro. Número grande de células.

Procedimento de ponderação	Característica	Vantagens	Desvantagens	Recomendação de uso
Linear	Ajusta a distribuição marginal da variável na amostra para a da população	Similar ao “raking”	Pode gerar pesos negativos	Ausência de resposta.
GREG	Ajusta a distribuição conjunta ou marginal da variável na amostra para a da população	Pesos positivos; uso de covariáveis contínuas na construção dos pesos		Ausência de resposta.
Regressão logística	Calcula a probabilidade de o indivíduo responder	Usar covariáveis qualitativas e quantitativas		Ausência de resposta.
Mistura de célula e outros métodos		Não há suposição em relação à origem da ausência de informação	Pode produzir pesos instáveis em situações onde célula apresenta número pequeno de observações	Ausência de resposta.

A seguir são apresentados os dois métodos de ponderação utilizados neste estudo.

O Vigitel adota o método de ponderação por célula, para suavizar o vício introduzido pela baixa taxa de cobertura de domicílios com telefone fixo. A figura 3 apresenta a distribuição conjunta (w_{HK}) e as marginais ($w_{H.}$ e $w_{.K}$) da população e da

amostra. Nesse método, os pesos são obtidos por meio da razão entre frequência relativa (ou absoluta) da população e da amostra, em cada célula (w_{HK}/q_{HK}).

Figura 3 – Distribuição de frequência bivariada da população e da amostra. Método de ponderação por célula

	População				Total		Amostra				Total
	1	2	...	K			1	2	...	K	
1	FR ₁₁	FR ₁₂	...	FR _{1K}	FR _{1.}	1	fr ₁₁	fr ₁₂	...	fr _{1K}	fr _{1.}
2	FR ₂₁	FR ₂₂	...	FR _{2K}	FR _{2.}	2	fr ₂₁	fr ₂₂	...	fr _{2K}	fr _{2.}
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
H	FR _{H1}	FR _{H2}	...	FR _{HK}	FR _{H.}	H	fr _{H1}	fr _{H2}	...	fr _{HK}	fr _{H.}
Total	FR _{.1}	FR _{.2}	...	FR _{.K}	1	Total	fr _{.1}	fr _{.2}	...	fr _{.K}	1

Figura extraída do livro de KALTON, 1983 pp.56

Pesos de pós-estratificação obtidos pelo método “raking” são amplamente utilizados para compensar ausência de resposta e baixa taxa de cobertura de domicílios com telefone fixo (BRICK et al., 1996; FLORES-CERVANTES et al., 2002; BATTAGLIA et al., 2004; BATTAGLIA et al., 2008; KERCKHOVE et al., 2008).

O método “raking” utiliza as distribuições marginais das variáveis qualitativas disponíveis na população e na amostra (figura 4). O objetivo do método, diferentemente do procedimento por célula, é calcular os pesos de pós-estratificação (w_{HK}) por meio de processos iterativos, de forma que, ao final do processo, as distribuições marginais das variáveis da amostra sejam iguais às da população.

Figura 4 – Tabela de contingência da população e da amostra

	População				Marginal		Amostra				Marginal
	1	2	...	K			1	2	...	K	
1					FR _{1.}	1	w ₁₁	w ₁₂	...	w _{1K}	fr _{1.}
2					FR _{2.}	2	w ₂₁	w ₂₂	...	w _{2K}	fr _{2.}
.					.	.	.	.	.	.	.
.					.	.	.	.	.	.	.
.					.	.	.	.	.	.	.
H					FR _{H.}	H	w _{H1}	w _{H2}	...	w _{HK}	fr _{H.}
Marginal	FR _{.1}	FR _{.2}		FR _{.K}	1	Marginal	fr _{.1}	fr _{.2}		fr _{.K}	1

A seguir é apresentada uma aplicação do método de ponderação “*raking*” utilizando o algoritmo implementado por IZRAEL et al. (2000) no programa SAS, denominada SAS Macro Raking. Usando os dados da amostra do Vigitel, realizada no município de Rio Branco, para obtenção dos pesos de pós-estratificação, segundo as variáveis faixas de escolaridade e etária, o algoritmo precisou de sete iterações para que as distribuições marginais das variáveis da amostra fossem iguais às da população. A figura 5 apresenta a saída da SAS Macro Raking, para cada iteração. Observam-se sete colunas de resultados para faixa de escolaridade e para faixa etária. A primeira coluna informa as categorias da variável, na segunda e na terceira constam as distribuições da variável na amostra e na população, na quarta coluna encontra-se a diferença entre a distribuição da população e da amostra, na quinta coluna encontra-se a nova distribuição da variável na amostra após o primeiro cálculo dos pesos, na sexta coluna encontra-se a distribuição da variável na população e na última coluna é a diferença entre as colunas cinco e quatro. Ao final do processo são obtidos os pesos de pós-estratificação (figura 6).

Figura 5 – Saída do algoritmo SAS macro Raking

Raking by Faixa de Escolaridade, iteration - 1						
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Faixa Escolar.	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[0 ; 8]	1566.17	45.7	-1520.47	30.207	45.700	-15.49
[9 ;11]	2157.17	35.2	-2121.97	41.605	35.200	6.405
[12;20]	1461.50	19.1	-1442.40	28.188	19.100	9.088
	=====	=====		=====	=====	
	5184.83	100.0		100.00	100.00	
Raking by Faixa Etária, iteration - 1						
Faixa Etária	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[18; 25]	18.960	27	8.04037	18.960	27.000	-8.040
[26; 35]	22.677	26	3.32324	22.677	26.000	-3.323
[36; 45]	19.555	20	0.44497	19.555	20.000	-0.445
[46; 55]	17.070	14	-3.07040	17.070	14.000	3.070
[56;100]	21.738	13	-8.73818	21.738	13.000	8.738
	=====	=====		=====	=====	
	100.000	100		100.00	100.00	
Raking by Faixa de Escolaridade, iteration - 2						
Faixa Escolar.	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[0 ; 8]	39.447	45.7	6.25317	39.447	45.700	-6.253
[9 ;11]	39.490	35.2	-4.28973	39.490	35.200	4.290
[12;20]	21.063	19.1	-1.96345	21.063	19.100	1.963
	=====	=====		=====	=====	
	100.000	100.0		100.00	100.00	
Raking by Faixa Etária, iteration - 2						
Faixa Etária	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[18; 25]	25.430	27	1.57026	25.430	27.000	-1.570
[26; 35]	25.350	26	0.65047	25.350	26.000	-0.650
[36; 45]	20.190	20	-0.18990	20.190	20.000	0.190
[46; 55]	14.599	14	-0.59912	14.599	14.000	0.599
[56;100]	14.432	13	-1.43172	14.432	13.000	1.432
	=====	=====		=====	=====	
	100.000	100		100.00	100.00	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Raking by Faixa de Escolaridade, iteration - 7						
Faixa Escolar.	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[0 ; 8]	45.699	45.7	0.000979692	45.699	45.700	-0.001
[9 ;11]	35.201	35.2	-0.000673107	35.201	35.200	0.001
[12;20]	19.100	19.1	-0.000306585	19.100	19.100	0.000
	=====	=====		=====	=====	
	100.000	100.0		100.00	100.00	
Raking by Faixa Etária, iteration - 7						
Faixa Etária	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
[18; 25]	27.000	27	0.000251243	27.000	27.000	-0.000
[26; 35]	26.000	26	0.000091602	26.000	26.000	-0.000
[36; 45]	20.000	20	-0.000040227	20.000	20.000	0.000
[46; 55]	14.000	14	-0.000097075	14.000	14.000	0.000
[56;100]	13.000	13	-0.000205542	13.000	13.000	0.000
	=====	=====		=====	=====	
	100.000	100		100.00	100.00	

Figura 6 – Pesos de pós-estratificação obtidos pelo método “*raking*”

Faixa etária	Faixa de escolaridade		
	[0 ; 8]	[9; 11]	[12;20]
[18; 25]	0.147	0.069	0.053
[26; 35]	0.107	0.046	0.033
[36; 45]	0.089	0.033	0.026
[46; 55]	0.074	0.027	0.020
[55;100]	0.044	0.021	0.014

Os pesos de pós-estratificação gerados pelos métodos de célula e de “*raking*” podem apresentar alta variabilidade, devido à presença de valores extremos. Nessa situação, recomenda-se o uso de procedimentos propostos por POTTER (1988) para reduzir a variabilidade dos pesos. IZRAEL et al. (2009) disponibilizaram a SAS macro para o cálculo dos pesos de pós- estratificação, com critério de poda de pesos com valores extremos. O algoritmo utiliza o método “*Individual and Global Cap Value*” (IGCV) para aparar os pesos.

Com o uso adequado do método de ponderação, aliado à escolha correta das variáveis utilizadas na construção dos pesos de pós estratificação, espera-se que o vício potencial do estimador ($\hat{\theta}$) seja reduzido. Mensurar o efeito dos pesos de pós-estratificação na redução do vício do estimador ($\hat{\theta}$) é objeto de estudo de pesquisadores. O limite desse estudo está no desconhecimento do valor populacional (θ), isto é, o valor de referência, no qual o pesquisador irá mensurar o efeito dos ajustes de pesos na estimativa do parâmetro ($\hat{\theta}$). Devido à ausência do valor verdadeiro , o pesquisador elege, a partir de pressupostos, o valor de referência que será usado na

mensuração do vício. O Erro Quadrático Médio do estimador ($\hat{\theta}$) (LITTLE, 1991; BRICK et al., 1996; KALSBECK, 2001; BATTAGLIA et al., 2008), a Razão de Vícios (BRICK et al., 2003; FLORES-CERVANTES et al., 2002) e o Teste t para diferença de médias (KERCKHOVE et al., 2008) são critérios utilizados para medir o efeito dos ajustes de pesos.

O erro quadrático médio (EQM) é amplamente utilizado como critério de validade do estimador $\hat{\theta}$ do parâmetro θ , e ainda, na comparação de estimadores. A expressão matemática que define o EQM de um estimador $\hat{\theta}$ do parâmetro θ é $EQM[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + B^2(\hat{\theta})$, sendo $B(\hat{\theta}) = E[\hat{\theta}] - \theta$ o vício do estimador $\hat{\theta}$. Se a $E[\hat{\theta}] = \theta$, então o estimador $\hat{\theta}$ é dito não viciado para θ . Na ausência de vício, o EQM é igual à variância; caso contrário, o EQM expressa a acurácia do estimador (KISH, 1965; COCHRAN, 1977).

1.2 Inquérito por telefone

Em 1987, ocorreu a 1ª Conferência Internacional sobre Metodologia em Inquéritos por Telefone em Charlotte, NC. Como resultado dessa conferência, foi publicado o livro *“Telephone Survey Metodology”* (GROVES et al., 1988), que reúne 32 trabalhos apresentados na conferência. Essa obra é uma referência bibliográfica para muitos pesquisadores e estudantes que atuam em inquéritos por telefone. Questões relacionadas ao cadastro de linhas telefônicas residenciais, desenho do plano de amostragem e implantação de inquérito, vícios de cobertura de linhas telefônicas, não respostas e erro de amostragem são objetos de estudos nele apresentados, além do uso de um sistema de entrevista automatizado, conhecido por *“Computer-assisted telephone interviewing systems”*.

MASSEY (1988) apresenta as definições de alguns termos utilizados na literatura: *taxa de cobertura*, que representa o percentual de domicílios atendidos por uma linha telefônica fixa em relação ao total de domicílios com e sem telefone; *vícios potenciais* devidos às baixas taxas de cobertura, que são valores decorrentes da diferença entre as estimativas de prevalências da população residentes em domicílios com telefone fixo e da população total.

Segundo TREWIN e LEE (1988), houve uma demora, por parte de pesquisadores e Governo, na utilização do inquérito por telefone, devido à incerteza nos resultados provenientes de entrevistas realizadas em regiões de baixa taxa de cobertura. No entanto, a elevação dos custos operacionais na realização de entrevistas domiciliares, no mundo todo, motivou pesquisadores a avaliarem o impacto desses vícios em regiões de

baixa taxa de cobertura. Os autores analisaram as taxas de cobertura de 22 países (Suécia, Canadá, Finlândia, Nova Zelândia, Estados Unidos, Alemanha (Occidental), Dinamarca, França, Holanda, Austrália, Inglaterra, Itália, Israel, Noruega, Grécia, Irlanda, Espanha, México e Polônia), sendo a variação da taxa de 16% na Polônia, México e Hungria e 99% na Suécia. As características sócio-demográficas e a localização geográfica das populações residentes em domicílios sem linha telefônica são bem similares entre os países estudados. Em 1986, a “*United Nations*” divulgou, para 64 países, a estatística do número de telefones a cada 100 habitantes. Essas taxas de cobertura foram agrupadas em seis classes: I (até 90%), II (80% a 90%), III (70% a 79%), IV (50% a 69%), V (25% a 49%) e VI (10% a 24%), formando um guia rústico (inicial) quanto ao uso do inquérito por telefone. Dessa forma, nos países classificados nas classes I e II recomenda-se o inquérito por telefone, enquanto nos países da terceira e quarta classe recomenda-se o uso de métodos alternativos, como uso de cadastros híbridos para a realização de inquérito por telefone e domiciliar.

Embora o livro “*Telephone Survey Methodology*” (GROVES et al., 1988) seja referência para muitos pesquisadores, a rápida mudança nas telecomunicações ocorrida desde então fez com que surgisse um volume muito grande de estudos voltados, principalmente, à questão de desenho de amostra, de forma que atendessem à nova realidade. Na década de 1990, a expansão do número de linhas telefônicas disponíveis e os códigos de áreas representaram um aumento em mais de 50% dos bancos de prefixos válidos no método “*random digital dialing*” (RDD). Entretanto, o crescimento no número de domicílios com linhas telefônicas foi de somente 10%. Uma das consequências foi o aumento na ineficiência do método RDD. Isso fez com que novas metodologias de sorteio de amostras de linhas telefônicas fossem necessárias. Além

disso, o surgimento das novas tecnologias como secretária eletrônica, identificador de chamada e telefone móvel contribuiu para a diminuição nas taxas de resposta nos inquéritos realizados por telefone fixo nos EUA e na Europa.

Estudos importantes ocorrem desde a 1ª Conferência, incluindo o desenho da amostra, em respostas às mudanças no sistema de telecomunicações.

Em 2006, ocorreu a 2ª Conferência Internacional sobre Metodologia em Inquéritos por Telefone em Miami, Flórida, que resultou na publicação do livro “*Advances in Telephone Survey Methodology*” (LEPKOWSKI et al., 2008). Essa conferência teve como principal meta reunir pesquisadores e profissionais interessados em metodologia de inquérito por telefone, estimular pesquisas com contribuições nas áreas voltadas a medição de erros ou a redução de erros atribuídos ao desenho da amostra, documentação dos inquéritos e estimular novas idéias e desenvolvimentos voltados à metodologia. De acordo com os estudos apresentados, há um interesse dos pesquisadores no desenvolvimento de alternativas metodológicas para a obtenção de amostras, utilizando, de forma híbrida, os cadastros de telefonia móvel, fixa, e-mail e domiciliar para aumentar a taxa de cobertura, assim como melhorar a taxa de resposta. Além disso, há preocupação na obtenção de estimativas válidas, redução de custos e qualidade nas respostas.

Segundo TUCKER e LEPKOWSKI (2008), após a primeira conferência, surgiram várias metodologias para obtenção de amostras para inquéritos por telefone. Com o avanço da tecnologia, surgiram outros meios de comunicação, como a Internet e o telefone móvel, que podem ser utilizados com o telefone fixo para realização de entrevistas. O método “*Dual Frame*” e “*Mixed-Mode Design*” (GROVES e LEPKOWSKI, 1985; SIRKEN e CASADY, 1988) utilizam os cadastros de telefones e

domicílios para compor a amostra. No entanto, há questões a serem resolvidas, como o efeito nas estimativas das variáveis de interesse, devido aos diferentes meios de comunicação e o uso da telefonia móvel, que implica mudança do elemento da amostra, de domicílio para indivíduo.

O inquérito por telefone, como visto, possui vários atrativos, como baixo custo operacional e rapidez do processo, quando comparado ao inquérito domiciliar. No entanto, a produção científica existente expõe três principais questões: validade das estimativas obtidas em inquérito por telefone, devido à exclusão de domicílios sem telefone fixo, ausência de resposta e procedimentos metodológicos para redução da variabilidade dos pesos utilizados na obtenção de estimativas válidas. Além disso, existem outras questões relevantes, como a adaptabilidade do processo à evolução tecnológica, em decorrência do surgimento de outros meios de comunicação.

No Brasil, o Sistema de Vigilância de doenças crônicas por entrevista telefônica – VIGITEL — completou cinco anos, em 2011. Esse sistema está consolidado como ferramenta de coleta de dados, nas 26 capitais brasileiras e no Distrito Federal.

Estudos que examinam a validade dos resultados, vícios introduzidos pela baixa taxa de cobertura de domicílios e tendência das prevalências divulgadas pelo Vigitel são objetos de estudos promovidos pela Coordenação Geral de Doenças e Agravos Não Transmissíveis (CGDANT) da Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS).

A motivação do estudo é verificar se inquérito por telefone realizado em municípios de baixa taxa de cobertura de domicílios com telefones fixos geram estimativas viciadas para as variáveis associadas à posse de telefone fixo e se os pesos de pós-estratificação podem reduzir os vícios.

2. Objetivos

Avaliar vícios potenciais nos resultados divulgados em inquérito telefônico de fatores de risco para doenças crônicas não transmissíveis (Vigitel), em município de baixa cobertura de domicílios com telefone fixo, é a intenção central desta tese, cujo desenvolvimento será orientado para alcançar os seguintes objetivos específicos:

- a) identificar vícios potenciais usando prevalências estimadas no inquérito domiciliar;
- b) apresentar o perfil sociodemográfico dos não usuários de telefone fixo;
- c) estimar vício de cobertura nas estimativas do Vigitel;
- d) usar procedimento de ponderação alternativo com vistas à redução dos vícios estimados no sistema Vigitel ano 2007 .

3. Métodos

Avaliação da presença de vícios potenciais nos resultados do Vigitel foi realizada por meio da comparação entre os resultados obtidos no Vigitel e no Inquérito Domiciliar. As prevalências das variáveis do Inquérito Domiciliar foram consideradas como valores de referências por tratar-se de amostra de domicílios com e sem telefone fixo. Para a realização deste estudo, foram adotados os seguintes pressupostos: as prevalências das variáveis levantadas no Inquérito Domiciliar apresentam vícios desprezíveis; as amostras do Vigitel e Inquérito Domiciliar são independentes.

O Vigitel e o Inquérito Domiciliar foram realizados no município de Rio Branco, em 2007, tendo 40% dos domicílios particulares permanentes, com pelo menos uma linha residencial, segundo os dados da Pesquisa Nacional por Amostras de Domicílios do Instituto Brasileiro de Geografia e Estatística (IBGE).

3.1 Bases de dados

3.1.1 Vigitel

Para avaliar a presença de vício nas estimativas das prevalências divulgadas pelo Sistema de Vigilância de Fatores de Risco para Doenças Crônicas (DCNT) por Inquérito Telefônico – VIGITEL – foi utilizada a base de dados do Vigitel, em 2007. A população de estudo é composta por adultos maiores de 18 anos residentes em domicílios com pelo menos uma linha residencial. Para alcançar as 2.000 entrevistas, foi sorteada uma amostra de 5.000 linhas telefônicas mediante sorteio sistemático dos

números de telefones disponíveis no cadastro eletrônico da empresa de telecomunicação. A partir dessa amostra, foram geradas 25 réplicas, de tamanho 200 cada, usando o mesmo procedimento de sorteio que gerou a amostra inicial (Figura 7a). As réplicas são utilizadas até atingir o número de entrevistas esperado (Figura 7b). Esse procedimento visa a facilitar tarefas operacionais, incluindo substituição de linhas comerciais, fora de serviço, não existente e recusas. A cada linha elegível, sorteia-se um morador adulto para entrevista. A pesquisa foi realizada no período de julho a dezembro de 2007, atingindo 71% de taxa de resposta (figura 8)

Figura 7- Distribuição do prefixo das linhas telefônicas residenciais segundo réplicas

Figura 8 – Fluxo operacional de campo do Vigitel. Município de Rio Branco-AC, 2007

O peso final atribuído a cada indivíduo entrevistado pelo Vigitel é resultado da multiplicação do inverso do número de linhas telefônicas no domicílio, número de adultos no domicílio e peso de pós-estratificação (Brasil, 2007). O inverso do número de linhas e número de adultos compõe o peso da amostra devido ao desenho. O peso de pós-estratificação é usado para corrigir possíveis vícios nas estimativas das variáveis, devido à exclusão de domicílios sem telefone fixo (KALTON, 1983).

3.1.2 Inquérito Domiciliar

A população de estudo é composta por adultos maiores de 18 anos residentes em domicílios particulares permanentes no município de Rio Branco. O inquérito domiciliar de Rio Branco adota o plano complexo de amostragem (PCA), que envolve estratificação geográfica por conglomerados em dois estágios, sendo o setor censitário a unidade primária de amostragem (UPA), e o domicílio, a unidade secundária de amostragem (USA) (KISH, 1965). Devido à restrição de recursos financeiros para a pesquisa, adotou-se a amostra mestra dos setores censitários da PNAD (IBGE),

estratégia apresentada por Silva et al. (2003). A amostra mestra é composta por 35 setores, sendo 31 deles localizados na área urbana e 4 na área rural. As amostras de domicílios foram obtidas por meio de sorteio sistemático de endereços provenientes de cópias cedidas pelo IBGE registrados pela contagem PNAD em 2005. Dessa maneira, reduz-se o custo de arrolamento dos setores censitários. Nos domicílios sorteados, foram entrevistados todos os moradores maiores de 18 anos. Essa amostra apresenta probabilidades desiguais de seleção, cuja variabilidade dos pesos amostrais é alta, em decorrência do processo de seleção, apresentando coeficiente de variação de 153%. Consequentemente, essa alta variabilidade dos pesos compromete a precisão das estimativas construídas. Para reduzir a variabilidade dos pesos amostrais (w_1), utilizou-se o procedimento de substituição dos pesos extremos pela média dos pesos amostrais. O critério utilizado para identificar os pontos extremos da variável w_1 foi o gráfico Box Plot (BUSSAB e MORETIN, 2002), ilustrado na figura 9. Verificam-se no gráfico três pontos extremos, que correspondem às UPAs 32, 33 e 34. Substituindo os pesos dessas UPAs pela média e mantendo os pesos das demais (w_2), obteve-se uma redução no coeficiente de variação de 153% para a 23%.

Figura 9 – Box plot da variável pesos amostrais (w1)

A pesquisa foi realizada no período de março de 2007 a setembro de 2008, tendo utilizado o mesmo questionário do Vigitel nos módulos de Atividade Física, Frequência Alimentar, Tabagismo, Álcool, Percepção e Morbidade Referida. No total, foram entrevistados 1.515 adultos maiores de 18 anos, atingindo 87% de taxa de resposta (Figura 10).

Para que a amostra tenha a mesma composição da população segundo sexo e escolaridade, utilizou-se o procedimento de pesos de pós-estratificação para ajustar a distribuição da amostra para a da população. Esses pesos foram construídos a partir dos dados da PNAD 2007 do município de Rio Branco. As frequências relativas foram obtidas considerando os respectivos planos de amostragem de cada inquérito. O peso de pós-estratificação é obtido por meio da divisão da frequência relativa da população sobre a amostra (FR/fr) para cada casela.

Figura 10- Fluxo operacional de campo do Inquérito Domiciliar. Município de Rio Branco-AC, 2007

3.2 Seleção das variáveis

O questionário aplicado no Inquérito Domiciliar é idêntico ao Vigitel nos módulos: Atividade Física, Frequência Alimentar, Tabagismo, Álcool, Percepção e Morbidade Referida. As variáveis (Y) selecionadas são qualitativas e foram transformadas em dicotômicas, recebendo 1 para Sim e 0 caso contrário.

O Inquérito Domiciliar foi realizado no período de março de 2007 a setembro de 2008, tendo sido realizadas 64% das entrevistas em 2007 e 36% em 2008. Para controlar a possível interferência, devido ao tempo de coleta das informações, foi

realizado o teste de hipótese (Teste t) para testar a igualdade dos parâmetros das variáveis obtidas em 2007 e 2008 ($H_0: P_{2007} = P_{2008}$). Dessa maneira, foram selecionadas somente as variáveis que não apresentaram diferença no tempo (Quadro 2).

Quadro 2 – Relação das variáveis no estudo utilizadas no estudo de comparação

Indicadores	Variáveis (Y)
Consumo alimentar	Feijão 5 ou mais dias por semana
	FLV Regularmente
	FLV Recomendado
	Carnes com gordura visível
	Leite com teor integral de gordura
	Refrigerante 5 ou mais dias por semana
Atividade física	Prática de atividade física no tempo livre
	Comportamento sedentário
Consumo de bebida alcoólica	Forma abusiva
Tabagismo	Fumante
	Ex-fumante
Auto- avaliação do estado de saúde	Ruim
Morbidade referida	Pressão alta
	Diabetes
	Infarto, derrame ou acidente vascular cerebral (AVC)
	Colesterol ou triglicérides elevado
	Asma, bronquite asmática, bronquite crônica ou enfisema
	Osteoporose

3.3 Análises Estatísticas

3.3.1 Comparação das populações usuárias e não usuárias de linhas telefônicas residenciais

Na base de dados do Inquérito Domiciliar, comparou-se a prevalência estimada da variável observada no grupo de usuários e não usuários, por meio de procedimento de teste de hipótese para diferença de média de duas populações independentes, considerando o nível de significância de 5%. As análises estatísticas foram processadas no programa STATA, versão 10.0, sendo utilizado o módulo “SVY” para obtenção de estimativas não viciadas em dados provenientes de planos complexos de amostragem.

3.3.2 Perfil do não usuário de linhas telefônicas residenciais

Para encontrar o perfil do não usuário de telefone fixo, utilizou-se a base de dados do Inquérito Domiciliar. Por se tratarem de dados provenientes de planos complexos de amostragem, as análises estatísticas foram processadas no programa STATA, versão 10.0, do qual se utilizou o módulo “SVY”, adequado para obtenção de estimativas não viciadas.

Em uma análise descritiva, a tabela de contingência e o teste qui-quadrado foram utilizados para identificar as variáveis associadas à posse de telefone fixo. Analisando-se conjuntamente as variáveis, o modelo de regressão logística múltipla (expressão 1) foi utilizado para encontrar as características que discriminassem os grupos de usuários e não usuários de telefone fixo. A variável resposta (Y) é dicotômica, sendo o valor igual a 1 atribuído para o grupo não usuário e 0 para o usuário de telefone fixo. O

quadro 3 apresenta a descrição das variáveis e seus respectivos valores. Para as variáveis categóricas, considerou-se a última categoria como referência.

Expressão 1 – Modelo de regressão logística (PAULA, 2004)

$$\log\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

em que $\pi(x)$ expressa a probabilidade de não acesso a LTR, dadas as características x_p .

Quadro 3 – Descrição das variáveis utilizadas no modelo de regressão logística múltipla, para identificar o perfil do não usuário de telefone fixo

Variável	Descrição	Domínio	Nome
Y	Usuário de telefone fixo	0 1	Sim Não
Faixa de escolaridade	Anos de estudo	1 2 4 5	0 a 4 5 a 8 9 a 11 12 ou mais
Branca	Cor da pele	0 1	não branca Branca
Faixa etária	Idade em anos	1 2 3 4	18 a 24 25 a 34 35 a 44 45 a 96
Número de moradores	Total de moradores adultos no domicílio	1 2 3	1 2 a 4 5 a 8

3.3.3 Estimação de vício de cobertura

Define-se como vício de cobertura para uma determinada variável de interesse a diferença entre as estimativas das prevalências obtidas no Vigitel (p_V) e no Inquérito Domiciliar (p_D).

Expressão 2 – Vício de cobertura

$$Vicio(p) = p_V - p_D$$

A estimativa da prevalência do Vigitel, será obtida por meio do estimador do parâmetro definido na expressão 3.

Expressão 3 – Estimador da proporção ($\hat{P}$).

$$\hat{P} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i},$$

sendo: $\hat{P}$ o estimador da proporção da variável (Y), n o número de indivíduos, w_i o peso final do i-ésimo indivíduo e y_i o valor observado da variável Y (1 se possui a característica e 0 caso contrário) para o i-ésimo indivíduo.

O estimador da variância de $\hat{P}$ é dado por:

$$V(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{\sum_{i=1}^n w_i - 1}$$

O erro quadrático médio (EQM) da prevalência é a medida do erro total do estimador (expressão 4). Na ausência de vício, o EQM expressa a variância da estimativa (KISH, 1965).

Expressão 4 – Erro quadrático médio

$$EQM_v(p) = \text{var}(p_v) + \text{vício}^2(p_v) \quad ,$$

sendo $\text{var}(p_v) = ep(p)^2$ a estimativa da variância da prevalência obtida no Vigitel.

A medida REQM Relativo (REQMR) do estimador, dada na expressão 5, mede o erro total da estimativa do Vigitel em relação à prevalência estimada no Inquérito Domiciliar (BATTAGLIA et al., 2008).

Expressão 5 – REQM Relativo

$$REQMR(p_v) = \frac{\sqrt{EQM(p_v)}}{p_D}$$

3.3.4 Redução de vício de cobertura

Os métodos de ponderação visam à correção do vício potencial, introduzido pela baixa taxa de cobertura de domicílios com telefone fixo, baseados em fontes externas.

O Vigitel adota o método de ponderação por célula (CDC, 2009), para corrigir a inferência estatística. Esse procedimento ajusta a distribuição da amostra de indivíduos residentes em domicílios com telefone fixo para a da população segundo as variáveis: sexo, faixa etária e faixa de escolaridade. Como fonte externa, foram utilizados os dados da PNAD (IBGE, 2007) para obtenção da distribuição conjunta das variáveis, na população, para construção dos pesos de pós-estratificação. As frequências relativas foram obtidas mediante uso das variáveis, disponíveis no banco de dados, que definem o plano de amostragem do inquérito. A seguir é apresentado o procedimento para o cálculo dos pesos de pós-estratificação por célula.

- 1) Calcular a frequência relativa da população do município de Rio Branco segundo sexo, idade e escolaridade, disponíveis na base de dados da PNAD (IBGE, 2007).

$$FR_{ijkl} = \frac{\sum_{l=1}^{n_{ijk}} w_{ijk} y_{ijkl}}{\sum_{i=1}^2 \sum_{j=1}^6 \sum_{k=1}^3 \sum_{l=1}^{n_{ijk}} w_{ijkl}}$$

,

sendo:

w_{ijk} o peso do indivíduo do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade;

y_{ijk} é 1 se o indivíduo é do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade; 0 caso contrário;

n_{ijk} é o número de indivíduos do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade;

i=M e F;

j=[18;24], [25;34], [35;44], [45;54], [55 ;64] e 65 anos e mais;

k= [0;8], [9;11] e 12 anos e mais.

l=1, 2, ..., n

Para identificar as entrevistas realizadas no município de Rio Branco, na base de dados da PNAD, foram executados os passos a seguir:

- a) merge dos arquivos Domicílios e Pessoas;
- b) seleção da variável UF igual a 12;
- c) seleção da variável Código da área censitária (V4107) igual 2 (autorepresentativo) e do número de municípios selecionados (V4604) igual a 1;
- d) seleção dos indivíduos com 18 anos ou mais de idade (V8005 ≥ 18).

No total, foram obtidas 36 caselas. A última casela (mulheres com 65 anos e mais e 12 e mais anos de estudo) não apresentou nenhuma ocorrência na amostra da PNAD e, por isso, essa casela foi agregada com a anterior, formando no total 35 caselas.

2) Calcular a frequência relativa na amostra do Vigitel segundo sexo, idade e escolaridade, considerando o peso da amostra, dado por:

$$w_{ijk} = \frac{1}{NLT_{ijk}} . NAD_{ijk} ,$$

sendo:

NLT_{ijk} o número de linhas telefônicas do adulto do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade;

NAD_{ijk} é o número de adultos maiores no domicílio do adulto do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade.

$$fr_{ijkl} = \frac{\sum_{l=1}^{n_{ijk}} w_{ijk} y_{ijkl}}{\sum_{i=1}^2 \sum_{j=1}^6 \sum_{k=1}^3 \sum_{l=1}^{n_{ijk}} w_{ijkl}}$$

sendo:

w_{ijk} o peso da amostra de indivíduos do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade;

y_{ijk} é 1 se indivíduo é do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade; 0 caso contrário;

n_{ijk} é o número de indivíduos do i-ésimo sexo, j-ésima faixa etária e k-ésima faixa de escolaridade;

i=M e F;

j=[18;24], [25;34]. [35;44], [45;54], [55 ;64] e 65 anos e mais;

k= [0;8], [9;11] e 12 anos e mais.

l=1, 2, ..., n

3) Cálculo do peso de pós-estratificação é dado por:

$$PesoPos_{ijk} = \frac{FR_{ijk}}{fr_{ijk}}$$

sendo:

i=M e F;

j=[18;24], [25;34], [35;44], [45;54], [55;64] e 65 anos e mais;

k= [0;8], [9;11] e 12 anos e mais.

Método de ponderação alternativo

Com vistas à redução do vício de cobertura, do Vigitel 2007, avaliou-se um método alternativo de ponderação para corrigir a inferência estatística. O método “*rake*” foi escolhido devido à facilidade de se obterem informações de fontes externas usadas na construção dos pesos. Diferentemente do método de ponderação por célula, o método “*rake*” utiliza a distribuição marginal de cada variável utilizada na construção dos pesos. Em geral, essas informações são facilmente obtidas em fontes externas. Utilizou-se a SAS macro “*rake_and_trimm*”, disponibilizada por IZRAEL et al (2009) para obter os pesos de pós-estratificação. Devido à presença de pesos com valores extremos, estes foram calibrados pelo método *Individual and Global Cap Value* (IGCV). As variáveis utilizadas na construção dos pesos encontram-se no quadro 4.

Quadro 4 – Variáveis qualitativas utilizadas na construção dos pesos de pós-estratificação

Variável	Descrição	Categorias	Nome
Faixa de escolaridade	Anos de estudo	1	0 a 8
		2	9 a 11
		3	12 ou mais
Branca	Cor da pele	1	Branca
		1	Não branca
Sexo	Sexo	1	Masculino
		2	Feminino
Faixa etária	Idade em anos	1	18 a 24
		2	25 a 34
		3	35 a 44
		4	45 a 54
		5	55 ou mais
Número de adultos	Total de adultos no domicílio	1	1
		2	2 a 4
		3	5 ou mais

Seleção das variáveis utilizadas na construção dos pesos de pós-estratificação

Escolher variáveis demográficas associadas a variável de interesse é fundamental para o bom desempenho dos pesos de pós-estratificação, para corrigir vícios de cobertura.

O modelo de regressão logística múltipla (expressão 1) foi utilizado para identificar as variáveis demográficas (quadro 5) associadas à variável de interesse (Y) (quadro 2). As variáveis explicativas são de natureza qualitativa, sendo a última categoria considerada como referência. Os modelos foram desenvolvidos com os dados do Vigitel, tendo utilizado o procedimento “*Forward selection*” para seleção dos modelos. Dessa maneira, foi possível identificar a ordem de importância das variáveis demográficas para discriminar a variável resposta.

Quadro 5 – Variáveis utilizadas no modelo de regressão logística múltipla

Variável	Descrição	Domínio	Nome
Y	Variáveis de interesse	0	Não
		1	Sim
Faixa de escolaridade	Anos de estudo	1	0 a 8
		2	9 a 11
		3	12 ou mais
Branca	Cor da pele	0	não branca
		1	Branca
Faixa etária	Idade em anos	1	18 a 24
		2	25 a 34
		3	35 a 44
		4	45 a 54
		5	55 a 64
		6	65 ou mais
Número de adultos	Total de adultos no domicílio	1	1
		2	2
		3	3
		4	4 ou mais

Para cada indivíduo foi estimada a probabilidade de ocorrer o sucesso da variável de interesse (categoria igual a Sim). Considerando a frequência relativa da categoria igual a um (1) como critério de ponto de corte na probabilidade. Os indivíduos foram divididos em dois grupos: o primeiro composto por aqueles com probabilidade inferior ao ponto de corte; e o segundo, pela probabilidade superior ou igual (figura 11). A avaliação do desempenho do modelo de regressão logística foi medida pelo acerto médio do modelo. Essa medida é dada pela razão entre a soma da quantidade de indivíduos classificados corretamente em cada categoria e o total de indivíduos (tabela 2).

Figura 11 – Ilustração do uso da probabilidade de ocorrência do sucesso para classificação do modelo de regressão logística

Tabela 2 – Classificação do modelo de regressão logística

Resposta Real	Grupo de Resposta do Modelo Regressão Logística		Total
	0	1	
0	n(0,0)		
1		n(1,1)	
Total			n

Avaliação do método alternativo de ponderação

A estratégia adotada neste trabalho foi avaliar o procedimento de ponderação pelo método “*rake*” na redução do vício das estimativas do Vigitel, apoiado no EQM como medida de desempenho do estimador.

A comparação entre as estimativas ponderadas pelos métodos “*rake*” e por célula é expressa pela razão do REQMR (expressão 6). Valores menores que 1 indicam redução de vício nas estimativas decorrente do método “*rake*”, para suavizar o vício da estimativa, devido à baixa taxa de cobertura de domicílios com telefone residencial fixo (IZRAEL et al., 2009).

Expressão 6 – Razão REQMR

$$Razão_{REQMR} = \frac{REQMR_{raking}}{REQMR_{vigitel}}$$

4. Resultados

A avaliação da presença de vícios potenciais nas prevalências estimadas do Vigitel foi realizada por meio da comparação entre os resultados obtidos nos Inquéritos por telefone e no Domiciliar. Por se tratar de amostra de domicílios com segmentos da população com e sem telefone fixo, espera-se que as prevalências do Inquérito Domiciliar apresentem vícios desprezíveis e, por isso, foram consideradas como valores de referência. Ainda possibilita identificar variáveis associadas à posse de telefone fixo e definir o perfil demográfico do não usuário de linha telefônica residencial (LTR).

No Inquérito Domiciliar foram alcançadas 1.515 entrevistas, de adultos maiores de 18 anos no período de março de 2007 a setembro de 2008, tendo sido realizadas em 2007 64% das entrevistas, enquanto o Vigitel entrevistou 2.010 adultos maiores de 18 anos no período de julho a dezembro de 2007. Os dois estudos utilizaram o mesmo questionário nos módulos de Atividade Física, Frequência Alimentar, Tabagismo, Álcool, Percepção e Morbidade Referida. Devido à diferença no período de coleta de dados do Inquérito Domiciliar, o efeito do tempo foi controlado mediante seleção de variáveis que não apresentaram diferença entre 2007 e 2008. No total, foram selecionadas 18 variáveis.

Na comparação entre as prevalências estimativas nos grupos que possuem e não possuem LTR, verificaram-se diferenças significativas, em sete das 18 variáveis analisadas (tabela 3). Notou-se que, o grupo que possui LTR é diferente nas

prevalências de consumo de frutas, legumes e verduras (FLV) regularmente, consumo de carnes com gordura visível, prática de atividade física no tempo livre, fumante, auto-avaliação de saúde considerada ruim e nas morbidades referidas de colesterol ou triglicérides elevado e asma, bronquite asmática, bronquite crônica ou enfisema.

Tabela 3 – Diferença de prevalências das variáveis de interesse segundo posse de telefone fixo levantadas no Inquérito Domiciliar. Município de Rio Branco-AC, 2007

Variáveis	Com telefone (n=644)	Sem telefone (n=871)	Diferença	
	Prevalência (p ₁)	Prevalência (p ₂)	p ₁ -p ₂	p-valor
Consumo alimentar				
Feijão 5 ou mais dias por semana	62.89	62.39	0.50	0.85
FLV Regularmente	24.33	11.21	13.13	0.00
FLV Recomendado	5.79	4.10	1.69	0.16
Carnes com gordura visível	28.74	34.21	-5.46	0.03
Leite com teor integral de gordura	68.45	71.93	-3.48	0.16
Refrigerante 5 ou mais dias por semana	40.16	40.60	-0.45	0.87
Atividade física				
Prática de atividade física no tempo livre	13.45	9.61	3.84	0.02
Assiste televisão por três ou mais horas em cinco ou mais dias da semana	20.10	23.60	3.50	0.11
Consumo de bebida alcoólica				
Forma abusiva	17.45	17.11	0.35	0.87
Tabagismo				
Fumante	13.49	22.13	-8.65	0.00
Ex-fumante	30.69	29.41	1.28	0.64
Auto avaliação do estado de saúde				
Ruim	3.39	8.30	-4.91	0.00
Morbidade referida				
Hipertensão arterial	28.82	24.56	4.25	0.08
Diabetes	4.78	4.85	-0.07	0.95
Infarto, derrame ou acidente vascular cerebral (AVC)	3.28	3.17	0.10	0.92
Colesterol ou triglicérides elevado	19.10	14.32	4.78	0.02
Asma, bronquite asmática, bronquite crônica ou enfisema	11.65	16.56	-4.92	0.01
Osteoporose	4.85	3.44	1.41	0.17

Explorando as relações entre as variáveis demográficas e a posse de LTR, por meio da distribuição bivariada, foram identificadas associações significativas mediante uso da estatística qui-quadrado de Pearson. Comparando as porcentagens das variáveis demográficas em relação ao total de não usuários de telefone fixo, verifica-se maior concentração de não usuários nas três primeiras faixas de escolaridade, na cor da pele declarada negra ou parda, nas faixas etárias de 18 a 34 anos e nos domicílios com um ou dois moradores (tabela 4).

Tabela 4 – Distribuição bivariada da posse de telefone fixo segundo características demográficas do Inquérito Domiciliar. Município de Rio Branco-AC, 2007

Variáveis	Posse de LTR				χ^2	p-valor
	Não		Sim			
	%	n	%	n		
Sexo						
Masculino	61.64	389	38.46	267	2.26	0.06
Feminino	56.73	480	43.27	373		
Faixa escolaridade						
[0;4]	74.21	197	25.79	76	107.59	0.00
[5;8]	73.14	218	26.86	92		
[9;11]	66.08	144	33.92	79		
[12;20]	45.48	310	54.52	393		
Faixa etária						
[18;24]	61.08	216	38.92	151	34.95	0.00
[25;34]	68.33	256	31.67	124		
[35;44]	59.47	178	40.53	120		
[45;54]	45.75	88	54.25	110		
[55;64]	51.16	73	48.84	75		
[65;90]	51.15	58	48.85	60		
Cor da pele						
Branca	42.79	108	57.21	143	34.04	0.00
Negra	67.58	45	32.42	27		
Parda	62.00	711	38.00	464		
Outras	45.17	5	54.83	6		
Número de moradores						
1	76.39	148	23.61	49	75.63	0.00
2	65.05	392	34.95	231		
3	46.66	161	53.34	193		
4	43.47	92	56.53	116		
5 e mais	59.82	76	40.18	51		
Total	59.03	869	40.97	640		

Os resultados do modelo de regressão logística múltipla, expressos pelas razões de chances, mostram que a chance de um adulto não possuir LTR diminui à medida que aumenta a escolaridade. A mesma relação é encontrada para o número de domicílios e para a faixa etária. Entre aqueles declarados com cor da pele não branca a chance de não possuir LTR é quase duas vezes a daqueles declarados de cor da pele branca (tabela 5).

Tabela 5 – Estimativa da razão de chance associada à não posse de LTR, obtida no modelo de regressão logística múltipla. Inquérito Domiciliar realizado no município de Rio Branco-AC, 2007

Variáveis	Razão de chance	Erro padrão	T	P>t
Faixa de escolaridade				
[0;4]	7.16	1.58	8.93	0.00
[5;8]	4.62	1.19	5.93	0.00
[9;11]	2.59	0.51	4.86	0.00
[12;20]	1.00			
Faixa de número de moradores				
1	3.74	0.78	6.33	0.00
2	1.97	0.39	3.38	0.00
[3;8]	1.00			
Faixa etária				
[18;24]	4.43	0.94	7.00	0.00
[25;34]	3.93	0.80	6.74	0.00
[35;44]	2.04	0.42	3.46	0.00
[45;96]	1.00			
Cor da pele				
Não branca	1.90	0.44	2.77	0.01
Branca	1.00			

Na avaliação de vícios potenciais do Vigitel, os resultados apontam em 12 das 18 prevalências, magnitudes de vícios superiores a 1% e inferiores a -1%. No que se refere ao módulo de consumo alimentar, cabem destacar as maiores magnitudes de vícios

encontradas e no fato de que o Vigitel subestima as prevalências, exceto para o consumo de feijão em cinco ou mais dias por semana. No que tange à prática de atividade física, o Vigitel superestima a prevalência em quase 7%, enquanto as prevalências do consumo abusivo de bebida alcoólica, de fumantes e de ex-fumantes são subestimadas. Para o módulo de morbidade referida, as prevalências são subestimadas, exceto osteoporose.

Devido à presença de vício, a validade da prevalência, expressa pela raiz quadrada do EQM, é menor para o consumo de feijão cinco ou mais dias por semana, para o consumo de leite com teor integral de gordura, para o consumo de refrigerantes em cinco ou mais dias por semana, para a prática de atividade física no tempo livre, para os diagnósticos de morbididades referidas de hipertensão arterial e de asma, bronquite asmática, bronquite crônica ou enfisema. No entanto, cabe ressaltar que a magnitude do vício da prevalência da prática de atividade física no tempo livre é um pouco mais da $\frac{1}{2}$ do valor de referência, medida expressa pela REQMR. E nas prevalências inferiores a 6%, do Vigitel, as magnitudes dos vícios representam quase $\frac{1}{3}$ do valor de referência (tabela 6).

Tabela 6 – Vício e erro quadrático médio (EQM) das prevalências das variáveis de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007

Variáveis	Inquérito	Vigitel (n=2.010)					
	Domiciliar	%	Variância	Vício ⁽²⁾	EQM	$\sqrt{EQM}$	REQMR
	%				$Var + Vicio^2$		$\frac{\sqrt{EQM}}{P_D} * 100$
Consumo alimentar							
Feijão 5 ou mais dias por semana	62.59	66.74	1.64	4.15	18.87	4.34	6.94
FLV Regularmente	16.58	14.96	1.02	-1.63	3.66	1.91	11.54
FLV Recomendado	4.79	4.01	0.43	-0.78	1.05	1.02	21.36
Carnes com gordura visível	31.97	30.46	1.76	-1.51	4.03	2.01	6.28
Leite com teor integral	70.51	61.98	1.87	-8.52	74.50	8.63	12.24
Refrigerante 5 ou mais dias por semana	40.42	31.62	1.75	-8.81	79.30	8.90	22.03
Atividade física							
Prática de atividade física no tempo livre	11.18	17.81	1.26	6.63	45.26	6.73	60.17
Comportamento sedentário	22.13	22.10	1.32	-0.03	1.32	1.15	5.20
Consumo de bebida alcoólica							
Forma abusiva	17.25	15.13	1.08	-2.12	5.56	2.36	13.67
Tabagismo							
Fumante	18.59	16.34	1.28	-2.25	6.34	2.52	13.55
Ex-fumante	29.97	26.66	1.50	-3.32	12.50	3.53	11.79
Auto avaliação do estado de saúde							
Ruim	6.29	6.02	0.41	-0.27	0.48	0.69	11.01
Morbidade referida							
Hipertensão arterial	26.30	20.33	1.13	-5.97	36.81	6.07	23.07
Diabetes	4.82	4.04	0.22	-0.78	0.82	0.91	18.83
Infarto, derrame ou acidente vascular	3.22	2.36	0.20	-0.85	0.93	0.97	30.02
Colesterol ou triglicérides elevado	16.28	15.71	0.92	-0.57	1.24	1.11	6.83
Asma, bronquite asmática, bronquite	14.55	11.40	0.88	-3.16	10.83	3.29	22.61
Osteoporose	3.79	5.03	0.37	1.24	1.90	1.38	36.38

⁽¹⁾ peso da amostra e pós estratificação segundo idade, sexo e anos de estudo (PNAD 2007), considerando 36 caselas

⁽²⁾ Vício é estimado como a diferença entre as estimativas do Vigitel e Domiciliar

⁽⁴⁾ Vício relativo é estimado como Vício dividido pela estimativa do Indicador obtido no Inquérito Domiciliar vezes 100

Devido à exclusão dos domicílios sem telefone fixo, observa-se que a pirâmide etária da amostra é diferente da população com e sem telefone fixo (figura 12). Essa diferença também ocorre para a variável anos de estudo (figura 13). Usando o procedimento de ponderação para corrigir a inferência estatística, a distribuição da amostra do Vigitel é ajustada para a da população, obtida na base de dados da PNAD (IBGE, 2007), mediante seleção das variáveis de sexo, idade e anos de estudo. O uso de pesos de pós-estratificação, nas análises estatísticas, faz com que a pirâmide etária da amostra seja igual à da população. O mesmo ocorre para a frequência da variável anos de estudo.

Figura 12 - Pirâmide etária. Município de Rio Branco-AC, 2007

Figura 13- Anos de estudo na amostra e na população. Município de Rio Branco-AC, 2007

Comparando as prevalências com e sem peso de pós-estratificação, é possível medir o impacto dos pesos. Verifica-se que pesos aumentam as prevalências nos módulos de tabagismo, no consumo de bebida alcoólica e no consumo alimentar, exceto para o consumo de FLV regularmente. Por outro lado, os pesos diminuem quatro das seis prevalências que compõem o módulo de morbidade referida. Ainda verifica-se que os pesos alteram somente as casas decimais das prevalências de atividade física, do consumo de leite com teor integral de gordura e da auto-avaliação do estado de saúde como ruim (tabela 7).

Tabela 7 – Diferença relativa entre as prevalências ponderadas e erros padrão segundo variável de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007

Variáveis	Peso amostral		Peso final ¹		Diferença relativa
	(%) (A)	Erro padrão	(%) (B)	Erro padrão	$\frac{(B - A)}{A}$
Consumo alimentar					
Feijão 5 ou mais dias por semana	64.01	1.17	66.74	1.28	0.04
FLV Regularmente	15.33	0.89	14.96	1.01	-0.02
FLV Recomendado	3.77	0.51	4.01	0.66	0.06
Carnes com gordura visível	28.44	1.12	30.46	1.33	0.07
Leite com teor integral	61.82	1.19	61.98	1.37	0.00
Refrigerante 5 ou mais dias por semana	30.07	1.13	31.62	1.32	0.05
Atividade física					
Prática de atividade física no tempo livre	17.71	0.95	17.81	1.12	0.01
Comportamento sedentário	22.60	1.00	22.10	1.17	-0.02
Consumo de bebida alcoólica					
Forma abusiva	14.58	0.87	15.13	1.04	0.04
Tabagismo					
Fumante	13.47	0.82	16.34	1.13	0.21
Ex-fumante	26.13	1.07	26.66	1.23	0.02
Auto avaliação do estado de saúde					
Ruim	6.01	0.56	6.02	0.64	0.00
Morbidade referida					
Hipertensão arterial	22.16	1.01	20.33	1.06	-0.08
Diabetes	4.78	0.52	4.04	0.47	-0.15
Infarto, derrame ou acidente vascular cerebral (AVC)	2.05	0.36	2.36	0.45	0.15
Colesterol ou triglicérides elevado	17.61	0.92	15.71	0.96	-0.11
Asma, bronquite asmática, bronquite crônica ou enfisema	11.07	0.77	11.40	0.94	0.03
Osteoporose	4.69	0.52	3.80	0.51	-0.19

⁽¹⁾ peso final = peso da amostra e pós estratificação segundo idade, sexo e anos de estudo (PNAD 2007), considerando 35 caselas

Na avaliação do efeito do método de ponderação, como procedimento para corrigir a inferência estatística, foram comparados os vícios nas prevalências com e sem ajustes de pesos de pós-estratificação. Das sete variáveis associadas à posse de telefone fixo, indicadas pelo Inquérito Domiciliar, três (Carnes com gordura visível, Fumante e Colesterol) apresentam redução do vício após uso dos pesos (figura 14). No entanto, as prevalências ponderadas de consumo de feijão em cinco ou mais dias por semana e de diagnóstico de morbididade referida de hipertensão arterial apresentam maior vício em relação às não ponderadas. Esse resultado mostra que os pesos de pós-estratificação eliminam parcialmente os vícios introduzidos pela baixa taxa de cobertura de domicílios com telefone fixo, mas, por outro lado, o uso desses pesos quando não necessário, aumenta o vício.

Figura 14 - Vício absoluto das prevalências não ponderadas e ponderadas pelo peso de pós-estratificação do Vigitel. Município de Rio Branco-AC, 2007

Com o intuito de melhorar o desempenho do método de ponderação, na redução do vício, foram selecionadas as variáveis demográficas associadas a cada variável de interesse. A escolha de variáveis externas utilizadas na construção dos pesos de pós-estratificação foi realizada mediante seleção de variáveis obtidas em cada modelo de regressão logística múltipla (tabela 8). A análise das variáveis demográficas selecionadas pelos modelos mostra que a variável Idade discrimina 16 das 18 variáveis respostas, seguida da variável Sexo em 11 e Escolaridade em 10. Nota-se que a combinação das variáveis Idade, Sexo e Escolaridade foi significativa em seis modelos, e as variáveis Idade e Sexo, em quatro modelos. Na avaliação do percentual de acerto médio de classificação correta de cada categoria resposta do modelo, observa-se, em seis modelos, o acerto médio inferior a 60%, considerado baixo poder de discriminação da variável resposta. As tabelas com as estimativas das variáveis e erro padrão para cada modelo ajustado encontram-se no Anexo 1.

Tabela 8 – Variáveis explicativas por ordem de entrada e acerto médio do modelo de regressão logística, segundo variável de resposta levantadas no Vigitel. Município de Rio Branco-AC, 2007

Variável resposta	Variáveis explicativas	Acerto médio (%)
Feijão 5 ou mais dias por semana	Sexo, escolaridade e idade	58
FLV Regularmente	Idade	41
FLV Recomendado	Idade e sexo	68
Carnes com gordura visível	Sexo e idade	65
Leite com teor integral de gordura	Idade	59
Refrigerante 5 ou mais dias por semana	Idade, sexo e escolaridade	58
Prática de atividade física no tempo livre	Idade, sexo e escolaridade	59
Comportamento sedentário	Escolaridade	61
Forma abusiva	Sexo, idade e escolaridade	68
Fumante	Escolaridade e idade	66
Ex-fumante	Idade e escolaridade	66
Ruim	Sexo, escolaridade, cor e adultos	61
Hipertensão arterial	Idade e sexo	75
Diabetes	Idade	70
Infarto, derrame ou acidente vascular cerebral (AVC)	Idade, escolaridade e cor	71
Colesterol ou triglicérides elevado	Idade e sexo	64
Asma, bronquite asmática, bronquite crônica ou Enfisema	Idade, sexo, escolaridade e cor	51
Osteoporose	Idade, sexo e escolaridade	82

Com vistas à redução dos vícios estimados no Vigitel, quatro pesos de pós-estratificação foram calculados pelo método de ponderação “rake”. Devido à presença de pesos com valores extremos, utilizou-se o método IGCV para aparar e redistribuir os pesos. Esses pesos foram obtidos mediante uso da SAS Macro “rake and trimm” e das variáveis demográficas apontadas pelo modelo de regressão logística. O primeiro peso (wr1) foi construído a partir das variáveis idade e sexo; no segundo peso (wr2), foi acrescentada a variável escolaridade; no terceiro (wr3), acrescentou-se a variável cor da pele declarada branca e, no último peso (wr4), acrescentou-se a faixa de número de adultos no domicílio. A tabela 9 apresenta as prevalências ponderadas pelos pesos de pós-estratificação de célula (wc) e “rake” (wr). Observam-se flutuações nas estimativas decorrentes do processo de ponderação dos dados (tabela 9). As saídas do programa encontram-se no Anexo 2.

Tabela 9 – Prevalências estimadas para diferentes pesos de pós-estratificação das variáveis de interesse levantadas no Vigitel. Município de Rio Branco-AC, 2007

Variáveis	P _{wc}	P _{wr(1)}	P _{wr(2)}	P _{wr(3)}	P _{wr(4)}
Consumo alimentar					
Feijão 5 ou mais dias por semana	66.74	64.24	66.74	64.16	66.20
FLV Regularmente	14.96	15.37	14.87	14.94	14.56
FLV Recomendado	4.01	3.48	3.82	3.48	3.69
Carnes com gordura visível	30.46	29.08	30.41	28.32	30.53
Leite com teor integral	61.98	61.50	61.82	61.12	61.34
Refrigerante 5 ou mais dias por semana	31.62	30.49	32.20	30.33	31.85
Atividade física					
Prática de atividade física no tempo livre	17.81	18.09	17.95	17.86	18.43
Assiste televisão por três ou mais horas em cinco ou mais dias da semana	22.10	22.65	21.99	22.51	22.62
Consumo de bebida alcoólica					
Forma abusiva	15.13	16.07	15.51	14.96	16.47
Tabagismo					
Fumante	16.34	14.27	16.66	14.98	16.50
Ex-fumante	26.66	26.38	26.75	26.37	25.32
Auto avaliação do estado de saúde					
Ruim	6.02	6.04	6.44	6.52	6.26
Morbidade referida					
Hipertensão arterial	20.33	21.98	20.22	22.07	20.26
Diabetes	4.04	4.83	4.23	4.84	4.23
Infarto, derrame ou acidente vascular cerebral (AVC)	2.36	1.84	2.21	1.97	1.88
Colesterol ou triglicérides elevado	15.71	17.66	16.05	18.12	16.05
Asma, bronquite asmática, bronquite crônica ou enfisema	11.40	10.74	11.29	11.06	10.94
Osteoporose	5.03	4.21	3.96	4.63	4.05

Notas: wc = pós estratificação do Vigitel

wr = peso de pós estratificação pelo método rake and trimm

(1) idade e sexo

(2) idade, sexo e escolaridade

(3) idade, sexo, escolaridade e cor da pele branca

(4) idade, sexo, escolaridade, cor da pele branca e número de adultos

A tabela 10 apresenta os resultados da comparação entre o REQMR das prevalências obtidas pelo método alternativo e pelo Vigitel, expresso pela razão. Valor inferior a um indica que a magnitude do vício da prevalência obtida pelo método alternativo é menor que a do Vigitel, valor igual a um indica empate e valor acima de um indica maior magnitude de vício. A análise das prevalências ajustadas por idade e sexo mostra (a) no módulo de consumo alimentar, valor inferior ou igual a um em cinco das seis prevalências, (b) no módulo de atividade física e na auto-avaliação do estado de saúde como ruim, que as magnitudes de vícios são maiores que a do Vigitel; (c) na prevalência do consumo de bebida alcoólica, menor magnitude de vício em relação ao do Vigitel e (d) nas prevalências de morbidade auto-referida, menores ou iguais magnitudes de vício em quatro das seis variáveis que compõem o módulo. Ao analisar a prevalência ajustada para idade, sexo e escolaridade, nota-se, para a maioria das variáveis que compõem o módulo de consumo alimentar, aumento na magnitude de vício, enquanto, para o módulo de atividade física e de tabagismo, a magnitude do vício é menor. No módulo de morbidade auto-referida, as prevalências ajustadas por idade, sexo e escolaridade apresentam menores magnitudes de vícios, exceto para colesterol ou triglicérides elevado.

Cabe ressaltar que, para as prevalências de consumo de feijão em cinco ou mais dias por semana e nas morbidades referidas de diabetes e de osteoporose, o método alternativo de ponderação produz prevalências com menor magnitude de vício em todas as combinações de variáveis usadas na construção dos pesos de pós-estratificação.

Tabela 10 – Razão entre REQMR da prevalência obtida para diferentes pesos de pós estratificação pelo método “*rake and trimm*” e Vigitel. Município de Rio Branco-AC, 2007

Variáveis	Pesos de pós estratificação			
	wr(1)	wr(2)	wr(3)	wr(4)
Consumo alimentar				
Feijão 5 ou mais dias por semana	0.6	0.7	0.8	0.9
FLV Regularmente	0.7	1.2	1.2	1.2
FLV Recomendado	1.0	1.4	1.2	1.2
Carnes com gordura visível	1.2	1.7	1.0	1.0
Leite com teor integral	1.0	1.2	1.1	1.1
Refrigerante 5 ou mais dias por semana	1.0	1.1	1.0	1.0
Atividade física				
Prática de atividade física no tempo livre	1.1	1.0	1.1	1.1
Assiste televisão por três ou mais horas em cinco ou mais dias da semana	1.1	0.9	1.1	1.1
Consumo de bebida alcoólica				
Forma abusiva	0.7	1.2	0.6	0.6
Tabagismo				
Fumante	2.0	0.9	0.9	0.9
Ex-fumante	1.5	0.6	1.3	1.4
Auto avaliação do estado de saúde				
Ruim	1.2	1.2	0.9	0.9
Morbidade referida				
Hipertensão arterial	1.0	0.6	1.0	1.0
Diabetes	0.8	0.6	0.8	0.8
Infarto, derrame ou acidente vascular cerebral (AVC)	1.5	1.0	1.4	1.4
Colesterol ou triglicérides elevado	0.8	1.8	0.8	0.8
Asma, bronquite asmática, bronquite crônica ou enfisema	1.2	1.1	1.1	1.1
Osteoporose	0.3	0.8	0.4	0.4

Notas: (1) idade e sexo

(2) idade, sexo e escolaridade

(3) idade, sexo, escolaridade e cor da pele branca

(4) idade, sexo, escolaridade, cor da pele branca e número de número de adultos

casela hachurada indica redução/igual no vício da estimativa do Vigitel

5. Discussão

Em 2007, apenas 41% dos domicílios particulares permanentes, localizados na capital de Rio Branco, possuíam telefone fixo, segundo dados do Inquérito Domiciliar. Os resultados desse inquérito apontam sete variáveis com potencial a vícios nas estimativas, devido à baixa taxa de cobertura de domicílios com telefone fixo. As variáveis são: Consumo de Frutas, Legumes e Verduras (FLV) Regularmente, Consumo de Carnes com gordura visível, Prática de atividade física no tempo livre, Fumante, Auto-avaliação de saúde considerada ruim e nas morbidades referidas de Colesterol ou triglicérides elevado e Asma, bronquite asmática, bronquite crônica ou enfisema.

Os resultados do Inquérito Domiciliar mostram que os não usuários de telefone fixo estão concentrados nas classes de menor escolaridade, naqueles declarados de cor não branca, entre os indivíduos de 18 a 34 anos de idade e nos domicílios com um ou dois moradores. Esses resultados são consistentes com outros estudos que identificaram o perfil do não usuário de telefone fixo (THORNBERRY e MASSEY, 1978; BATTAGLIA et al., 2008).

Para corrigir a inferência estatística, o Vigitel adota o método de ponderação por célula (CDC, 2009) para suavizar os vícios, devido à baixa taxa de cobertura de domicílios com linha telefônica. Das 18 variáveis selecionadas para o estudo, em 12 delas foram detectadas magnitudes de vícios absolutos, superiores a 1%. Devido à presença de vícios, a prevalência estimada pelo Vigitel apresenta baixa acurácia para o consumo de feijão em cinco ou mais dias por semana, para o consumo de leite com teor integral de gordura, para o consumo de refrigerantes em cinco ou mais dias por semana,

para a prática de atividade física no tempo livre, para os diagnósticos de morbidade auto-referida de hipertensão arterial e de asma, bronquite asmática, bronquite crônica ou enfisema.

Em geral, o Vigitel subestima as prevalências que compõem o grupo de fator de risco. Este resultado concorda com outros estudos que avaliaram o efeito dos pesos para corrigir vício das estimativas devido à ausência de resposta (BATTAGLIA et al., 2008). BELÁND e ST-PIERRE (2008) encontraram resultados semelhantes para prevalência de Fumantes e Consumo de bebida alcoólica.

Nas variáveis indicadas pelo Inquérito Domiciliar com potencial a vício nas estimativas, os pesos de pós-estratificação do Vigitel eliminam parcialmente os vícios introduzidos pela baixa taxa de cobertura de domicílios com telefone fixo, em três das sete variáveis (consumo de carnes com gordura visível, de fumantes e do diagnóstico auto-referido de colesterol ou triglicérides elevado).

Cabe ressaltar que o uso dos pesos quando não necessário potencializou o vício, fato observado nas prevalências de consumo de feijão cinco ou mais dias por semana, refrigerante cinco ou mais dias por semana e diagnóstico de hipertensão arterial.

No trabalho apresentado por BATTAGLIA et al. (2008), para melhorar o desempenho dos pesos de pós-estratificação na redução do vício do “*Behavioral Risk Factor Surveillance System*” (BRFSS) (CDC, 2009), devido à alta taxa de não resposta, os autores selecionaram variáveis sociodemográficas associadas às variáveis de interesse e usaram o método “rake” para construção dos pesos de pós-estratificação. Esses pesos foram calibrados pelo método “*trimming*”, para suavizar os pesos com valores extremos. Esse procedimento é amplamente utilizado para diminuir a variabilidade dos pesos. Com isso, os autores conseguiram redução no vício das

estimativas, devido à ausência de resposta de 50%, cuja taxa de cobertura de domicílios sem telefone fixo corresponde a 16%.

Seguindo o procedimento proposto por BATTAGLIA et al. (2008), foram identificadas as variáveis demográficas associadas às variáveis de interesse. Devido à grande quantidade de variáveis levantadas no Vigitel, observou-se que a combinação das variáveis demográficas associadas a cada variável de interesse não é constante. Observou-se ainda, para a maioria dos modelos ajustados, que as variáveis demográficas apresentam baixo poder de discriminação das variáveis de interesse. Destaca-se o modelo ajustado para o diagnóstico auto-referido de osteoporose com acerto médio acima de 80%.

Os resultados do desempenho do método de ponderação “*rake and trim*”, como método alternativo com vistas à redução do vício do Vigitel, mostram que as prevalências ajustadas para idade e sexo apresentam igual ou melhor desempenho para os módulos de consumo alimentar e de bebida alcoólica, enquanto as prevalências ajustadas para idade, sexo e escolaridade são iguais ou melhores para os módulos de atividade física e de tabagismo. O uso de único peso de pós-estratificação pelo método “*rake and trim*” reduzem o vício para 8 das 18 variáveis analisadas. No entanto, ao considerar dois pesos de pós-estratificação, o número sobe para 13 variáveis. Assim, para reduzir o vício do Vigitel, seriam necessários pelo menos dois pesos de pós-estratificação. O primeiro peso, composto por idade e sexo para o conjunto de variáveis: consumo de feijão 5 ou mais dias por semana, FLV regularmente, leite com teor integral de gordura, carne com gordura visível, consumo de bebida alcoólica de forma abusiva, diagnóstico de diabetes, auto-avaliação do estado de saúde ruim, colesterol e osteoporose. O segundo peso composto por idade, sexo e escolaridade para o conjunto

de variáveis: prática de atividade física no tempo livre, comportamento sedentário, fumante, ex-fumante e diagnóstico de hipertensão, infarto e asma.

Para eliminar completamente os vícios nas estimativas em municípios de baixa taxa de cobertura de domicílios com telefone fixo, torna-se necessária a inclusão dos domicílios sem telefone fixo. No entanto, alto custo, tempo e questões operacionais de campo são obstáculos que inviabilizam a utilização do Inquérito Domiciliar para o monitoramento anual dos fatores de riscos e proteção para DCNT. Métodos alternativos que utilizam mais de um cadastro para fins de sorteio, conhecidos como inquérito “*dual-frame*” (GROVES e LEPKOWSKI, 1985) e “*multiple mode and frame*” (BRICK e LEPKOWSKI, 2008) apresentam algumas vantagens em relação ao Inquérito por telefone. Mas FLORES-CERVANTES et al. (2002) alertam que, para inquéritos com baixa taxa de domicílios com telefone fixo, o inquérito “*dual frame*” tem um custo relativamente alto, o que também inviabiliza seu uso.

6. Conclusão

Passados cinco anos desde sua implantação, o Vigitel está consolidado como ferramenta de coleta de dados. Suas estimativas de prevalências são utilizadas para o planejamento de políticas públicas e diretrizes de Programas de Promoção de Saúde e Prevenção de Riscos e Doenças nas capitais brasileiras.

Em Rio Branco, os não usuários de telefones fixos estão concentrados nas classes sociais menos favorecidas, apresentam maiores prevalências nas variáveis associadas a fatores de risco, como tabagismo e consumo de carne com gordura visível. No que se refere a fatores de proteção, os não usuários de telefone fixo apresentam menores prevalências nas variáveis de prática de atividade física no tempo livre e consumo de FLV regularmente.

Em geral, o Vigitel subestima as prevalências das variáveis para a população. Os pesos de pós-estratificação eliminam parcialmente o vício, cuja origem é proveniente de baixa taxa de cobertura de domicílios com telefone fixo. Por outro lado, o uso desses pesos, quando não necessário, potencializou o vício. Fato observado na prevalência de consumo de feijão em cinco ou mais dias por semana e diagnóstico de hipertensão arterial, cujas variáveis não estão associadas à posse de telefone.

Tendo em vista que as prevalências obtidas no inquérito por telefone, em municípios de baixa taxa de cobertura de domicílios com telefone fixo, são subestimadas, torna-se necessária a implementação de novo método de ponderação e estratégia de seleção de variáveis externas, para construção de pesos de pós-estratificação, que minimizem o vício.

7. Referências Bibliográficas

BATTAGLIA, M.P. et al. Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing). Paper presented at the annual meeting of the American Association for Public Opinion Research, Pointe Hilton Tapatio Cliffs, Phoenix, Arizona, 2004. Disponível em <URL <http://www.amstat.org/sections/srms/proceedings/y2004/files/Jsm2004-000074.pdf>> [2010 dez 20].

BATTAGLIA, M.P. et al. Improving Standard Poststratification Techniques for Random-Digit-Dialing Telephone Surveys. Survey Research Methods (2008). Vol.2, No.1, pp. 11-19. ISSN 1864-3361. Disponível em <URL <http://w4.ub.uni-konstanz.de/srm/article/view/597/1295>> [2010 nov 24].

BELÁND Y.; ST-PIERRE M. Mode Effects in the Canadian Community Health Survey: A Comparison of CATI and CAPI. **Advances in Telephone Survey Methodology**. Edited by James M. Lepkowski et al. A John Wiley & Sons, INC. 2008.

BERNAL, R.T.I.; SILVA, N.N. da. Linhas Telefônicas Residenciais: uso em inquéritos epidemiológicos no Brasil. São Paulo; 2006 [Dissertação de Mestrado – Faculdade de Saúde Pública – USP].

BOLFARINE, H.; BUSSAB, W.O. **Elementos de Amostragem**. São Paulo: Editora Blucher, 2005.

BRASIL. VIGITEL Brasil 2006. Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sócio-demográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2006. Brasília: Ministério da Saúde; 2007.

BRICK, J.M. et al. Adjusting for Coverage Bias Using Telephone Service Interruption Data, NCES 97-336, U.S. Department of Education. National Center for Education Statistics, project otlcer. Washington, DC: 1996. Disponível em <URL <http://nces.ed.gov/pubs/97336.pdf>> [2010 nov 24].

BRICK, J.M; LEPKOWSKI J.M. **Multiple Mode and Frame Telephone Survey. Advanced in Telephone Survey Methodology.** Edited by James M. Lepkowski, Clyder Tucker, J. Michael Brick, Edith de Leew, Lilli Japac, Paul J. Lavrakas, Michael W. Link and Roberta L. Sangster. John Wiley & sons, Inc. 2008.

BUSSAB, W. de O.; MORETTIN, P.A. **Estatística Básica.** 5. Ed. – São Paulo: Saraiva, 2002.

CDC. Center for Disease Control and Prevention. Behavioral Risk Factor Surveillance System – BRFSS. Survey Data and Documentaion, 2009. Disponível em <URL http://www.cdc.gov/brfss/technical_infodata/weighting.htm> [2010 dez 12].

COCHRAN, W.G. **Sampling Techniques.** 3.ed. John Wiley & Sons. New York, 1977.

FLORES-CERVANTES, I.; BRICK, J.M.; JONES, M.E. Weighting for nontelephone household in the 2001 California Health Interview Survey. Joint Statistical Meetings – Section on Survey Research Methods, 2002.

FREITAS, M.P.S. Cadastro para seleção de amostras em pesquisas domiciliares. II ESAMP Escola de Amostragem e Metodologia de Pesquisa; 2009. Natal-RN; 2009. Disponível em <URL <http://www.fundaj.gov.br/geral/2esamp/cadastro-2esamp.pdf>> [2011 abril 26].

GROVES, R.M.; LEPKOWSKI, J.M.. Dual Frame, Mixed Mode Survey Designs. Journal of Official Statistics. Vol.1, No. 3, 1985, PP. 263-286. Disponível em <URL <http://www.jos.nu/Articles/article.asp>> [2010 dez 12].

GROVES, R.M. et al.. **Telephone Survey Methodology**. John Wiley & Sons, Inc, New York, 1988.

GROVES, R.M. Nonresponse rates and nonresponse bias in household survey. Public Opinion Quarterly, vol. 70, No. 5, Special Issue 2006, PP. 646-675. Disponível em <URL <http://poq.oxfordjournals.org/content/70/5/646.full>> [2009 dez 12].

IBGE – Instituto Brasileiro de Geografia e Estatística. Microdados da Pesquisa Nacional por Amostra de Domicílios, 2007.

ISAKSSON, A. et al. On the Number of Call Attempts in a Telephone Survey. In: ISI 2007 Book Abstracts: 56th Session of the International Statistical Institute; 2007 Aug 22-29; Lisboa, Portugal. CPM 085: Survey Methodology p. 431.

IZRAEL, D. et al. "A SAS Macro for Balancing a Weighted Sample." Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference, Paper 275, 2000. Disponível em <URL <http://www2.sas.com/proceedings/sugi29/207-29.pdf>> [2010 dez 12].

IZRAEL, D. et al. Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a. Raking). SAS Global Forum paper, 2009. Disponível em <URL <http://www.abtassociates.com/Page.cfm?PageID=40858&FamilyID=8600>> [2010 dez 12].

KALSBEEK W. D. et al. Effects of Nonresponse on the Mean Squared Error of Estimates from a Longitudinal Study. Proceedings of the Annual Meeting of the American Statistical Associations, August 5-9, 2003. Disponível em <URL <http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00194.pdf> > [2010 out 25].

KALTON, G. **Compensating for Missing Survey Data**. Institute for Social Research, The University of Michigan, An Arbor, Michigan, 1983.

KALTON, G.; FLORES-CERVANTES, I. Weighting Methods. Journal of Official Statistics, Vol. 19, No. 2, pp. 81-97, 2003.

KERCKHOVE, V. de. et al. An Evaluation of Bias in the 2007 National Household Education Surveys Program: Results From a Special Data Collection Effort (NCES 2009-029). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. 2008.

KISH, L. **Survey Sampling**. John Wiley & Sons, Inc. New York, 1965.

LEE, S. ; VALLIANT, R. **Weighting Telephone Samples Using Propensity Scores. Advanced in Telephone Survey Methodology.** Edited by James M. Lepkowski, Clyder Tucker, J. Michael Brick, Edith de Leew, Lilli Japac, Paul J. Lavrakas, Michael W. Link and Roberta L. Sangster. John Wiley & sons, Inc. 2008.

LEPKOWSKI, J.M. et al.. **Advanced in Telephone Survey Methodology.** John Wiley & Sons, Inc, New York, 2008.

LESSER, V.; NEWTON, L. Comparison of Delivery Methods in a Survey Distributed by Internet, Mail and Telephone. In: ISI 2007 Book Abstracts: 56th Session of the International Statistical Institute; 2007 Aug 22-29; Lisboa, Portugal. CPM 093: Internet Surveys p. 452.

LITTLE, R.J.A, Inference with Survey Weights. Journal of Official Statistics. Vol.7, No. 4, 1991, pp. 405-424 ©Statistics Sweden. Disponível em <URL <http://www.jos.nu/Articles/abstract.asp?article=74405>> [2010 out 25].

MASSEY, J,T. An Overview of Telephone Coverage. **Telephone Survey Methodology.** John Wiley & Sons, Inc, New York, 1988. Páginas 3 – 8.

MONTEIRO, C.A. et al. Monitoramento de fatores de risco para doenças crônicas por entrevistas telefônicas. Revista Saúde Pública 2005; 39(1):47-57.

MOURA, E. C. et al. Vigilância de Fatores de Risco para Doenças Crônicas por Inquérito Telefônico nas capitais dos 26 estados brasileiros e no Distrito Federal (2006). Rev. bras. epidemiol. [online]. 2008, vol.11, suppl.1, pp. 20-37. ISSN 1415-790X.

PAULA, G.A. **Modelos de Regressão com apoio computacional**. IME - USP; 2004.

POTTER, F. Survey of Procedures to Control Extreme Sampling Weights. Research Institute Research Triangle Park, NC 27709, 1988. Disponível em <URL http://www.amstat.org/sections/srms/proceedings/papers/1988_083.pdf> [2010 out 25].

SILVA, N.N da. **Amostragem Probabilística**. 2ª ed. São Paulo: EDUSP; 2001.

SILVA, N. N. da et al. Amostra mestra de Geoprocessamento: tecnologias para inqueritos domiciliares. Revista de Saúde Pública / Journal of Public Health, São Paulo, v. 37, n. 4, p. 494-502, 2003.

SILVA, N.N da. Coverage Rates of Telephone Household Lines and Potencial Bias in Epidemiological Surveys, Brazil, 2005. In: ISI 2007 Book Abstracts: 56th Session of the International Statistical Institute; 2007 Aug 22-29; Lisboa, Portugal. CPM 086: Sampling p.624.

SILVA, P.L.N. Chapter XI. Reporting and compensating for non-sampling errors for surveys in Brazil: current practice and future challenges. Household Sample Surveys in Developing and Transition Countries, 2005 (United Nations Publications. Sales No. E.05. XVII.6)

SIRKEN, M. G.; CASADY, R.J. Sampling Variance and Nonresponse Rates in Dual Frame, Mixes Mode Survey. **Telephone Survey Methodology**. John Wiley & Sons, Inc, New York, 1988. Páginas 175 – 88.

THOMAS, P. Using Telephone Data Entry as a Data Collection Mode for Business Survey. In: ISI 2007 Book Abstracts: 56th Session of the International Statistical Institute; 2007 Aug 22-29; Lisboa, Portugal. STCPM 25: The Modern Way of Making Business Statistics p.207.

THORNBERRY JR, O. T. ; MASSEY, J.T. Correcting for Undercoverage Bias in Random Digit Dialed National Health Survey. National Center for Health Statistics, 1978. Disponível em <http://www.amstat.org/sections/srms/proceedings/papers/1978_045.pdf> [2010 out 25].

TREWIN, D. e LEE, G. International Comparisons of Telephone Coverage. **Telephone Survey Methodology**. John Wiley & Sons, Inc, New York, 1988. Páginas 25 – 49.

TUCKER, C.;LEPKOWSKI, J.M. Telephone Survey: Adapting to Change. **Advanced in Telephone Survey Methodology**. Edited by James M. Lepkowski, Clyder Tucker, J. Michael Brick, Edith de Leew, Lilli Japiec, Paul J. Lavrakas, Michael W. Link and Roberta L. Sangster. John Wiley & sons, Inc. 2008.

ÜSTUN et al. Chapter X. Quality assurance in surveys: standards, guidelines and procedures. Household Sample Surveys in Developing and Transition Countries, 2005 (United Nations Publications. Sales No. E.05. XVII.6)

VASCONCELLOS, M.T.L. Cadastros e amostragem probabilística de populações finitas. II ESAMP Escola de Amostragem e Metodologia de Pesquisa; 2009. Natal-RN; 2009. Disponível em <URL <http://www.fundaj.gov.br/geral/2esamp/cadastros-mauriciomasconcellos.pdf>> [2011 abril 26].

WALDMAN, E. A. et al. Inquéritos populacionais: aspectos metodológicos, operacionais e éticos. *Rev. bras. epidemiol.* [online]. 2008, vol.11, suppl.1, pp. 168-179. ISSN 1415-790X.

8. Anexo 1

8.1 Resultados dos modelos de regressão logística pelo método “*forward*” de seleção de variáveis

8.1.1 Consumo de feijão em cinco ou mais dias por semana

8.1.2 Consumo de feijão em cinco ou mais dias por semana

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Sexo									
	Masculino	0.63	0.06	106.32	1	0.00	1.87	45.8	68.9	54.1
	Feminino									
	Constante	0.34	0.04	84.67	1	0.00	1.40			
2	Faixa escolaridade									
	0 a 8	0.54	0.08	49.59	1	0.00	1.72			
	9 a 11	0.36	0.07	26.32	1	0.00	1.43			
	12 ou +							45.8	68.9	54.1
	Sexo									
	Masculino	0.62	0.06	102.68	1	0.00	1.86			
	Feminino									
	Constante	0.03	0.06	0.31	1	0.58	1.03			
	Faixa etária			48.02	5	0.00				
	18 a 24	0.48	0.13	12.94	1	0.00	1.62			
3	25 a 34	0.46	0.13	12.41	1	0.00	1.59			
	35 a 44	0.46	0.13	11.91	1	0.00	1.58			
	45 a 54	0.19	0.13	2.05	1	0.15	1.21			
	55 a 64	-0.19	0.14	1.76	1	0.18	0.83			
	65 ou +									
	Faixa escolaridade			78.10	2	0.00		54.4	64.1	58.0
	0 a 8	0.74	0.08	77.01	1	0.00	2.10			
	9 a 11	0.36	0.07	25.99	1	0.00	1.43			
	12 ou +									
	Sexo									
	Masculino	0.60	0.06	94.77	1	0.00	1.82			
	Feminino									
	Constante	-0.34	0.13	7.07	1	0.01	0.71			

8.1.3 Consumo de frutas, verduras e hortaliças regularmente

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	0.84	0.19	19.43	1	0.00	2.31			
	25 a 34	0.76	0.19	16.20	1	0.00	2.14			
	35 a 44	0.66	0.19	11.56	1	0.00	1.93	77.4	34.8	41.4
	45 a 54	0.26	0.21	1.62	1	0.20	1.30			
	55 a 64	0.00	0.23	0.00	1	0.99	1.00			
	65 ou +									
	Constante	-2.29	0.17	170.96	1	0.00	0.10			

8.1.4 Consumo de frutas, verduras e hortaliças recomendado

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	0.93	0.31	8.67	1	0.00	2.52			
	25 a 34	0.17	0.33	0.27	1	0.60	1.19			
	35 a 44	-0.53	0.37	2.00	1	0.16	0.59	43.4	78.7	77.4
	45 a 54	0.02	0.35	0.00	1	0.95	1.02			
	55 a 64	-0.78	0.48	2.61	1	0.11	0.46			
	65 ou +									
	Constante	-3.45	0.29	138.55	1	0.00	0.03			
2	Faixa etária									
	18 a 24	0.87	0.32	7.59	1	0.01	2.38			
	25 a 34	0.13	0.33	0.16	1	0.68	1.14			
	35 a 44	-0.57	0.37	2.30	1	0.13	0.57			
	45 a 54	-0.01	0.35	0.00	1	0.98	0.99			
	55 a 64	-0.81	0.48	2.81	1	0.09	0.45	57.1	68.7	68.3
	65 ou +									
	Sexo									
	Masculino	0.31	0.15	4.57	1	0.03	1.37			
	Feminino									
	Constante	-3.55	0.30	142.20	1	0.00	0.03			

8.1.5 Consumo de carnes com gordura visível

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Sexo									
	Masculino	0.83	0.06	176.12	1	0.00	2.30	55.1	65.2	62.3
	Feminino									
	Constante	-1.30	0.04	877.79	1	0.00	0.27			
2	Faixa etária									
	18 a 24	0.94	0.15	37.72	1	0.00	2.57			
	25 a 34	0.94	0.15	38.39	1	0.00	2.57			
	35 a 44	0.77	0.16	24.62	1	0.00	2.17			
	45 a 54	0.26	0.16	2.53	1	0.11	1.30			
	55 a 64	-0.26	0.19	1.82	1	0.18	0.77	49.7	71.9	65.0
	65 ou +									
	Sexo									
	Masculino	0.81	0.06	160.45	1	0.00	2.24			
	Feminino									
	Constante	-1.95	0.14	187.31	1	0.00	0.14			

8.1.6 Consumo de leite com teor integral de gordura

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	0.79	0.12	43.84	1	0.00	2.20			
	25 a 34	0.64	0.12	30.31	1	0.00	1.90			
	35 a 44	0.63	0.12	27.34	1	0.00	1.88	71.4	40	59.4
	45 a 54	0.31	0.12	6.35	1	0.01	1.36			
	55 a 64	0.09	0.14	0.46	1	0.50	1.10			
	65 ou +									
	Constante	-0.03	0.10	0.11	1	0.74	0.97			

8.1.7 Consumo de refrigerante em cinco ou mais dias por semana

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	1.28	0.15	72.89	1	0.00	3.59			
	25 a 34	1.01	0.15	46.12	1	0.00	2.75			
	35 a 44	0.76	0.15	24.36	1	0.00	2.13			
	45 a 54	0.39	0.16	6.00	1	0.01	1.48	59.1	58	58.3
	55 a 64	0.13	0.18	0.50	1	0.48	1.14			
	65 ou +									
	Constante	-1.64	0.14	142.83	1	0.00	0.19			
2	Faixa etária									
	18 a 24	1.22	0.15	66.42	1	0.00	3.40			
	25 a 34	0.98	0.15	42.80	1	0.00	2.66			
	35 a 44	0.72	0.15	21.85	1	0.00	2.05			
	45 a 54	0.36	0.16	5.02	1	0.03	1.43			
	55 a 64	0.10	0.18	0.28	1	0.59	1.10	68.6	50.4	55.9
	65 ou +									
	Sexo									
	Masculino	0.33	0.06	28.74	1	0.00	1.40			
	Feminino									
	Constante	-1.74	0.14	157.40	1	0.00	0.17			
3	Faixa etária									
	18 a 24	1.40	0.16	75.74	1	0.00	4.05			
	25 a 34	1.14	0.16	52.33	1	0.00	3.13			
	35 a 44	0.86	0.16	28.94	1	0.00	2.36			
	45 a 54	0.46	0.16	7.87	1	0.01	1.58			
	55 a 64	0.15	0.18	0.69	1	0.41	1.16			
	65 ou +									
	Faixa escolaridade							63.9	55.8	58.2
	0 a 8	0.40	0.09	20.30	1	0.00	1.49			
	9 a 11	0.25	0.08	10.98	1	0.00	1.28			
	12 ou +									
	Sexo									
	Masculino	0.32	0.06	25.63	1	0.00	1.37			
	Feminino									
	Constante	-2.09	0.16	172.88	1	0.00	0.12			

8.1.8 Prática de atividade física no tempo livre

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Sexo									
	Masculino	0.77	0.07	108.90	1	0.00	2.16	56.1	62.8	61.6
	Feminino									
	Constante	-1.90	0.05	1257.08	1	0.00	0.15			
2	Faixa etária			106.88	5	0.00				
	18 a 24	1.51	0.21	52.60	1	0.00	4.52			
	25 a 34	0.81	0.21	14.59	1	0.00	2.24			
	35 a 44	0.69	0.22	10.03	1	0.00	1.99			
	45 a 54	0.93	0.22	18.25	1	0.00	2.54			
	55 a 64	0.76	0.24	10.33	1	0.00	2.13	65.4	53	55.2
	65 ou +									
	Sexo									
	Masculino	0.72	0.07	93.77	1	0.00	2.06			
	Feminino									
	Constante	-2.82	0.20	199.47	1	0.00	0.06			

8.1.9 Comportamento sedentário

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa escolaridade									
	0 a 8	-0.28	0.10	8.96	1	0.00	0.75			
	9 a 11	0.49	0.08	36.43	1	0.00	1.63	53.6	61.8	60.0
	12 ou +									
	Constante	-1.38	0.07	447.45	1	0.00	0.25			
2	Faixa etária									
	18 a 24	0.17	0.15	1.24	1	0.26	1.19			
	25 a 34	-0.13	0.15	0.71	1	0.40	0.88			
	35 a 44	-0.22	0.16	2.03	1	0.15	0.80			
	45 a 54	-0.44	0.16	7.39	1	0.01	0.64			
	55 a 64	-0.53	0.18	8.40	1	0.00	0.59	53.7	63.0	60.8
	65 ou +									
	Faixa escolaridade			65.10	2	0.00				
	0 a 8	-0.18	0.10	3.21	1	0.07	0.83			
	9 a 11	0.48	0.08	34.36	1	0.00	1.61			
	12 ou +									
	Constante	-1.26	0.15	68.50	1	0.00	0.28			

8.1.10 Consumo de bebida alcoólica de forma abusiva

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Sexo									
	Masculino	1.22	0.08	218.35	1	0.00	3.39	65.9	63.8	64.07
	Feminino									
	Constante	-2.39	0.06	1353.30	1	0.00	0.09			
2	Faixa etária									
	18 a 24	3.71	0.71	27.05	1	0.00	40.94			
	25 a 34	3.57	0.71	24.99	1	0.00	35.44			
	35 a 44	3.63	0.71	25.76	1	0.00	37.60			
	45 a 54	3.17	0.72	19.47	1	0.00	23.76			
	55 a 64	2.63	0.73	12.95	1	0.00	13.89	61.3	69.7	68.5
	65 ou +									
	Sexo									
	Masculino	1.18	0.08	198.83	1	0.00	3.25			
	Feminino									
	Constante	-5.77	0.71	65.80	1	0.00	0.00			
3	Faixa etária									
	18 a 24	3.45	0.72	23.16	1	0.00	31.56			
	25 a 34	3.34	0.72	21.68	1	0.00	28.10			
	35 a 44	3.42	0.72	22.77	1	0.00	30.59			
	45 a 54	3.02	0.72	17.58	1	0.00	20.42			
	55 a 64	2.54	0.73	12.05	1	0.00	12.69			
	65 ou +									
	Faixa escolaridade							62.2	69.1	68.1
	0 a 8	-0.50	0.12	17.28	1	0.00	0.61			
	9 a 11	-0.17	0.09	3.46	1	0.06	0.84			
	12 ou +									
	Sexo									
	Masculino	1.21	0.08	205.18	1	0.00	3.34			
	Feminino									
	Constante	-5.38	0.72	56.31	1	0.00	0.00			

8.1.11 Fumante

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa escolaridade									
	0 a 8	1.36	0.12	128.82	1	0.00	3.88			
	9 a 11	0.55	0.12	19.58	1	0.00	1.72	50.3	70.9	69.8
	12 ou +									
	Constante	-2.60	0.10	636.12	1	0.00	0.07			
Step 2(b)	Faixa etária									
	18 a 24	1.13	0.23	24.72	1	0.00	3.10			
	25 a 34	0.80	0.23	12.79	1	0.00	2.24			
	35 a 44	1.51	0.22	48.57	1	0.00	4.51			
	45 a 54	1.68	0.21	60.83	1	0.00	5.35			
	55 a 64	1.21	0.23	27.85	1	0.00	3.34			
	65 ou +							66.5	66.3	66.5
	Faixa escolaridade									
	0 a 8	1.43	0.13	130.40	1	0.00	4.18			
	9 a 11	0.55	0.12	19.48	1	0.00	1.73			
	12 ou +									
	Constante	-3.81	0.23	280.56	1	0.00	0.02			

8.1.12 Ex-fumante

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	-2.78	0.15	360.23	1	0.00	0.06			
	25 a 34	-2.42	0.13	327.07	1	0.00	0.09			
	35 a 44	-1.51	0.13	144.67	1	0.00	0.22			
	45 a 54	-0.88	0.13	48.91	1	0.00	0.41			
	55 a 64	-0.40	0.14	8.34	1	0.00	0.67			
	65 ou +									
	Constante	0.46	0.10	19.71	1	0.00	1.59			
2	Faixa etária									
	18 a 24	-2.50	0.16	254.63	1	0.00	0.08			
	25 a 34	-2.15	0.14	229.49	1	0.00	0.12			
	35 a 44	-1.28	0.13	93.06	1	0.00	0.28			
	45 a 54	-0.70	0.13	28.84	1	0.00	0.50			
	55 a 64	-0.30	0.14	4.47	1	0.03	0.74			
	65 ou +							76.6	63.9	67.2
	Faixa escolaridade									
	0 a 8	0.87	0.10	77.46	1	0.00	2.39			
	9 a 11	0.63	0.10	42.83	1	0.00	1.87			
	12 ou +									
	Constante	-0.31	0.14	5.07	1	0.02	0.74			
3	Faixa etária									
	18 a 24	-2.59	0.16	266.77	1	0.00	0.08			
	25 a 34	-2.22	0.14	239.84	1	0.00	0.11			
	35 a 44	-1.33	0.13	99.91	1	0.00	0.26			
	45 a 54	-0.75	0.13	32.26	1	0.00	0.47			
	55 a 64	-0.33	0.14	5.59	1	0.02	0.72			
	65 ou +									
	Faixa escolaridade							69.5	71.4	70.8
	0 a 8	0.86	0.10	74.50	1	0.00	2.35			
	9 a 11	0.62	0.10	41.45	1	0.00	1.86			
	12 ou +									
	Sexo									
	Masculino	0.35	0.07	24.30	1	0.00	1.42			
	Feminino									
	Constante	-0.39	0.14	8.16	1	0.00	0.68			

8.1.13 Estado de saúde considerada ruim

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Sexo									
	Masculino	-0.89	0.14	41.08	1.00	0.00	0.41			
	Feminino							77.1	41.7	43.8
	Constant	-2.48	0.07	1351.28	1.00	0.00	0.08			
2	Faixa escolaridade									
	0 a 8	0.81	0.16	25.97	1.00	0.00	2.25			
	9 a 11	0.27	0.16	2.73	1.00	0.10	1.31			
	12 ou +							61.9	59.5	59.7
	Sexo									
	Masculino	-0.90	0.14	41.90	1.00	0.00	0.41			
	Feminino									
	Constante	-2.88	0.13	459.36	1.00	0.00	0.06			
3	Cor da pele									
	Branca	0.48	0.15	11.12	1.00	0.00	1.62			
	Demais									
	Faixa escolaridade									
	0 a 8	0.74	0.16	21.39	1.00	0.00	2.10			
	9 a 11	0.21	0.16	1.68	1.00	0.20	1.24	67.9	56.0	56.9
	12 ou +									
	Sexo									
4	Masculino	-0.92	0.14	43.03	1.00	0.00	0.40			
	Feminino									
	Constante	-3.19	0.17	362.53	1.00	0.00	0.04			
	Número de adultos									
	1	0.49	0.25	3.79	1.00	0.05	1.64			
	2	-0.09	0.15	0.33	1.00	0.57	0.92			
	3	0.29	0.15	3.72	1.00	0.05	1.33			
	4 ou +									
	Cor da pele									
	Branca	0.49	0.15	11.11	1.00	0.00	1.62			
	Demais							65.3	60.6	60.8
	Faixa escolaridade									
	0 a 8	0.74	0.16	21.30	1.00	0.00	2.10			
	9 a 11	0.22	0.17	1.74	1.00	0.19	1.24			
	12 ou +									
	Sexo									
	Masculino	-0.89	0.14	39.89	1.00	0.00	0.41			
	Feminino									
	Constante	-3.28	0.19	288.43	1.00	0.00	0.04			

8.1.14 Diagnóstico de hipertensão arterial

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	-3.63	0.17	453.32	1	0.00	0.03			
	25 a 34	-3.14	0.15	451.62	1	0.00	0.04			
	35 a 44	-2.27	0.14	279.08	1	0.00	0.10	68.4	77.2	75.3
	45 a 54	-1.33	0.13	103.82	1	0.00	0.26			
	55 a 64	-0.97	0.14	46.85	1	0.00	0.38			
	65 ou +									
	Constante	0.77	0.11	49.22	1	0.00	2.15			
2	Faixa etária									
	18 a 24	-3.59	0.17	439.67	1	0.00	0.03			
	25 a 34	-3.11	0.15	442.61	1	0.00	0.04			
	35 a 44	-2.24	0.14	270.17	1	0.00	0.11			
	45 a 54	-1.31	0.13	98.86	1	0.00	0.27			
	55 a 64	-0.95	0.14	43.86	1	0.00	0.39	68.4	77.2	75.3
	65 ou +									
	Sexo									
	Masculino	-0.39	0.08	24.02	1	0.00	0.68			
	Feminino									
	Constante	0.88	0.11	61.86	1	0.00	2.42			

8.1.15 Diagnóstico de diabetes

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	-4.39	0.59	54.70	1	0.00	0.01			
	25 a 34	-2.89	0.29	98.03	1	0.00	0.06			
	35 a 44	-2.10	0.24	77.20	1	0.00	0.12	82.3	69.3	70.2
	45 a 54	-0.95	0.19	25.56	1	0.00	0.39			
	55 a 64	-0.16	0.18	0.78	1	0.38	0.85			
	65 ou +									
	Constante	-1.56	0.13	136.07	1	0.00	0.21			

8.1.16 Diagnóstico de infarto, derrame ou acidente vascular cerebral

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	-2.65	0.46	33.98	1	0.00	0.07			
	25 a 34	-2.55	0.42	37.45	1	0.00	0.08			
	35 a 44	-1.77	0.34	26.64	1	0.00	0.17			
	45 a 54	-0.52	0.26	3.95	1	0.05	0.59	75.1	68.0	68.2
	55 a 64	-0.67	0.31	4.54	1	0.03	0.51			
	65 ou +									
	Constante	-2.59	0.20	169.11	1	0.00	0.07			
2	Faixa etária									
	18 a 24	-1.76	0.48	13.32	1	0.00	0.17			
	25 a 34	-1.83	0.43	17.65	1	0.00	0.16			
	35 a 44	-1.20	0.36	11.41	1	0.00	0.30			
	45 a 54	-0.09	0.27	0.11	1	0.74	0.91			
	55 a 64	-0.46	0.31	2.14	1	0.14	0.63	77.0	71.4	71.5
	65 ou +									
	Faixa escolaridade			27.45	2	0.00				
	0 a 8	0.91	0.29	10.04	1	0.00	2.49			
	9 a 11	-0.61	0.37	2.79	1	0.09	0.54			
	12 ou +									
	Constante	-3.36	0.34	98.57	1	0.00	0.03			
3	Faixa etária									
	18 a 24	-1.75	0.48	13.21	1	0.00	0.17			
	25 a 34	-1.81	0.44	17.21	1	0.00	0.16			
	35 a 44	-1.19	0.36	11.08	1	0.00	0.31			
	45 a 54	-0.05	0.27	0.03	1	0.86	0.95			
	55 a 64	-0.43	0.32	1.89	1	0.17	0.65			
	65 ou +									
	Cor da pele							81.7	70.8	71.0
	Branca	1.05	0.30	12.24	1	0.00	2.87			
	Demais									
	Faixa escolaridade			25.06	2	0.00				
	0 a 8	0.79	0.29	7.37	1	0.01	2.20			
	9 a 11	-0.71	0.37	3.75	1	0.05	0.49			
	12 ou +									
	Constante	-4.13	0.42	96.68	1	0.00	0.02			

8.1.17 Diagnóstico de colesterol ou triglicérides elevado

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária									
	18 a 24	-2.37	0.16	209.85	1.00	0.00	0.09			
	25 a 34	-2.17	0.15	207.65	1.00	0.00	0.11			
	35 a 44	-0.95	0.13	54.17	1.00	0.00	0.39	82.9	53.6	58.8
	45 a 54	-0.64	0.13	24.03	1.00	0.00	0.53			
	55 a 64	-0.08	0.14	0.32	1.00	0.57	0.92			
	65 ou +									
	Constante	-0.42	0.10	16.76	1.00	0.00	0.65			
2	Faixa etária									
	18 a 24	-2.30	0.16	195.50	1.00	0.00	0.10			
	25 a 34	-2.13	0.15	198.12	1.00	0.00	0.12			
	35 a 44	-0.90	0.13	47.67	1.00	0.00	0.41			
	45 a 54	-0.60	0.13	20.50	1.00	0.00	0.55			
	55 a 64	-0.03	0.14	0.04	1.00	0.84	0.97	74.8	61.8	64.1
	65 ou +									
	Sexo									
	Masculino	-0.55	0.08	44.18	1.00	0.00	0.58			
	Feminino									
	Constante	-0.27	0.11	6.52	1.00	0.01	0.76			

8.1.18 Diagnóstico de asma, bronquite asmática, bronquite crônica ou enfisema

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária							76.1	37.4	41.7
	18 a 24	-0.53	0.16	11.12	1	0.00	0.59			
	25 a 34	-0.49	0.15	10.02	1	0.00	0.61			
	35 a 44	-1.04	0.18	35.05	1	0.00	0.35			
	45 a 54	-1.10	0.19	34.32	1	0.00	0.33			
	55 a 64	-0.36	0.18	3.90	1	0.05	0.69			
	65 ou +									
	Constante	-1.47	0.13	128.26	1	0.00	0.23			
2	Faixa etária							57.6	58	58
	18 a 24	-0.47	0.16	8.61	1	0.00	0.63			
	25 a 34	-0.45	0.16	8.44	1	0.00	0.64			
	35 a 44	-1.00	0.18	32.27	1	0.00	0.37			
	45 a 54	-1.07	0.19	32.25	1	0.00	0.34			
	55 a 64	-0.33	0.19	3.24	1	0.07	0.72			
	65 ou +									
	Sexo									
	Masculino	-0.37	0.09	14.99	1	0.00	0.69			
	Feminino									
	Constante	-1.37	0.13	108.00	1	0.00	0.25			
3	Faixa etária							59.7	56.3	56.6
	18 a 24	-0.20	0.18	1.23	1	0.27	0.82			
	25 a 34	-0.21	0.17	1.58	1	0.21	0.81			
	35 a 44	-0.80	0.18	19.02	1	0.00	0.45			
	45 a 54	-0.91	0.19	22.50	1	0.00	0.40			
	55 a 64	-0.25	0.19	1.86	1	0.17	0.77			
	65 ou +									
	Faixa escolaridade			11.85	2	0.00				
	0 a 8	0.35	0.13	7.63	1	0.01	1.42			
	9 a 11	-0.04	0.11	0.09	1	0.76	0.97			
	12 ou +									
	Sexo									
	Masculino	-0.38	0.10	15.81	1	0.00	0.69			
	Feminino									
	Constante	-1.66	0.17	92.37	1	0.00	0.19			
4	Faixa etária							65.5	49.1	50.9
	18 a 24	-0.20	0.18	1.28	1	0.26	0.82			
	25 a 34	-0.22	0.17	1.71	1	0.19	0.80			
	35 a 44	-0.81	0.18	19.29	1	0.00	0.44			
	45 a 54	-0.93	0.19	23.38	1	0.00	0.39			
	55 a 64	-0.27	0.19	2.06	1	0.15	0.76			
	65 ou +									
	Cor da pele									
	Branca	-0.28	0.10	8.73	1	0.00	0.75			
	Demais									
	Faixa escolaridade			13.50	2	0.00				
	0 a 8	0.40	0.13	9.82	1	0.00	1.49			
	9 a 11	0.00	0.11	0.00	1	0.97	1.00			
	12 ou +									
	Sexo									
	Masculino	-0.37	0.10	15.30	1	0.00	0.69			
	Feminino									
	Constante	-1.49	0.18	68.15	1	0.00	0.22			

8.1.19 Diagnóstico de osteoporose

Modelo	Variável	B	S.E.	Wald	df	Sig.	Exp(B)	Classificação correta		
								Resposta		Global
								Sim	Não	
1	Faixa etária							91.8	70.0	71.0
	18 a 24	-3.81	0.40	91.24	1	0.00	0.02			
	25 a 34	-3.93	0.40	97.25	1	0.00	0.02			
	35 a 44	-3.86	0.43	81.38	1	0.00	0.02			
	45 a 54	-1.32	0.18	51.16	1	0.00	0.27			
	55 a 64	-0.32	0.17	3.50	1	0.06	0.72			
	65 ou +									
	Constante	-1.29	0.12	109.65	1	0.00	0.28			
2	Faixa etária			244.31	5	0.00		86.8	86.8	82.5
	18 a 24	-3.64	0.40	82.50	1	0.00	0.03			
	25 a 34	-3.86	0.40	92.83	1	0.00	0.02			
	35 a 44	-3.78	0.43	77.29	1	0.00	0.02			
	45 a 54	-1.24	0.19	42.61	1	0.00	0.29			
	55 a 64	-0.20	0.18	1.22	1	0.27	0.82			
	65 ou +									
	Sexo									
	Masculino	-2.58	0.30	73.22	1	0.00	0.08			
	Feminino									
	Constante	-0.89	0.13	46.79	1	0.00	0.41			
3	Faixa etária							86.8	82.3	82.5
	18 a 24	-3.36	0.42	64.29	1	0.00	0.03			
	25 a 34	-3.65	0.42	77.04	1	0.00	0.03			
	35 a 44	-3.59	0.44	67.34	1	0.00	0.03			
	45 a 54	-1.06	0.20	27.47	1	0.00	0.35			
	55 a 64	-0.11	0.19	0.35	1	0.56	0.90			
	65 ou +									
	Faixa escolaridade									
	0 a 8	-0.09	0.19	0.24	1	0.63	0.91			
	9 a 11	-0.98	0.24	16.50	1	0.00	0.38			
	12 ou +									
	Sexo									
	Masculino	-2.59	0.30	73.77	1	0.00	0.07			
	Feminino									
	Constante	-0.73	0.21	11.61	1	0.00	0.48			
4	Faixa etária							86.8	82.3	82.5
	18 a 24	-3.37	0.42	64.69	1	0.00	0.03			
	25 a 34	-3.66	0.42	77.39	1	0.00	0.03			
	35 a 44	-3.60	0.44	67.37	1	0.00	0.03			
	45 a 54	-1.09	0.20	28.66	1	0.00	0.34			
	55 a 64	-0.12	0.19	0.41	1	0.52	0.89			
	65 ou +									
	Cor da pele									
	Branca	-0.31	0.15	4.09	1	0.04	0.73			
	Demais									
	Faixa escolaridade									
	0 a 8	-0.05	0.19	0.07	1	0.79	0.95			
	9 a 11	-0.94	0.24	15.20	1	0.00	0.39			
	12 ou +									
	Sexo									
	Masculino	-2.57	0.30	72.67	1	0.00	0.08			
	Feminino									
	Constante	-0.54	0.23	5.42	1	0.02	0.58			

9. Anexo 2

9.1 Saídas do programa “rake and trimm”

9.1.1 Pesos de pós estratificação segundo sexo e idade

size of completed interviews: **2003**

input weight adjusted to population total: **ADULFONE_ATPT**

blue of raking input weight adjusted to population total: **99.27**

m value of raking input weight: **19.17**

im value of raking input weight: **383.50**

ent of variation of raking input weight: **0.46**

ow weight cap value (GLCV): **49.63**

ow weight cap value factor: Mean input weight times **0.5**

nigh weight cap value (GHCV): **496.35**

nigh weight cap value factor: Mean input weight times **5**

ial low weight cap value (ILCV) factor: Respondent's weight times **0.2**

ial high weight cap value (IHCV) factor: Respondent's weight times **5**

of respondents who have an individual high weight cap value less than the global low weight cap value

used in weight trimming): **0**

of respondents who have an individual low weight cap value greater than the global high weight cap value

used in weight trimming): **0**

Weighted Distribution Prior To Raking, Iteration 0

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	44025.28	47207	-3181.72	22.142	23.742	-1.600
25 a 34	49707.40	54621	-4913.60	24.999	27.470	-2.471
35 a 44	39710.95	38883	827.95	19.972	19.555	0.416
45 a 54	31753.42	29938	1815.42	15.970	15.057	0.913
55 ou +	33638.94	28187	5451.94	16.918	14.176	2.742

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
sexo						
Masculino	80674.66	92020	-11345.34	40.573	46.280	-5.706
Feminino	118161.34	106815	11346.34	59.427	53.720	5.706

Raking by Q7, iteration - 1

sexo	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
Masculino	80743.24	92020	-11276.76	40.608	46.280	-5.672
Feminino	118092.76	106815	11277.76	59.392	53.720	5.672
	198836.00	198835		100.00	100.00	

Weighted Distribution After Iteration 1

fxidl	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
18 a 24	47702.41	47207	495.41	23.991	23.742	0.249
25 a 34	54485.55	54621	-135.45	27.402	27.470	-0.068
35 a 44	38932.98	38883	49.98	19.581	19.555	0.025
45 a 54	29848.54	29938	-89.46	15.012	15.057	-0.045
55 ou +	27865.51	28187	-321.49	14.014	14.176	-0.162

sexo	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
Masculino	92020.00	92020	0.00	46.280	46.280	0.000
Feminino	106815.00	106815	-0.00	53.720	53.720	-0.000

Raking by FXID1, iteration - 2

FXID1	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
18 a 24	47702.41	47207	495.412	23.991	23.742	0.249
25 a 34	54485.55	54621	-135.448	27.402	27.470	-0.068
35 a 44	38932.98	38883	49.978	19.581	19.555	0.025
45 a 54	29848.54	29938	-89.458	15.012	15.057	-0.045
55 ou +	27865.51	28187	-321.485	14.014	14.176	-0.162
	198835.00	198836		100.00	100.00	

Raking by Q7, iteration - 2

sexo	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
Masculino	91958.63	92020	-61.3703	46.248	46.280	-0.031
Feminino	106877.37	106815	62.3703	53.752	53.720	0.031
	198836.00	198835		100.00	100.00	

Weighted Distribution After Iteration 2

fxid1	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
18 a 24	47210.37	47207	3.37	23.743	23.742	0.002
25 a 34	54620.10	54621	-0.90	27.470	27.470	-0.000
35 a 44	38883.92	38883	0.92	19.556	19.555	0.001
45 a 54	29936.68	29938	-1.32	15.056	15.057	-0.001
55 ou +	28183.91	28187	-3.09	14.175	14.176	-0.001

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	92020.00	92020	-0.00	46.280	46.280	0.000
Feminino	106815.00	106815	-0.00	53.720	53.720	-0.000

*** Program terminated at iteration 2 because all current percents differ from target percents by less than 0.025 ***

Weighted Distribution After Raking

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	47210.37	47207	3.37	23.743	23.742	0.002
25 a 34	54620.10	54621	-0.90	27.470	27.470	-0.000
35 a 44	38883.92	38883	0.92	19.556	19.555	0.001
45 a 54	29936.68	29938	-1.32	15.056	15.057	-0.001
55 ou +	28183.91	28187	-3.09	14.175	14.176	-0.001

sexo	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
Masculino	92020.00	92020	-0.00	46.280	46.280	0.000
Feminino	106815.00	106815	-0.00	53.720	53.720	-0.000

Number of Respondents Who Had Their Weights Decreased by the Trimming: **0.**

Number of Respondents Who Had Their Weights Increased by the Trimming: **266.**

Raking output weight: **rake_trim**

Weight	Mean	Min	Max	CV
ADULFONE_ATPT	99.27	19.17	383.50	0.463
rake_trim	99.27	49.63	874.22	0.706

9.1.2 Peso de pós estratificação segundo sexo, idade e escolaridade

RAKING WITH TRIMMING WEIGHT BY INDIVIDUAL AND GLOBAL CAP VALUE METHOD

Sample size of completed interviews: 2003

Raking input weight adjusted to population total: ADULFONE_ATPT

Mean value of raking input weight adjusted to population total: 99.27

Minimum value of raking input weight: 19.17

Maximum value of raking input weight: 383.50

Coefficient of variation of raking input weight: 0.46

Global low weight cap value (GLCV): 49.63

Global low weight cap value factor: Mean input weight times 0.5

Global high weight cap value (GHCV): 1091.96

Global high weight cap value factor: Mean input weight times 11

Individual low weight cap value (ILCV) factor: Respondent's weight times 0.2

Individual high weight cap value (IHCV) factor: Respondent's weight times 5

Number of respondents who have an individual high weight cap value less than the global low weight cap value

(GLCV used in weight trimming): 0

Number of respondents who have an individual low weight cap value greater than the global high weight cap value

(GHCV used in weight trimming): 0

Weighted Distribution Prior To Raking. Iteration 0

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	44025.28	47207	-3181.72	22.142	23.742	-1.600
25 a 34	49707.40	54621	-4913.60	24.999	27.470	-2.471
35 a 44	39710.95	38883	827.95	19.972	19.555	0.416
45 a 54	31753.42	29938	1815.42	15.970	15.057	0.913
55 ou +	33638.94	28187	5451.94	16.918	14.176	2.742

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
sexo						
Masculino	80674.66	92020	-11345.34	40.573	46.280	-5.706
Feminio	118161.34	106815	11346.34	59.427	53.720	5.706

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	60061.78	90932	-30870.22	30.207	45.732	-15.526
9 a 11	82726.36	70078	12648.36	41.605	35.244	6.361
12 ou +	56047.86	37825	18222.86	28.188	19.023	9.165

Raking by FXID1, iteration - 1

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXID1						
18 a 24	44025.28	47207	-3181.72	22.142	23.742	-1.600
25 a 34	49707.40	54621	-4913.60	24.999	27.470	-2.471
35 a 44	39710.95	38883	827.95	19.972	19.555	0.416
45 a 54	31753.42	29938	1815.42	15.970	15.057	0.913
55 ou +	33638.94	28187	5451.94	16.918	14.176	2.742
	198836.00	198836		100.00	100.00	

Raking by Q7, iteration - 1

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
sexo						
Masculino	80743.24	92020	-11276.76	40.608	46.280	-5.672
Feminio	118092.76	106815	11277.76	59.392	53.720	5.672
	198836.00	198835		100.00	100.00	

Raking by FXESC2, iteration - 1

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXESC2						
0 a 8	56465.11	90932	-34466.89	28.398	45.732	-17.33
9 a 11	85065.87	70078	14987.87	42.782	35.244	7.538
12 ou +	57304.02	37825	19479.02	28.820	19.023	9.797
	198835.00	198835		100.00	100.00	

Weighted Distribution After Iteration 1

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	39776.61	47207	-7430.39	20.005	23.742	-3.737
25 a 34	49417.65	54621	-5203.35	24.854	27.470	-2.617
35 a 44	38777.47	38883	-105.53	19.502	19.555	-0.053
45 a 54	33022.25	29938	3084.25	16.608	15.057	1.551
55 ou +	37841.02	28187	9654.02	19.031	14.176	4.855

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	91322.41	92020	-697.59	45.929	46.280	-0.351
Feminio	107512.59	106815	697.59	54.071	53.720	0.351

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	90932.00	90932	0.00	45.732	45.732	-0.000
9 a 11	70078.00	70078	0.00	35.244	35.244	0.000
12 ou +	37825.00	37825	0.00	19.023	19.023	-0.000

Raking by FXID1, iteration - 2

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXID1						
18 a 24	39776.61	47207	-7430.39	20.005	23.742	-3.737
25 a 34	49417.65	54621	-5203.35	24.854	27.470	-2.617
35 a 44	38777.47	38883	-105.53	19.502	19.555	-0.053
45 a 54	33022.25	29938	3084.25	16.608	15.057	1.551
55 ou +	37841.02	28187	9654.02	19.031	14.176	4.855
	198835.00	198836		100.00	100.00	

Raking by Q7, iteration - 2

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
sexo						
Masculino	92562.66	92020	542.662	46.552	46.280	0.273
Feminio	106273.34	106815	-541.662	53.448	53.720	-0.273
	198836.00	198835		100.00	100.00	

Raking by FXESC2, iteration - 2

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXESC2						
0 a 8	83320.39	90932	-7611.61	41.904	45.732	-3.828
9 a 11	75002.39	70078	4924.39	37.721	35.244	2.477
12 ou +	40512.22	37825	2687.22	20.375	19.023	1.351
	198835.00	198835		100.00	100.00	

Weighted Distribution After Iteration 2

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	45099.18	47207	-2107.82	22.682	23.742	-1.060
25 a 34	53393.33	54621	-1227.67	26.853	27.470	-0.617
35 a 44	39270.63	38883	387.63	19.750	19.555	0.195
45 a 54	30935.02	29938	997.02	15.558	15.057	0.502
55 ou +	30136.83	28187	1949.83	15.157	14.176	0.981

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	92029.82	92020	9.82	46.285	46.280	0.005
Feminio	106805.18	106815	-9.82	53.715	53.720	-0.005

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	90932.00	90932	0.00	45.732	45.732	-0.000
9 a 11	70078.00	70078	0.00	35.244	35.244	0.000
12 ou +	37825.00	37825	-0.00	19.023	19.023	-0.000

Raking by FXID1, iteration - 3

FXID1	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
18 a 24	45099.18	47207	-2107.82	22.682	23.742	-1.060
25 a 34	53393.33	54621	-1227.67	26.853	27.470	-0.617
35 a 44	39270.63	38883	387.63	19.750	19.555	0.195
45 a 54	30935.02	29938	997.02	15.558	15.057	0.502
55 ou +	30136.83	28187	1949.83	15.157	14.176	0.981
	198835.00	198836		100.00	100.00	

Raking by Q7, iteration - 3

sexo	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
Masculino	92310.74	92020	290.743	46.426	46.280	0.146
Feminio	106525.26	106815	-289.743	53.574	53.720	-0.146
	198836.00	198835		100.00	100.00	

Raking by FXESC2, iteration - 3

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXESC2						
0 a 8	89024.84	90932	-1907.16	44.773	45.732	-0.959
9 a 11	71282.04	70078	1204.04	35.850	35.244	0.606
12 ou +	38528.12	37825	703.12	19.377	19.023	0.354
	198835.00	198835		100.00	100.00	

Weighted Distribution After Iteration 3

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxid1						
18 a 24	46648.34	47207	-558.66	23.461	23.742	-0.281
25 a 34	54308.39	54621	-312.61	27.313	27.470	-0.157
35 a 44	39036.74	38883	153.74	19.633	19.555	0.077
45 a 54	30185.87	29938	247.87	15.181	15.057	0.125
55 ou +	28655.67	28187	468.67	14.412	14.176	0.236

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	92025.65	92020	5.65	46.282	46.280	0.003
Feminio	106809.35	106815	-5.65	53.718	53.720	-0.003

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	90932.00	90932	-0.00	45.732	45.732	0.000
9 a 11	70078.00	70078	-0.00	35.244	35.244	-0.000
12 ou +	37825.00	37825	-0.00	19.023	19.023	-0.000

Raking by FXID1, iteration - 4

	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
FXID1						
18 a 24	46648.34	47207	-558.665	23.461	23.742	-0.281
25 a 34	54308.39	54621	-312.611	27.313	27.470	-0.157
35 a 44	39036.74	38883	153.740	19.633	19.555	0.077
45 a 54	30185.87	29938	247.866	15.181	15.057	0.125
55 ou +	28655.67	28187	468.670	14.412	14.176	0.236
	198835.00	198836		100.00	100.00	

Raking by Q7, iteration - 4

sexo	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
Masculino	92093.30	92020	73.3036	46.316	46.280	0.037
Feminio	106742.70	106815	-72.3036	53.684	53.720	-0.037
	198836.00	198835		100.00	100.00	

Raking by FXESC2, iteration - 4

FXESC2	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
0 a 8	90441.64	90932	-490.363	45.486	45.732	-0.247
9 a 11	70388.52	70078	310.518	35.400	35.244	0.156
12 ou +	38004.85	37825	179.846	19.114	19.023	0.090
	198835.00	198835		100.00	100.00	

Weighted Distribution After Iteration 4

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxidl						
18 a 24	47061.07	47207	-145.93	23.668	23.742	-0.073
25 a 34	54539.95	54621	-81.05	27.430	27.470	-0.041
35 a 44	38924.20	38883	41.20	19.576	19.555	0.021
45 a 54	30003.31	29938	65.31	15.090	15.057	0.033
55 ou +	28306.47	28187	119.47	14.236	14.176	0.060

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	92022.10	92020	2.10	46.281	46.280	0.001
Feminio	106812.90	106815	-2.10	53.719	53.720	-0.001

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	90932.00	90932	-0.00	45.732	45.732	0.000
9 a 11	70078.00	70078	-0.00	35.244	35.244	-0.000
12 ou +	37825.00	37825	-0.00	19.023	19.023	0.000

Raking by FXID1, iteration - 5

FXID1	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
18 a 24	47061.07	47207	-145.929	23.668	23.742	-0.073
25 a 34	54539.95	54621	-81.054	27.430	27.470	-0.041
35 a 44	38924.20	38883	41.205	19.576	19.555	0.021
45 a 54	30003.31	29938	65.306	15.090	15.057	0.033
55 ou +	28306.47	28187	119.473	14.236	14.176	0.060
	198835.00	198836		100.00	100.00	

Raking by Q7, iteration - 5

sexo	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
Masculino	92041.40	92020	21.3954	46.290	46.280	0.011
Feminio	106794.60	106815	-20.3954	53.710	53.720	-0.011
	198836.00	198835		100.00	100.00	

Raking by FXESC2, iteration - 5

FXESC2	Current Sum of Weights	Target Total	Sum of Weights Difference	Current % of Weights	Target % of Weights	Difference in %
0 a 8	90802.64	90932	-129.358	45.667	45.732	-0.065
9 a 11	70160.87	70078	82.867	35.286	35.244	0.042
12 ou +	37871.49	37825	46.491	19.047	19.023	0.023
	198835.00	198835		100.00	100.00	

Weighted Distribution After Iteration 5

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxidl						
18 a 24	47167.78	47207	-39.22	23.722	23.742	-0.020
25 a 34	54599.17	54621	-21.83	27.460	27.470	-0.011
35 a 44	38893.57	38883	10.57	19.561	19.555	0.005
45 a 54	29956.07	29938	18.07	15.066	15.057	0.009
55 ou +	28218.40	28187	31.40	14.192	14.176	0.016

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
sexo						
Masculino	92020.99	92020	0.99	46.280	46.280	0.000
Feminio	106814.01	106815	-0.99	53.720	53.720	-0.000

	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
fxesc2						
0 a 8	90932.00	90932	-0.00	45.732	45.732	-0.000
9 a 11	70078.00	70078	-0.00	35.244	35.244	0.000
12 ou +	37825.00	37825	0.00	19.023	19.023	0.000

**** Program terminated at iteration 5 because all current percents differ from target percents by less than 0.025 ******

Weighted Distribution After Raking

fxid1	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
18 a 24	47167.78	47207	-39.22	23.722	23.742	-0.020
25 a 34	54599.17	54621	-21.83	27.460	27.470	-0.011
35 a 44	38893.57	38883	10.57	19.561	19.555	0.005
45 a 54	29956.07	29938	18.07	15.066	15.057	0.009
55 ou +	28218.40	28187	31.40	14.192	14.176	0.016

sexo	Output Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Output Weights	Target % of Weights	Difference in %
Masculino	92020.99	92020	0.99	46.280	46.280	0.000
Feminio	106814.01	106815	-0.99	53.720	53.720	-0.000

Iteration Number	Maximum Absolute Value of Difference in Sum of Weights	Maximum Absolute Value of Difference in %	Coefficient of Variation of Weights at the Completion of the Iteration
1	9654.02	4.8554	0.60420
2	2107.82	1.0600	0.67562
3	558.66	0.2808	0.69830
4	145.93	0.0733	0.70438
5	39.22	0.0196	0.70595

Number of Respondents Who Had Their Weights Decreased by the Trimming: **0**.

Number of Respondents Who Had Their Weights Increased by the Trimming: **667**.

Raking output weight: **rake_trim**

Weight	Mean	Min	Max	CV
ADULFONE_ATPT	99.27	19.17	383.50	0.463
rake_trim	99.27	49.63	478.35	0.483