

Pesquisa Nacional de Demografia e Saúde
da Criança e da Mulher
– PNDS 2006 –

ANEXO 11
ERROS NÃO AMOSTRAIS

CEBRAP – Centro Brasileiro de Análise e Planejamento

DECIT – Departamento de Ciência e Tecnologia / MS

Junho 2008

Anexo 11

Erros não Amostrais – PNDS 2006

Suzana Cavenaghi

Declaração de Idade e Sexo

A informação sobre idade na PNDS é essencial para marcar a ocorrência dos eventos ligados à reprodução. Na PNDS de 2006 utilizou-se pela primeira vez três perguntas para determinar a idade das pessoas residentes no domicílio: a data de nascimento (mês e ano), a idade em anos completos e, na falta destas, a idade presumida. Com isto, foi possível determinar a idade para todas as pessoas residentes nos domicílios contactados e, portanto, a idade em anos completos para 100% das mulheres de 15 a 49 anos, sem precisar de imputação.

A qualidade da declaração desta informação é reconhecidamente melhor quando se pergunta pela data de nascimento, ao invés dos anos completos, para se evitar o efeito de arredondamento, amplamente conhecido. O mês de nascimento não foi preenchido em somente 2,6% das 56.365 pessoas residentes nos domicílios abordados. O ano de nascimento não foi preenchido em 2,7% dos casos. Apesar de baixa porcentagem de informações ignoradas para estas perguntas, parte desta falta de informação se deu nos domicílios onde a entrevista não foi realizada por completo no domicílio (para domicílio sem mulheres elegíveis, o questionário de domicílio foi preenchido somente até a pergunta sobre a educação do responsável pelo domicílio). Nos domicílios com seu questionário completo, se obteve o mês de nascimento e ano de nascimento em respectivamente 1,9% e 2,0%. Para as mulheres em idade reprodutiva, o mês e o ano de nascimento foi ignorado para somente 0,1% das 15.765 mulheres entrevistadas (13 mulheres). Este resultado foi decorrência da insistência no treinamento sobre a importância destas variáveis e o trabalho de supervisão de campo, que colocou muita atenção nestes campos de idade.

A informação sobre sexo foi obtida para todas as pessoas residentes nos domicílios abordados. Somente para as crianças que morreram muito jovens ou nasceram mortas esta é uma variável mais difícil de ser coletada, mas para as 27.477 crianças declaradas pelas mães, no módulo de história de nascimentos, no questionário de

mulheres, somente para duas das crianças não foi possível obter a informação sobre o sexo e para aquelas que se conhece o sexo, 51,6% eram meninos.

No entanto, apesar de boa, a qualidade da declaração de idade e sexo não acompanhou as expectativas iniciais. Antes de se passar para esta análise, deve-se mencionar primeiro que houve uma tendência maior à seleção de domicílios com mulheres de 15 a 49 anos. Isto pode ser explicado pelo fato de que após três visitas sem sucesso, a entrevistadora deveria substituir o domicílio pelo próximo selecionado. Como a probabilidade de se encontrar pessoas no momento da visita é maior nos domicílios com mulheres, houve uma leve tendência a realizar entrevistas em domicílios com mulheres de 15 a 49 residentes. Este viés foi resolvido com a expansão dos dados amostrais utilizando os dados da Pesquisa Nacional por Amostra de Domicílios, conforme explicado no capítulo Metodológico deste relatório. Desta forma, para a análise da qualidade da idade, utiliza-se os dados ponderados e expandidos.

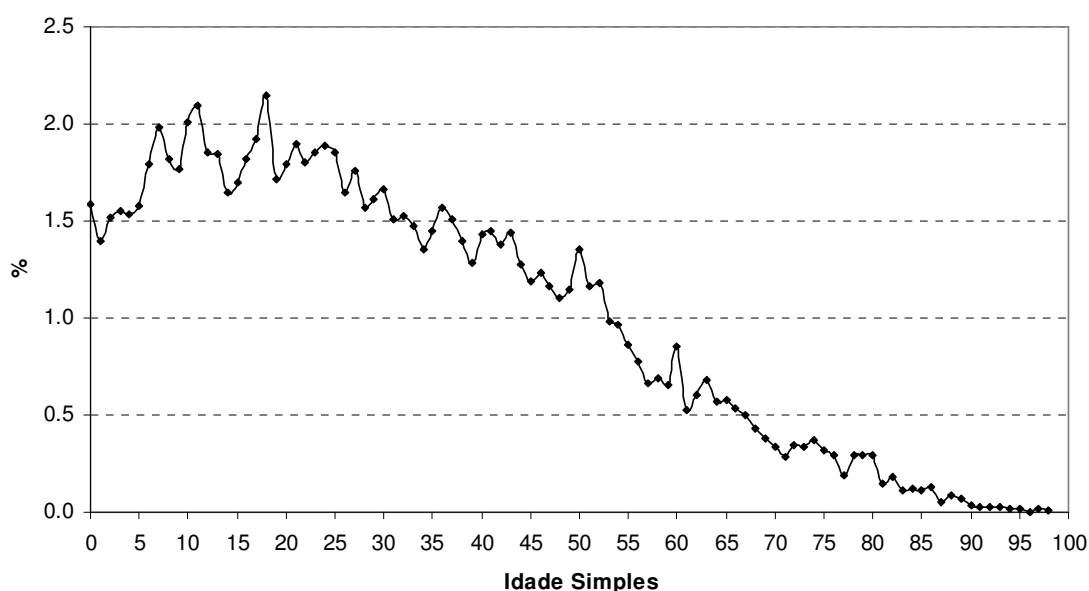
Qualidade da declaração de idade e sexo

Como se pode verificar no gráfico 11.1 existem picos de frequência em algumas idades, que apesar de esperadas, pela preferência de dígitos, aparentemente deveriam ter sido minimizadas pela informação sobre data de nascimento. Durante a etapa de consistência dos dados se verificou que algumas supervisoras e até as coordenadoras de equipes faziam consistência manual sobre a variável data de nascimento e idade em anos completos e, inadvertidamente corrigiram no próprio questionário a informação sobre o ano de nascimento para ficar compatível com a idade declarada em anos completos. Em pesquisas futuras esta informação deve ser melhor trabalhada no treinamento para que a informação não seja alterada em campo.

Outro fato intrigante mostrado neste gráfico é a sub-enumeração de pessoas de 15 anos e sobre-enumeração de pessoas de 50 anos de idade. Na PNDS de 1996 se relatou fato, explicando que *“Uma falha comum com relação a este fato refere-se à tendência de alguns entrevistadores em excluir mulheres cujas idades estão próximas aos limites fixados para a condição de elegibilidade, a fim de diminuir a carga do trabalho de campo. Como, no presente caso, as mulheres elegíveis*

situaram-se entre as idades de 15 a 49 anos, poderia esperar-se uma tendência de se excluir jovens ligeiramente acima de 15 anos e mulheres ligeiramente abaixo de 50 anos” (BEMFAM, 1997, p. 172)¹. As entrevistadoras têm que aplicar o questionário completo para todas as mulheres de 15 a 49 anos no domicílio, e completar 12 domicílios por setor, portanto, é bastante difícil controlar estes efeitos, a não ser com um acompanhamento em tempo real da distribuição por idade e sexo da população entrevistada.

Gráfico 11.1: Distribuição etária da população residente nos domicílios particulares permanentes, por idade simples . Brasil, 2006.

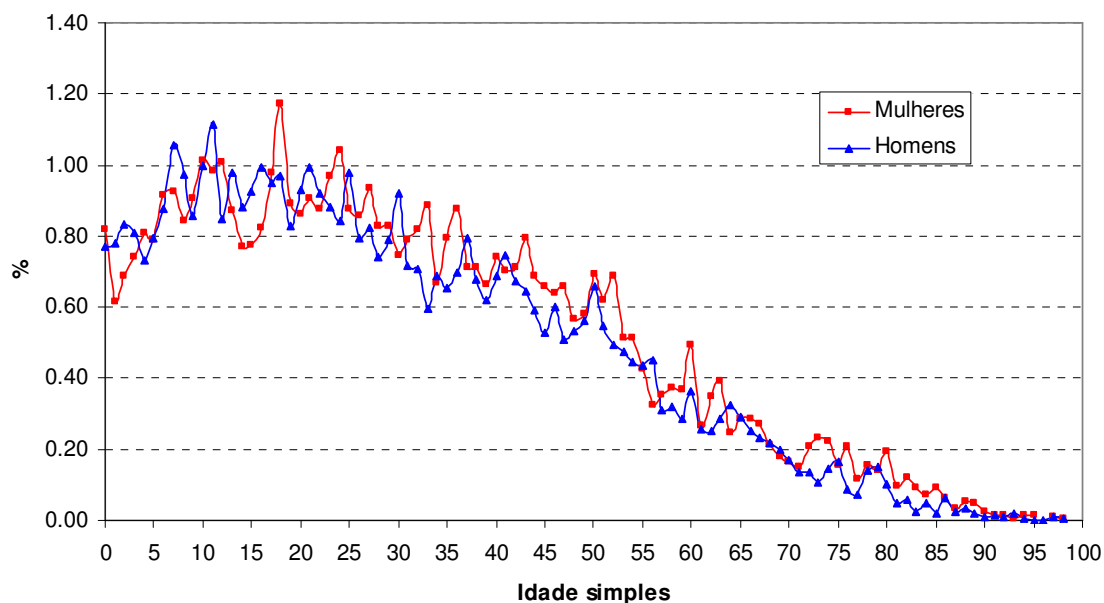


Fonte: PNDS, 2006

O Gráfico 11.2, que apresenta a mesma distribuição do gráfico anterior por sexo, mostra claramente que este efeito aconteceu mais para as mulheres do que para os homens, confirmando este viés trazido devido ao grupo de elegibilidade para o questionário completo. Apesar dos picos em algumas idades, para os homens a série segue uma tendência decrescente, enquanto para as mulheres, a tendência é interrompida para a idade de 14 e 15 anos e novamente para o grupo de de 50 e 51 anos de idade.

¹ BEMFAM, 1997. “Brasil: Pesquisa Nacional sobre Demografia e Saúde.” Macro International, Inc., Maryland.

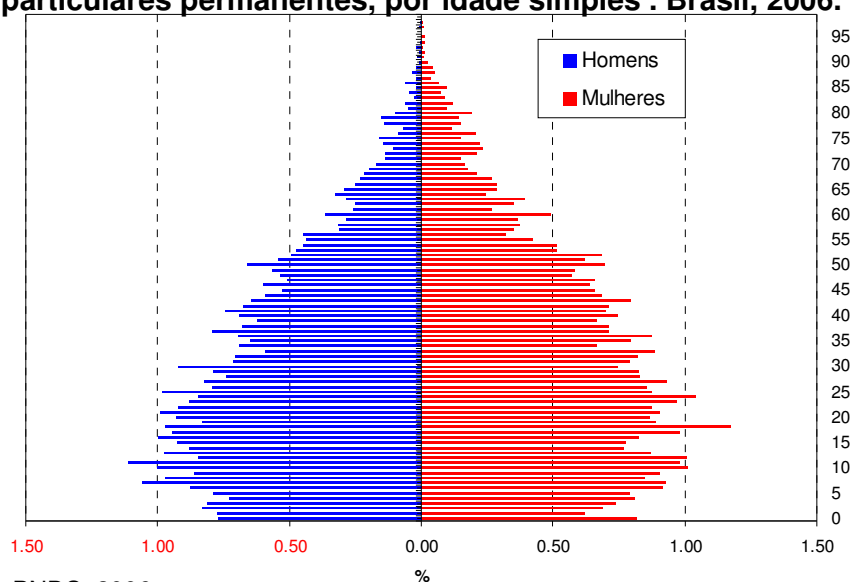
Gráfico 11.2: Distribuição etária da população residente nos domicílios particulares permanentes, por idade simples, segundo sexo . Brasil, 2006.



Fonte: PNDS, 2006

Igualmente, a partir da Pirâmide etária por idade simples, disponível no Gráfico 11.3, podemos observar que a atração de dígitos na declaração da idade ocorre tanto para homens quanto para mulheres. Este problema pode ser resolvido com análises por grupos quinquenais, o que é feito para a maioria dos eventos ocorridos no período reprodutivo feminino, de 15 a 49 anos de idade.

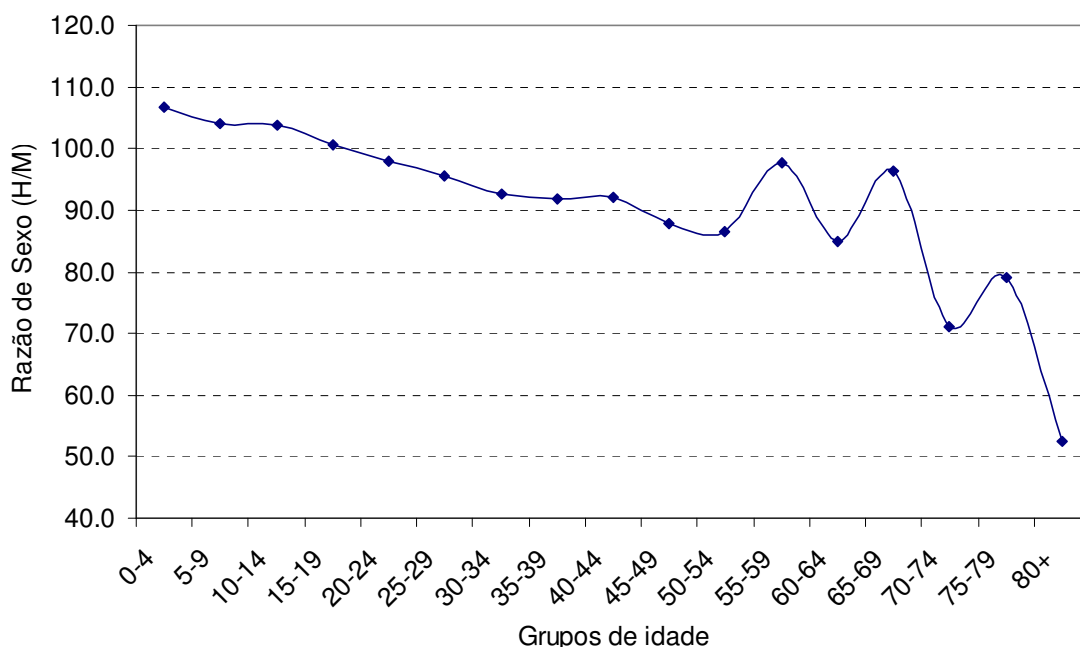
Gráfico 11.3: Distribuição etária por sexo da população residente nos domicílios particulares permanentes, por idade simples . Brasil, 2006.



Fonte: PNDS, 2006.

A razão de sexo, apresentada no gráfico 11.4, por grupos quinquenais de idade apresenta irregularidades para as idades a partir de 55 anos de idade, mas a tendência geral é esperada. Até 15 anos de idade existem mais homens do que mulheres na população e a partir daí o número de mulheres é crescentemente maior do que o número de homens. Nas idade mais avançada isto é esperado devido ao diferencial na mortalidade, favorecendo as mulheres que sobrevivem mais do que os homens. Os dados são coerentes com os dados observados para os censos demográficos e, conseqüentemente, com os dados da PNAD. A variação observada para as idades mais avançadas é devido ao tamanho da amostra, que para representar o Brasil é bastante reduzida e, de qualquer forma, os dados de interesse do estudo, que são as mulheres de 15 a 49 anos de idade são bastante consistentes.

Gráfico 11.4: Razão de sexo da população residente nos domicílios particulares permanentes, por grupos quinquenais de idade. Brasil, 2006.



Fonte: PNDS, 2006.

Como nem todas as mulheres residentes no domicílio aceitaram fazer a entrevista, é importante analisar se existem vieses na recusa à entrevista. A distribuição etária das mulheres em idade reprodutiva em grupos quinquenais pode ser observada na Tabela 11.1, com dados não ponderados no painel superior e dados ponderados no painel inferior. Apresenta-se a distribuição de número de mulheres residentes nos

domicílios e aquelas que foram entrevistadas, por grupo etário. A recusa de mulheres elegíveis para a média do Brasil foi de 10,8%, portanto, a proporção de mulheres elegíveis entrevistadas é considerada bastante boa (89,2%) considerando a complexidade da pesquisa, que além do questionário muito grande e antropometria, como nas pesquisas anteriores, teve coleta de sangue em crianças e mulheres. Adicionalmente, nos últimos 10 anos se esperaria, principalmente com o aumento da violência, que a recusa fosse maior. Entretanto, nas pesquisas anteriores as taxas de resposta ficaram pouco abaixo do observado em 2006. Conforme relatado (BEMFAM, 1997, p. 174), “...a proporção de mulheres entrevistadas fica acima de 85% na maioria dos casos, principalmente depois da idade de 25 anos. Similar proporção foi encontrada na PNSMIPF de 1986. Nesse ano, 87% das mulheres da amostra dos domicílios responderam e/ou foram consideradas elegíveis. Em 1991, a proporção é de 85%.”

Tabela 11.1: Distribuição percentual de mulheres entrevistadas de 15-49 anos e porcentagem de mulheres elegíveis que foram entrevistadas (não ponderadas e ponderadas). Brasil, 2006.

Grupos etários	Sem ponderação – valores observados				Porcentagem de elegíveis entrevistadas
	Mulheres elegíveis		Mulheres entrevistadas		
	Número	Porcentagem	Número	Porcentagem	
15-19	3129	17.9	2488	16.0	79.5
20-24	2910	16.7	2508	16.1	86.2
25-29	2665	15.3	2435	15.6	91.4
30-34	2419	13.9	2301	14.8	95.1
35-39	2235	12.8	2099	13.5	93.9
40-44	2161	12.4	1975	12.7	91.4
45-49	1946	11.1	1769	11.4	90.9
Total	17465	100.0	15575	100.0	89.2
Grupos etários	Dados Ponderados				Porcentagem de elegíveis entrevistadas
	Mulheres elegíveis		Mulheres entrevistadas		
	Número	Porcentagem	Número	Porcentagem	
15-19	8694000	16.6	7815218	15.5	89.9
20-24	8725810	16.6	8323243	16.5	95.4
25-29	8092000	15.4	7894867	15.7	97.6
30-34	7312000	13.9	7185063	14.3	98.3
35-39	7035000	13.4	6865483	13.6	97.6
40-44	6805000	13.0	6585015	13.1	96.8
45-49	5817000	11.1	5664993	11.3	97.4
Total	52480810	100.0	50333882	100.0	95.9

Fonte: PNDS, 2006.

A porcentagem de mulheres entrevistadas por grupo etário, no entanto, mostra que houve uma recusa maior das mulheres mais jovens, O grupo de 15 a 19 anos de

idade apresenta uma média de recusa de quase 20%. Segundo os dados de campo esta recusa se deu com maior frequência em domicílios onde uma mulher mais velha já tinha sido entrevistada e as jovens, que estudam ou trabalham, apresentaram uma resistência maior a responder ao questionário. A expansão da amostra, que considera uma reponderação pela recusa dentro do domicílio e uma calibração por idade e sexo, diminui o viés introduzido por grupo de idade.

Grau de omissão das respostas em variáveis selecionadas

O questionário completo da PNDS é muito extenso e existem inúmeras variáveis. Por este motivo não cabe aqui fazer uma análise completa do grau de omissão de todas as respostas. Sabe-se que alguns dados são mais difíceis de serem obtidos dos entrevistados, mas a maioria das variáveis devem apresentar em média graus de omissão baixo se o cuidado no treinamento e na supervisão de campo foram dadas a todos os módulos do questionário. Por este motivo, seleciona-se aqui algumas variáveis que são de importância para análises transversais e outras que acredita-se de antemão que teriam maior grau de omissão dada a natureza da pergunta. A Tabela 11.2 apresenta a porcentagem de casos onde não se obteve uma resposta válida, por recusa em responder a pergunta, por não saber, ou por ter apresentado resposta inconsistente com outras informações do questionário. As variáveis selecionadas para esta análise foram: data de nascimento (mês e ano), idade ao morrer, idade à primeira união, esterilização, educação dos entrevistados, estatura da criança ao nascer, antropometria e ocorrência de diarreia em crianças, dentre outras variáveis.

Como pode ser observado na Tabela 11.2, a maioria das variáveis apresentadas têm baixa porcentagem de informações ignoradas. A maioria delas fica abaixo dos 2%. Algumas excessões são visíveis, principalmente no que se refere às informações obtidas a partir da coleta de sangue. Para as mulheres houve uma omissão de ao redor de 10%, que inclui recusa, falta de informação por inconsistência ou amostra de sangue coletado insuficiente para análise. Para as crianças a falta de dados vindos da coleta de sangue é bem maior, ao redor de 20%. Além da inadequação da amostra para análise, a maior parte foi devido à recusa, que muitas vezes foi de uma segunda ou terceira criança abordada no domicílio. Com relação aos dados de peso

e altura, novamente a omissão para as mulheres é bem mais baixa do que para as crianças, mas mesmo assim, o falta de informação de peso e altura para as crianças é menor do que 8%.

Tabela 11.2: Porcentagem de informações ignoradas observadas para variáveis e grupos populacionais selecionados. Brasil, 2006.

Variáveis e grupos populacionais selecionados	Número pessoas na categoria	Pessoas com Informação ignorada	Porcentagem de omissão
Data de nascimento toda população no questionário de domicílio			
mês	56.365	1.465	2,6
ano	56.355	1.503	2,7
Data de nascimento dos filhos <=15 anos na história nascimento			
mês	18.401	195	1,1
ano	18.401	209	1,1
Idade ao morrer para crianças <=15 anos	1.384	84	6,1
Anos de estudo da população	56.365	748	1,3
Anos de estudo das mulheres de 15-49	15.575	121	0,8
Data da esterilização			
mês	4.098	386	9,4
ano	4.098	203	5,0
mês e ano	4.098	160	3,9
Idade do cônjuge/companheiro atual	9.989	14	0,1
Idade/data primeira união	11.754	60	0,5
Peso da mulher	15.575	168	1,1
Altura da mulher	15.575	184	1,2
Cintura da mulher	15.575	865	5,6
Retinol em mulheres	6.673	729	10,9
Hemoglobina em mulheres	6.673	758	11,4
Retinol em crianças	4.817	1.006	20,9
Hemoglobina crianças	4.817	1.072	22,3
Diarréia em crianças < 59 meses nas últimas duas semanas	4.817	96	2,0
Diarréia em crianças < 59 meses nos últimos 3 meses	4.817	91	1,9
Crianças pesada ao nascer	4.957	15	0,3
Peso declarado	4.864	3.393	69,8
Peso no cartão	4.864	1.373	28,2
Peso das crianças	4.817	304	6,3
Altura das crianças	4.817	383	8,0
Peso e altura das crianças	4.817	290	6,0

Fonte: PNDS, 2006.

A informação sobre idade ao morrer para as crianças com 15 anos ou menos ficou ao redor de 6%. Apesar de ser somente 84 casos de omissão, como a mortalidade já é um evento muito raro, o ideal seria conseguir menor índice de omissão nesta variável, no entanto, dependendo da situação ou do tempo da ocorrência do evento é muito difícil conseguir estas informações com a mãe.

A data da esterilização é outra variável importante para a análise, dada a alta porcentagem de mulheres esterilizadas e por se saber que a esterilização se dá em idade muito precoces, ter boas informações sobre esta variável é importante. Verifica-se que em se lembrar do mês da ocorrência da cirurgia da esterilização é mais difícil. O ano no entanto ficou com uma porcentagem de omissão baixa, ao redor de 5%, ainda mais quando estamos falando de eventos que a mulher deve se recordar de um passado há mais de 20 anos.

Uma informação apresentada na Tabela 11.2 que mostra inadequação completa pela alta taxa de omissão é sobre o peso ao nascer. Apesar de se somente 0,3% ignorar se a criança tinha sido pesada ao nascer e, poucas responderem que não foi pesado (1,6%), a grande maioria não soube dizer o peso. A partir dos dados anotados no cartão da criança o nível de omissão diminui, mas ainda assim, chega ao redor de 30% de omissão. Desta forma, a análise de peso ao nascer a partir destas variáveis não é adequado.

Considerações Finais

A breve análise apresentada mostra que os dados coletados na PNDS 2006 são bastante consistentes, apesar da complexidade do questionário e do tamanho da amostra, reduzido para muitos eventos raros importantes investigados na pesquisa. O controle e supervisão de campo foi bastante rigoroso, mostrando resultados consistentes, mas como era de se esperar em uma pesquisa desta magnitude, apresenta alguns vieses pequenos e que ainda puderam ser minimizados a partir de técnicas de estatística.

A qualidade das informações mais relevantes podem ser atestadas pela análise de omissão das informações apresentadas aqui e, também, pela análise de tendência

com as pesquisas anteriores e comparações com valores esperados de outras pesquisas. As análises de consistência realizadas mostram que algumas variáveis ainda podem apresentar aparentes incoerências, mas isto é inevitável em uma base de dados tão complexa, mas estas são mínimas e não colocam em risco a qualidade das informações como um todo.

Finalmente deve-se ressaltar que os dados devem ser utilizados com a ponderação informada na base de dados. Esta, como já mencionado, foi calculada a partir das estimativas da PNAD 2006, a qual por sua vez é baseada nas estimativas populacionais oficiais divulgadas pelo IBGE. Caso haja alguma revisão oficial destas estimativas, será necessário fazer estudos analíticos sobre a influência destas mudanças nas estimativas obtidas com a PNDS 2006. Chama-se atenção para que o usuário em suas análises sempre busque estimativas ponderadas, mas que não se esqueça de verificar o tamanho da amostra a que se refere suas estimativas, ou seja, o número de entrevistas efetivamente realizadas, para não fazer inferências que não sejam estatisticamente significantes devido ao tamanho reduzido da amostra.