

Pesquisa sobre as Condições de Saúde Bucal da População Brasileira (SB Brasil 2003): seus dados não produzem estimativas populacionais, mas há possibilidade de correção

Brazilian Oral Health Survey (SB Brazil 2003): data do not allow for population estimates, but correction is possible

*Rejane Christine de Sousa Queiroz¹
Margareth Crisóstomo Portela¹
Maurício Teixeira Leite de Vasconcellos²*

Abstract

The Brazilian Oral Health Survey (SB Brazil 2003) was the most comprehensive study on oral health conditions ever conducted in Brazil. Probabilistic sampling methods were applied in order for the collected data to represent the population age groups selected in the 5 regions of the country. However, this was not possible because the sampling process was never concluded, which would require estimation of the sample weights and identification of the sample's other structural variables (selection strata and primary sampling units). This paper describes the SB Brazil 2003 sample design, formulates the inclusion probabilities in the multiple selection stages, and proposes strategies for estimating the sample weights. The strategy to define the sample weights and identify the sample's structural variables hinges on retrieving data that should have been recorded in the reports produced during the survey, and in their absence, in information available from the Brazilian Institute of Geography and Statistics (IBGE) and Ministry of Education and Culture (MEC), as valid proxies.

Oral Health; Sampling Studies; Health Status; Population

Introdução

O processo de inserção da área de saúde bucal na agenda da saúde pública brasileira trouxe a necessidade de ampliar o conhecimento das condições de saúde bucal da população brasileira e de quantificar os eventos de interesse para apoiar o planejamento de ações na área. A partir da década de 80 do século passado, a área começou a acumular experiência em grandes inquéritos epidemiológicos que incluíam exames clínicos bucais dos entrevistados, especialmente escolares.

O primeiro inquérito sobre a saúde bucal no Brasil foi realizado em 1986¹ pela Divisão Nacional de Saúde Bucal (DNSB) do Ministério da Saúde, com o apoio do Instituto de Planejamento Econômico e Social da Secretaria de Planejamento da Presidência da República. As limitações orçamentárias foram determinantes para definir a abrangência geográfica, a população alvo e a abrangência do exame bucal. Realizou-se um inquérito na área urbana de 16 capitais, distribuídas entre as cinco macrorregiões do país, incluindo crianças (6 a 12 anos), adolescentes (15 a 19 anos) e adultos (35 a 44 anos e 50 a 59 anos). A justificativa para não considerar idosos entre 65 e 74, como proposto pela Organização Mundial da Saúde (OMS) e adotado em países desenvolvidos (65 a 74 anos), foi baseada no fato de a esperança de vida ao nascer ser, à época, de 65,5 anos no Brasil². Ocorrência de cárie e doença periodon-

¹ Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

² Escola Nacional de Ciências Estatísticas, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brasil.

Correspondência

M. T. L. Vasconcellos
Escola Nacional de Ciências Estatísticas, Instituto Brasileiro de Geografia e Estatística.
Rua André Cavalcanti 106,
Rio de Janeiro, RJ
20231-050, Brasil.
mauricio.vasconcellos@ibge.gov.br

tal, necessidade e uso de prótese dentária foram observados, além de informações sobre acesso aos serviços de saúde.

Nesse inquérito, a idade de 12 anos foi utilizada para fins de comparação internacional, obedecendo a critérios propostos pela Federação Dentária Internacional (FDI) e pela OMS, que também lançaram metas com índices a serem alcançados por volta do ano de 2000³.

O relatório desse inquérito aponta várias dificuldades enfrentadas para a sua efetivação, incluindo questões operacionais (dimensão e a diversidade regional do país) e a necessidade de obter dados mais acurados para o planejamento governamental, diante da magnitude da necessidade de tratamento odontológico² pela população.

Em relação ao procedimento de amostragem, o relatório informa que foram utilizadas duas amostras, uma para a faixa etária entre 6 e 12 anos e outra para as demais faixas etárias. As amostras foram selecionadas em dois estágios. Para estudantes (6 a 12 anos) foram selecionados estabelecimentos de ensino (1ª e 2ª graus) e, neles, os estudantes que foram entrevistados e examinados. Para as demais faixas etárias, foram selecionados setores censitários; neles, os domicílios onde foram entrevistados e examinados todos os moradores das demais faixas etárias contempladas.

Embora restrito à área urbana de algumas capitais, este primeiro inquérito representou um passo importante, na medida em que somou conhecimentos específicos das áreas de Odontologia e Saúde Pública aos métodos e técnicas do Instituto de Pesquisa Econômica Aplicada (IPEA) e do Instituto Brasileiro de Geografia e Estatística (IBGE).

Dez anos depois, em 1996, com o Sistema Único de Saúde (SUS) já implantado, foi realizado o segundo inquérito sob coordenação da Área Técnica de Saúde Bucal do Ministério da Saúde. O seu objetivo foi verificar as alterações ocorridas no perfil de saúde bucal da população brasileira, tomando como referência o inquérito de 1986, e obter um referencial para as ações preventivas do SUS⁴. Apesar de ampliar a abrangência geográfica (27 capitais das Unidades da Federação), o inquérito limitou-se ao exame de cárie na faixa etária de 6 a 12 anos, com crianças selecionadas em escolas públicas e privadas⁵. Ainda que houvesse dados disponíveis no Departamento de Informática do SUS (DATASUS; <http://tabnet.datasus.gov.br/cgi/sbucal/sbdescr.htm>, acessado em 12/Dez/2007), a pesquisa sequer gerou um relatório final.

Embora obtivessem dados em diferentes regiões do país, os dois inquéritos não foram sufi-

cientes para a construção de um quadro abrangente sobre as condições de saúde bucal da população brasileira. Eles representavam as características dos subgrupos populacionais definidos pelas faixas etárias consideradas e residentes nas cidades pesquisadas; não eram representativos da população das regiões e, muito menos, da população brasileira.

O reconhecimento dessas limitações conduziu o Ministério da Saúde a conceber um projeto para enfrentar o desafio de realizar um diagnóstico de saúde bucal do país, contemplando um número maior de eventos e de grupos etários. Nasce, então, o projeto SB Brasil 2003, com abrangência realmente nacional, incorporando, pela primeira vez, as faixas etárias de 18 a 36 meses e 65 a 74 anos, e a área rural. Apesar do título, esse terceiro inquérito foi realizado entre 2002 e 2003, tendo sido concebido para obter informações epidemiológicas capazes de subsidiar a elaboração das diretrizes para uma política nacional de saúde bucal, bem como ações para o fortalecimento da gestão dos serviços públicos de saúde bucal nas diferentes esferas de governo.

Comparado aos inquéritos anteriores, o SB Brasil 2003 incluiu ainda uma gama maior de dados relativos à situação sócio-econômica, auto-avaliação da saúde bucal e acesso aos serviços odontológicos. Além disso, foram realizados exames clínicos que, em princípio, permitiriam comparar a situação observada com as metas de saúde bucal preconizadas para o ano 2000 pela OMS⁶.

O planejamento deste inquérito considerou a seleção de uma amostra probabilística por conglomerados, representativa de faixas etárias da população das cinco macrorregiões. O processo de amostragem, entretanto, não foi concluído. Faltou o cálculo dos pesos amostrais necessários para a produção de estimativas válidas para a população brasileira.

Em 2004, houve a publicação dos seus principais resultados pelo Ministério da Saúde⁶. A disseminação do banco de dados associado à descrição de que a pesquisa "*permitiu a produção de inferências para cada uma das macrorregiões brasileiras, por porte do município e para cada idade ou grupo etário*"⁶ (p. 13) talvez tenha induzido o uso dos dados do SB Brasil 2003 por diversos pesquisadores, que publicaram artigos científicos em revistas nacionais e internacionais^{7,8,9,10}, sem perceberem que os resultados não eram inferências estatisticamente válidas para os grupos etários da população de pesquisa (brasileira ou por macrorregião) porque eram referidos e limitados ao grupo estudado (estatísticas da amostra).

Este artigo tem por objetivo demonstrar que o procedimento de amostragem do SB Brasil 2003 não foi concluído, alertando seus potenciais usuários de que estimativas produzidas não são representativas da população brasileira, além de apontar para as possibilidades de correção dos problemas identificados. O artigo não avança na correção dos problemas indicados, pois ela depende do resgate de informações constantes dos relatórios das equipes de campo, o que necessariamente passa pelo apoio do Ministério da Saúde e das coordenações estaduais e municipais do projeto SB Brasil 2003.

Por que as estimativas do SB Brasil 2003 não são válidas para a população?

Desde que Kiar¹¹ apresentou suas idéias sobre o método representativo, hoje conhecido como amostragem probabilística, pesos amostrais são utilizados para a produção de inferências sobre a população de pesquisa ou população amostrada, como definiu Cochran¹². Na amostragem probabilística cada unidade da amostra tem uma probabilidade de inclusão na amostra, representada por π_i , cujo inverso é o peso amostral (w_i), ou seja, $w_i = \pi_i^{-1}$.

Como a base de dados do SB Brasil 2003 não tem pesos amostrais, os resultados desta pesquisa são representativos do grupo de 108.921 pessoas examinadas, não podendo ser extrapolados para a população de pesquisa (população brasileira dos grupos etários estudados).

É fato que alguns desenhos de amostra são concebidos de forma que as probabilidades de seleção sejam constantes para todas as unidades da amostra ($\pi_i = \pi \forall i \in S$, onde S é a amostra). Esses planos, ditos autoponderados¹², têm como propriedade principal o fato de médias, razões, proporções e prevalências amostrais serem estimativas não enviesadas da população, o que não vale para totais. No entanto, este não é o caso da amostra do SB Brasil 2003, cujo desenho forçosamente conduz a pesos desiguais.

Na Tabela 1, são apresentados pesos relativos que, apesar de grosseiros, dão uma idéia de quantas pessoas da população cada pessoa entrevistada representaria se o SB Brasil 2003 tivesse sido realizado em 2000 com um esquema de seleção equiprovável em cada estrato da amostra. Os pesos relativos, calculados dividindo a população observada no *Censo Demográfico de 2000* (IBGE; <http://www.ibge.gov.br>) pelo tamanho da amostra em cada estrato, variam entre 17,8 (12 anos, Centro-Oeste em municípios com população entre 50.001 a 100.000 habitantes) e

16.137,5 (35 a 44 anos, em municípios de mais de 100.000 habitantes do Sudeste).

Para cada grupo etário, a variação dos pesos relativos é bem grande, conforme indica a Tabela 2. Mesmo para estimativas de um único grupo etário e uma única macrorregião, a variação dos pesos relativos não é desprezível, fato que indica que as estatísticas da amostra constituem-se em estimativas enviesadas da população correspondente, com maior ou menor viés em função da variabilidade observada para cada variável.

É certo que para variáveis com reduzida variação na amostra as estatísticas amostrais, embora enviesadas, não serão muito diferentes das estimativas obtidas de forma correta; esse é possivelmente o caso da variável *cárie dentária*, de elevada prevalência na maioria dos grupos estudados. Contudo, quanto maior for a variabilidade da variável estudada, maior será o viés da estatística amostral em relação à estimativa correspondente. Além deste problema, há que se considerar que a falta dos pesos amostrais inviabiliza a estimação de totais populacionais.

Infelizmente, os problemas da amostra do SB Brasil 2003 não se limitam à ausência dos pesos amostrais. Para estimar variâncias e demais estatísticas dela dependentes (erros padrão, intervalos de confiança, significância de parâmetros de modelos estatísticos, dentre outras) é necessário, também, conhecer as outras informações estruturais do desenho da amostra: o estrato de seleção e a unidade primária de amostragem. Isto deriva do fato de o plano de amostragem do SB Brasil 2003 ser complexo¹³, combinando estratificação, conglomeração e seleção com probabilidades desiguais. Ou seja, um plano amostral que não permite estimação baseada na inferência estatística clássica, visto não ser aplicável um de seus pressupostos básicos: o de que todas as unidades da amostra derivam de observações independentes e igualmente distribuídas.

Entretanto, partindo-se do pressuposto de que as instruções de coleta foram estritamente observadas por toda a equipe de campo, pode-se buscar informações exatas, previstas para constarem dos relatórios das equipes de campo, ou aproximações válidas derivadas de dados do IBGE e do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) do Ministério da Educação (MEC), para determinar os pesos amostrais.

Tabela 1

Tamanho da amostra, população em 2000 e peso relativo por faixas etárias, segundo as macrorregiões e estratos de porte populacional.

Macrorregiões e estratos de porte populacional (habitantes)	18-36 meses			Faixa etária 5 anos			12 anos		
	Amostra	População *	Peso relativo **	Amostra	População *	Peso relativo **	Amostra	População *	Peso relativo **
Total	12.117	6.474.246	534,3	26.641	3.445.580	129,3	34.550	3.524.814	102,0
Norte	3.305	651.066	197,0	4.678	335.414	71,7	6.208	314.984	50,7
Até 5.000	329	20.765	63,1	300	10.724	35,7	515	10.285	20,0
5.001-10.000	526	45.113	85,8	557	23.379	42,0	957	22.031	23,0
10.001-50.000	861	254.101	295,1	1.034	132.369	128,0	1.338	123.084	92,0
50.001- 100.000	840	96.657	115,1	1.371	49.412	36,0	1.683	47.104	28,0
100.000 ou +	749	234.430	313,0	1.416	119.530	84,4	1.715	112.480	65,6
Nordeste	2.557	1.986.462	776,9	4.580	1.073.304	234,3	7.322	1.136.394	155,2
Até 5.000	533	45.604	85,6	476	25.850	54,3	680	27.948	41,1
5.001-10.000	473	133.965	283,2	601	73.882	122,9	1.075	78.420	72,9
10.001-50.000	425	895.366	2.106,7	660	491.400	744,5	1.451	516.364	355,9
50.001- 100.000	634	268.025	422,8	1.443	144.236	100,0	2.114	151.974	71,9
100.000 ou +	492	643.502	1.307,9	1.400	337.936	241,4	2.002	361.688	180,7
Sudeste	2.342	2.510.795	1.072,1	7.291	1.314.709	180,3	8.052	1.357.615	168,6
Até 5.000	447	55.179	123,4	457	30.882	67,6	604	33.195	55,0
5.001-10.000	545	105.913	194,3	586	58.906	100,5	1.030	62.855	61,0
10.001-50.000	660	472.142	715,4	1.334	257.414	193,0	1.871	271.194	144,9
50.001- 100.000	255	253.497	994,1	2.445	135.976	55,6	2.216	142.885	64,5
100.000 ou +	435	1.624.064	3.733,5	2.469	831.531	336,8	2.331	847.486	363,6
Sul	2.725	873.180	320,4	6.042	480.144	79,5	7.119	478.579	67,2
Até 5.000	449	44.116	98,3	442	26.026	58,9	456	27.227	59,7
5.001-10.000	545	74.531	136,8	583	43.186	74,1	699	43.491	62,2
10.001-50.000	652	263.888	404,7	1.105	147.911	133,9	1.592	146.777	92,2
50.001- 100.000	629	131.156	208,5	1.805	71.343	39,5	1.995	70.386	35,3
100.000 ou +	450	359.489	798,9	2.107	191.678	91,0	2.377	190.698	80,2
Centro-Oeste	1.188	452.743	381,1	4.050	242.009	59,8	5.849	237.242	40,6
Até 5.000	145	18.978	130,9	287	10.622	37,0	479	10.737	22,4
5.001-10.000	110	33.137	301,2	231	18.788	81,3	675	18.683	27,7
10.001-50.000	286	127.530	445,9	463	70.559	152,4	1.337	70.068	52,4
50.001- 100.000	387	58.300	150,6	1.414	30.096	21,3	1.516	26.984	17,8
100.000 ou +	260	214.798	826,1	1.655	111.944	67,6	1.842	110.770	60,1

(continua)

Desenho da amostra do SB Brasil 2003 e estratégias de cálculo dos pesos amostrais

O SB Brasil 2003 utilizou uma amostra probabilística com estratificação e conglomeração. A população de referência é constituída pelas pessoas dos seguintes grupos etários: 18 a 36 meses; 5 anos; 12 anos; 15 a 19 anos; 35 a 44 anos; e 65 a 74 anos. A amostra foi estratificada por macrorregião e, em cada uma delas, por porte populacional dos municípios (até 5.000 habitantes; 5.001 a 10.000 habitantes; 10.001 a 50.000 habitantes;

50.001 a 100.000 habitantes; e mais de 100.000 habitantes). Em cada estrato de porte populacional foram, a princípio, selecionados 10 municípios e, em cada município foram escolhidos os seguintes conglomerados: setores censitários, quadras ou vilas, domicílios e estabelecimentos de ensino (escolas e creches).

No entanto, o manual do coordenador ¹⁴, ao descrever a seleção das unidades primárias de amostragem (UPA), indica que as capitais de todas as Unidades da Federação foram incluídas na amostra, não fazendo parte do sorteio de municípios. Portanto, no estrato de maior por-

Tabela 1 (continuação)

Tamanho da amostra, população em 2000 e peso relativo por faixas etárias, segundo as macrorregiões e estratos de porte populacional.

Macrorregiões e estratos de porte populacional (habitantes)	15-19 anos			Faixa etária 35-44 anos			65-74 anos		
	Amostra	População *	Peso relativo **	Amostra	População *	Peso relativo **	Amostra	População *	Peso relativo **
Total	16.833	17.939.815	1.065,8	13.431	22.808.223	1.698,2	5.349	6.323.408	1.182,2
Norte	3.877	1.524.484	393,2	2.486	1.420.318	571,3	746	303.763	407,2
Até 5.000	680	45.175	66,4	458	41.217	90,0	122	10.916	89,5
5.001-10.000	727	97.213	133,7	430	83.794	194,9	160	21.855	136,6
10.001-50.000	852	553.086	649,2	518	466.225	900,0	150	111.822	745,5
50.001- 100.000	838	220.581	263,2	575	191.049	332,3	166	42.286	254,7
100.000 ou +	780	608.429	780,0	505	638.033	1.263,4	148	116.884	789,8
Nordeste	3.998	5.571.708	1.393,6	3.370	5.513.073	1.635,9	1.446	1.667.384	1.153,1
Até 5.000	711	128.633	180,9	548	114.200	208,4	249	43.366	174,2
5.001-10.000	776	359.952	463,9	551	310.459	563,4	253	117.759	465,5
10.001-50.000	749	2.416.235	3.225,9	618	2.095.534	3.390,8	259	763.970	2.949,7
50.001- 100.000	969	742.010	765,7	881	722.961	820,6	416	217.288	522,3
100.000 ou +	793	1.924.878	2.427,3	772	2.269.919	2.940,3	269	525.001	1.951,7
Sudeste	2.981	7.155.091	2.400,2	2.340	10.580.065	4.521,4	1.052	2.994.665	2.846,6
Até 5.000	677	162.113	239,5	487	209.345	429,9	205	78.096	381,0
5.001-10.000	683	307.811	450,7	513	388.064	756,5	238	138.388	581,5
10.001-50.000	679	1.359.052	2.001,5	534	1.821.075	3.410,3	286	574.498	2.008,7
50.001- 100.000	391	736.638	1.884,0	366	1.061.078	2.899,1	158	288.090	1.823,4
100.000 ou +	551	4.589.477	8.329,4	440	7.100.503	16.137,5	165	1.915.593	11.609,7
Sul	3.841	2.451.895	638,3	3.189	3.702.974	1.161,2	1.374	1.025.853	746,6
Até 5.000	704	131.064	186,2	516	196.436	380,7	231	69.386	300,4
5.001-10.000	665	212.672	319,8	407	307.423	755,3	184	103.677	563,5
10.001-50.000	751	728.887	970,6	633	1.060.592	1.675,5	290	311.578	1.074,4
50.001- 100.000	896	363.703	405,9	769	553.576	719,9	291	144.581	496,8
100.000 ou +	825	1.015.569	1.231,0	864	1.584.947	1.834,4	378	396.631	1.049,3
Centro-Oeste	2.136	1.236.637	578,9	2.046	1.591.793	778,0	731	331.743	453,8
Até 5.000	298	52.154	175,0	278	68.127	245,1	103	18.432	179,0
5.001-10.000	273	88.919	325,7	269	114.003	423,8	95	28.674	301,8
10.001-50.000	506	343.407	678,7	555	439.176	791,3	195	105.315	540,1
50.001- 100.000	512	135.616	264,9	467	166.813	357,2	172	32.623	189,7
100.000 ou +	547	616.541	1.127,1	477	803.674	1.684,9	166	146.699	883,7

* População em 2000 (fonte: *Censo Demográfico* de 2000);** Razão entre o número de pessoas na população em 1^a de setembro de 2000 e na amostra efetiva do SB Brasil 2003.

te populacional (mais de 100 mil habitantes) de cada macrorregião, os municípios de capital não correspondem à UPA e sim a estratos de seleção, uma vez que foram incluídos na amostra, sem procedimento de seleção aleatória. Além disso, no estrato de maior porte populacional da macrorregião Nordeste, as nove capitais foram incluídas, tendo sido selecionado apenas um município, o que viola um princípio básico de amostragem: o tamanho da amostra deve ser maior que um, para permitir o cálculo da variabilidade. Este problema pode ser corrigido pelo agrupamento dos municípios deste estrato (exceto os de

capital) com os do estrato de porte populacional imediatamente inferior (50.001 a 100.000 habitantes). Dessa forma, são 51 os estratos da amostra, correspondentes a: 27 capitais; mais 20 estratos correspondentes à combinação das regiões Norte, Sudeste, Sul e Centro-Oeste com 5 estratos de porte populacionais; mais quatro estratos de porte populacional na Região Nordeste.

Selecionados ou incluídos os municípios, os demais estágios de seleção foram diferenciados em dois grupos de estratos: (1) os de até 50 mil habitantes; e (2) os de mais de 50 mil habitantes. Dentro dos municípios com até 50 mil ha-

Tabela 2

Varição, média e coeficiente de variação do peso relativo, segundo os grupos etários.

Grupos etários	Varição do peso relativo	Média	Coeficiente de variação (CV)
18-36 meses	63,1 e 3.733,5	534,3	3.187%
5 anos	21,3 e 744,5	129,3	3.413%
12 anos	17,8 a 363,6	102,0	3.676%
15-19 anos	66,4 a 8.329,4	1.065,8	3.865%
35-44 anos	90,0 a 16.137,5	1.698,2	3.962%
65- 74 anos	89,5 a 11.609,7	1.182,2	2.520%

bitantes, foram selecionadas quadras e escolas (segundo estágio). Dentro das quadras, foram selecionados os domicílios (terceiro estágio) em que foram entrevistados os indivíduos. E, dentro das escolas, foram selecionados, no terceiro estágio, os alunos com a idade de interesse (12 anos). Para os municípios com mais de 50 mil habitantes foram selecionados setores, escolas e creches (primeiro estágio, nos municípios de capital, e segundo estágio nos demais). Dentro dos setores, foram selecionadas quadras (segundo ou terceiro estágio) e posteriormente domicílios (terceiro ou quarto estágio) em que foram entrevistados os indivíduos. E, dentro das escolas e creches, foram selecionados os alunos com as idades de interesse (5 e 12 anos). A Tabela 3 fornece indicação dos estratos, das unidades elementares da amostra e dos distintos estágios de seleção para os diferentes grupos definidos no plano de amostragem.

Para o cálculo dos pesos amostrais, definidos como o inverso das probabilidades de inclusão das unidades elementares na amostra, é necessário identificar as unidades de seleção de cada estágio da amostra, a saber: municípios, escolas, creches, setores, quadras/vilas, domicílios e alunos. Tais unidades, por sua vez, variam de acordo com o porte do município, com o fato de o município ser de capital ou não, e com o fato de o indivíduo ter sido entrevistado em domicílio ou estabelecimento de ensino, como indica a Tabela 3.

Partindo do pressuposto de que a coleta de dados foi feita exatamente como o manual do coordenador¹⁴ estabelece, é possível obter as informações necessárias para cálculo dos pesos amostrais. De fato, o referido manual previu um modelo de relatório de campo no qual deveriam constar informações que possibilitariam a recuperação dos dados necessários ao cálculo das probabilidades de inclusão de quadras urbanas e vilas rurais sorteadas, além do número total de domicílios identificados, e do período de coleta

de dados nos domicílios e nas escolas, indicando que esses relatórios seriam a fonte preferencial para obtenção dos dados requeridos. Porém, na ausência desses dados nos relatórios ou no caso dos relatórios não serem recuperados, a opção que resta é buscar informações em fontes exógenas (IBGE e MEC) e pressupor que as regras do referido manual foram estritamente seguidas. Além disso, como não foi prevista a inclusão de informações sobre as escolas selecionadas nem sobre os respectivos números de alunos nos relatórios ou na folha do questionário de coleta de dados, só restam para esses casos os dados do Censo Escolar do MEC.

Para municípios, as informações necessárias para determinar as probabilidades de inclusão constam do manual do coordenador¹⁴. No caso dos municípios de capital, a probabilidade de inclusão é igual a um, pois eles foram incluídos na amostra sem seleção. Para os municípios selecionados por processo aleatório, a probabilidade de inclusão é proporcional à estimativa populacional dos municípios para 1º de julho de 1999. As referidas probabilidades de inclusão constam das expressões 1 e 2 da Figura 1.

No caso da seleção de setores, aplicável apenas a municípios com mais de 50 mil habitantes, embora não conste do manual do coordenador, foi feita uma seleção equiprovável e a probabilidade de inclusão consta da expressão 3 da Figura 1. Os dados da Base Operacional Geográfica (BOG) do IBGE são suficientes para determinar as probabilidades de cada setor selecionado. Os códigos desses setores constam, exceto por alguns erros de digitação, da base de dados do SB Brasil 2003. A expectativa é que o relatório da equipe técnica responsável pela seleção da amostra permita corrigir esses poucos erros de digitação.

A seleção de quadras/vilas foi feita com reposição e equi-probabilidade, e as probabilidades de inclusão constam das expressões 4 e 5 da Figura 1. Há pequena variação na forma de

Tabela 3

Estratos, unidades elementares e estágios de seleção por porte do município, condição do município (capital ou não) e local da entrevista.

Estratos, unidades elementares e estágios de seleção	Município até 50.000 habitantes		Municípios com mais de 50.000 habitantes			
	Entrevista domiciliar	Entrevista na escola	Capitais		Outros municípios	
			Entrevista domiciliar	Entrevista na escola	Entrevista domiciliar	Entrevista na escola
Estratos de seleção	Macrorregião e porte populacional		Município		Macrorregião e porte populacional	
Unidades elementares	Moradores	Alunos (12 anos)	Moradores	Alunos (5 ou 12 anos)	Moradores	Alunos (5 ou 12 anos)
Unidade primária de amostragem	Município		Setor censitário	Escola ou creche *	Município	
Segundo estágio de seleção	Quadra ou vila	Escola (se mais de 20) ou aluno (se até 20 escolas)	Quadra ou vila	Aluno	Setor censitário	Escola (se mais de 20) ou aluno (se até 20 escolas)
Terceiro estágio de seleção	Domicílio **	Aluno (se mais de 20 escolas)	Domicílio **		Quadra ou vila	Aluno (se mais de 20 escolas)
Quarto estágio de seleção					Domicílio **	

* Supõe-se que existam mais de 20 escolas e mais de 20 creches em todas as capitais, caso contrário valem as regras para outros municípios com mais de 50.000 habitantes. Este é, em princípio, o único caso em que será necessário usar simulação estatística para avaliar o efeito de diferentes agrupamentos de alunos por escola ou creche;

** De fato, os domicílios não são sempre selecionados. Para alguns grupos etários, todos os moradores de todos os domicílios são entrevistados (os que se recusarem são considerados como não-resposta). No entanto, para certos grupos etários mais frequentes na população foi estabelecida uma fração de amostragem, que implica a seleção de um a cada dois domicílios. Neste caso, há seleção de domicílios e são entrevistados todos os moradores do grupo etário, desde que dêem seu consentimento.

calcular numerador e denominador em função da maneira como é calculada a média de domicílios por quadra. Nos municípios até 50 mil habitantes, onde não há seleção de setores, essa média parte do pressuposto de que o tamanho médio do domicílio é igual a quatro pessoas (expressão 5.1 da Figura 1). Nos demais municípios, onde os setores foram selecionados, essa média parte do pressuposto de que os setores têm, em média, 300 domicílios (expressão 5.3 da Figura 1).

Há, também, uma certa imprecisão na definição do tamanho da amostra de domicílios (usados nas expressões 5.2 e 5.4 da Figura 1) no manual do coordenador¹⁴, em função de ajustes feitos (correção de população finita, população média dos municípios em cada estrato etc.). Essa imprecisão fica evidente nos dados apresentados na Tabela 4. Mas isso não é importante para calcular as probabilidades de inclusão, pois será usado o tamanho efetivo da amostra, determinado com base do banco de dados do SB Brasil 2003. O uso do tamanho efetivo da amostra corresponde a uma correção de não-resposta, visto que irá alterar o peso amostral para com-

pensar a diferença entre a amostra desejada e a efetiva (que, no caso, é a não-resposta).

Voltando à estratégia para calcular as probabilidades de inclusão, serão usados os dados dos relatórios das equipes de campo. Caso eles não existam, o número de quadras/vilas pode ser obtido nos mapas dos setores selecionados (municípios com mais de 50.000 habitantes) ou nos Mapas Municipais para fins Estatísticos (MME), disponíveis na BOG do IBGE.

Para a seleção dos domicílios, deve-se considerar duas possibilidades que dependem do estrato da amostra e da faixa etária em consideração, como indica a expressão 6 da Figura 1: (1) não há seleção de domicílios e todas as unidades domiciliares existentes na quadra entram na amostra (probabilidade de inclusão igual a um), ou (2) a seleção é realizada de forma sistemática, sendo pesquisado um a cada dois domicílios (probabilidade de inclusão igual a 1/2). As frações de amostragem de 100% ou 50% estão indicadas na Tabela 4. Também nesse estágio de seleção não é necessário contar os domicílios das quadras e vilas selecionados, bastando contá-los na base de dados do SB Brasil 2003 e aplicar a fração

Figura 1

Esquema probabilístico da amostra do SB Brasil 2003.

A. Probabilidades de inclusão na amostra para pessoas entrevistadas nos domicílios

Probabilidade de inclusão de municípios

(1) $P(\text{município de capital}) = 1$

(2) $P(\text{outros municípios}) = \frac{n \times \text{população do município em 01/07/99}}{\text{população do estrato em 01/07/99}}$, onde

n é o número de municípios selecionados no estrato e, de um modo geral igual a 10, exceto para o maior estrato de porte populacional das macrorregiões Norte, Sudeste, Sul e Centro-Oeste, em que n é igual a 3, 6, 7 e 6, respectivamente, bem como para o quarto estrato de porte populacional (50.001 a 100.000 habitantes) do Nordeste onde n é igual 11 (pela inclusão do único município de mais de 100 mil habitantes selecionado).

Probabilidade de inclusão de setores (apenas para municípios com mais de 50.000 habitantes):

(3) $P(\text{setor}) = \frac{10}{\text{número de setores do município}}$, se seleção com equi-probabilidade; ou

Probabilidade de inclusão de quadras ou vilas rurais:

(4) $P(\text{quadras/vila}) = 1$, se o município tiver até 10 quadras; ou

(5) $P(\text{quadras/vila}) = \frac{TA_{qv}}{TP_{qv}}$, onde

TA_{qv} é o tamanho da amostra de quadras e vilas; e

TP_{qv} é o tamanho da população de quadras e vilas.

Para o cálculo de TA_{qv} e TP_{qv} as regras mudam de acordo com a população municipal.

Para municípios com até 50 mil habitantes, não há seleção de setores e o tamanho da amostra depende do número médio de domicílios (MDQ) por quadra/vila, como indicado nas expressões 5.1 e 5.2:

$$(5.1) \text{MDQ} = \frac{\text{População total do município}}{4 \times \text{Número de quadras/vilas do município}}$$

$$(5.2) TA_{qv} = \frac{TA_d}{\text{MDQ}}$$
 onde TA_d é o tamanho da amostra de domicílios indicado na Tabela 4.

Para municípios com mais de 50 mil habitantes, há seleção de setores e o tamanho da amostra depende do número médio de domicílios (MDQ) por quadra/vila nos 10 setores selecionados, como indicado nas expressões 5.3 e 5.4:

$$(5.3) \text{MDQ} = \frac{300 \times 10}{\text{Número de quadras/vilas nos 10 setores selecionados}}$$

$$(5.4) TA_{qv} = \frac{TA_d}{\text{MDQ}}$$
 onde TA_d é o tamanho da amostra de domicílios indicado na Tabela 4.

Probabilidade de inclusão dos domicílios

(6) $P(\text{domicílio}) = 1$ ou $\frac{1}{2}$, de acordo com as frações de amostragem indicadas na Tabela 4.

Probabilidade de inclusão das unidades elementares (moradores)

Como os moradores não são alvo de seleção, pois todos os residentes em domicílios selecionados que aceitaram a pesquisa foram examinados, sua probabilidade de inclusão corresponde ao produto das probabilidades de inclusão nos diferentes estágios de seleção.

B. Probabilidades de inclusão para crianças selecionadas em estabelecimentos de ensino

Probabilidade de inclusão de municípios

São as mesmas definidas anteriormente nas expressões (1) e (2).

(continua)

Figura 1 (continuação)

Probabilidade de inclusão de escolas (para seleção de crianças de 12 anos)

(7) $P(\text{escola}) = 1$, se o município tiver até 20 escolas de 1ª grau freqüentadas por crianças de 12 anos; ou

$$(8) P(\text{escola}) = \frac{20 \times \text{número de alunos da escola}}{\text{Total de alunos de 12 nas escolas do município}}$$

Probabilidade de inclusão de creches (crianças de 5 anos e municípios com população > 50.000)

(9) $P(\text{creche}) = 1$, se o município tiver até 20 creches freqüentadas por crianças de 5 anos; ou

$$(10) P(\text{creche}) = \frac{\text{Tamanho da amostra de alunos}}{\text{Total de alunos de 5 anos nas creches do município}}$$

Probabilidade de inclusão de crianças de 5 e 12 anos

$$(11) P(\text{criança}) = \frac{\text{Tamanho de amostra de alunos}}{\text{Total de alunos nas escolas selecionadas (ou incluídas)}}$$

Probabilidade de inclusão das unidades elementares (alunos)

A probabilidade de inclusão dos alunos corresponde ao produto das probabilidades de inclusão do município, da escola (ou creche) e da criança.

de amostragem correspondente à idade das pessoas entrevistadas.

As pessoas de 12 anos foram sempre selecionadas em escolas. As crianças de 5 anos foram selecionadas nos domicílios, se o município tinha até 50.000 habitantes, ou em creches no caso de municípios com mais de 50.000 habitantes. Para os estabelecimentos de ensino (escolas e creches), as probabilidades de inclusão são dadas pelas expressões 7 a 10 da Figura 1. As expressões 7 e 9 aplicam-se aos municípios com até 20 escolas ou creches, respectivamente, caso no qual os estabelecimentos de ensino são incluídos na amostra com probabilidade igual a um. As expressões 8 e 10 da Figura 1 indicam que, no caso de mais de 20 escolas e mais de 20 creches, elas foram selecionadas com probabilidade proporcional ao seu número de alunos. Para determinar as probabilidades de inclusão das expressões 8 e 10 serão preferencialmente usadas as informações dos relatórios das equipes de campo. Na ausência dessas informações, a melhor aproximação possível é usar o número de alunos de 6ª série de cada escola (ou de cinco anos na pré-escola) selecionada e o número de alunos de 6ª série de todas as escolas (ou de 5 anos de todas as pré-escolas) do município, disponibilizados pelo Censo Escolar de 2003 (MEC).

Para crianças de 5 e 12 anos selecionadas em estabelecimentos de ensino, foi adotado um procedimento de seleção sistemática com equi-probabilidade, independentemente do número

de estabelecimentos de ensino do município. A probabilidade de inclusão de um aluno consta da expressão 11 da Figura 1. Também, neste caso, será usado o tamanho da amostra efetiva como numerador (para correção das não-respostas) e o denominador, obtido da mesma fonte usada para o numerador das expressões 8 e 10 da Figura 1, corresponde à soma de alunos de 12 anos (ou 5 anos) das escolas (ou creches) selecionadas ou incluídas na amostra.

Enfim, a fonte de informações preferencial para o cálculo das probabilidades de inclusão são os relatórios das equipes de campo. Todavia, não sendo possível recuperar todas as informações necessárias nesses relatórios, o uso de informações do IBGE e do INEP permitirá obter probabilidades de inclusão aproximadas em cada estágio de seleção. Com essas probabilidades de inclusão, exatas ou aproximadas, o peso amostral de cada unidade elementar da amostra pode ser obtido pelo inverso do produto das suas probabilidades de inclusão em cada estágio da amostra, como indicado na Figura 1.

Após o cálculo exato ou aproximado dos pesos amostrais, um processo de calibração desses pesos deve ser aplicado para assegurar que as estimativas reproduzam os totais populacionais por sexo e faixa de idade, conhecidos para todos os estratos de seleção mediante a projeção da população feita pelo IBGE, para o ano de 2003.

Entretanto, calcular os pesos amostrais não é suficiente. É preciso identificar cada estrato de

Tabela 4

Número previsto de domicílios na amostra, tamanho da amostra e fração de amostragem por faixa etária, segundo as macrorregiões e os estratos de porte populacional.

Macrorregiões e porte populacional (habitantes)	Número previsto de domicílios na amostra	Faixa etária					
		18-36 meses	5 anos	12 anos	15-19 anos	35-44 anos	65-74 anos
Tamanho da amostra							
Norte		100	145	178	91	51	15
Nordeste		100	163	240	97	59	23
Sudeste		100	329	297	76	43	14
Sul		100	218	238	92	40	13
Centro-Oeste		100	214	218	70	40	14
Fração de amostragem							
Norte							
Até 5.000	157	1	1	EE	1	1	1
5.001-10.000	175	1	1	EE	1	1	1
10.001-50.000	187	1	1	EE	1	1	1
Mais de 50.000	191	1	EE	EE	1	1	1
Nordeste							
Até 5.000	172	1	1	EE	1	1	1
5.001-10.000	188	1	1	EE	1	1	1
10.001-50.000	201	1	1	EE	1	1	1
Mais de 50.000	208	1	EE	EE	1	1	1
Sudeste							
Até 5.000	160	1	1	EE	1	1/2	1
5.001-10.000	172	1	1	EE	1	1/2	1
10.001-50.000	184	1	1	EE	1	1/2	1
Mais de 50.000	190	1	EE	EE	1	1/2	1/2
Sul							
Até 5.000	184	1	1	EE	1	1/2	1
5.001-10.000	210	1	1	EE	1	1/2	1
10.001-50.000	230	1	1	EE	1	1/2	1
Mais de 50.000	236	1	EE	EE	1	1/2	1/2
Centro-Oeste							
Até 5.000	144	1	1	EE	1	1/2	1
5.001-10.000	150	1	1	EE	1	1/2	1
10.001-50.000	155	1	1	EE	1	1	1
Mais de 50.000	160	1	EE	EE	1	1/2	1

EE: pesquisado em estabelecimento de ensino.

Adaptado das Tabelas 2, 4 e 5 do manual do coordenador ¹⁴.

seleção e todas as suas unidades primárias de amostragem, a fim de viabilizar o cálculo de variâncias e demais estatísticas delas dependentes.

O manual do coordenador ¹⁴ esclarece que a amostra tem 25 estratos (5 macrorregiões x 5 estratos de porte populacional). Já o texto anterior mostra que, na realidade, a amostra do SB Brasil 2003 foi selecionada de uma população estratificada em 51 estratos (= 27 capitais + 20 estratos de porte populacional nas macrorregiões Norte, Sudeste, Sul e Centro-Oeste + 4 estratos de porte

populacional na macrorregião Nordeste, onde o estrato de maior porte populacional foi agrupado ao estrato anterior, como indicado anteriormente).

Também não são 250 UPA (10 municípios x 25 estratos), como está descrito no referido manual, apesar de o número exato de UPA ainda não ter sido determinado. A Tabela 3 mostra que, de acordo com o estrato de seleção, a UPA pode ser o município, o setor censitário, a escola ou a creche. Como o número de escolas e creches ain-

da não foi determinado, o número total de UPA ainda é desconhecido. O problema, nesse caso, é que o arquivo de dados não identifica cada estabelecimento de ensino (escola ou creche) selecionado, apenas identificando o setor, a quadra ou vila e o tipo de estabelecimento de ensino. Os relatórios das equipes de campo constituem-se na fonte ideal para determinar quais estabelecimentos de ensino foram selecionados em cada município da amostra. Na sua ausência, a única alternativa é fazer uso da técnica de simulação estatística para identificar o impacto de diferentes agrupamentos (aleatoriamente construídos em cada município) de alunos por escola (ou creche) sobre as variâncias das principais variáveis da pesquisa. Se o impacto não for grande, pode-se determinar – e identificar no arquivo – os agrupamentos a serem utilizados no processo de estimação de variâncias e estatísticas delas derivadas.

Conclusão

Trata-se, portanto, de uma proposta de correção trabalhosa, que requer a busca de informações junto às coordenações municipais, estaduais e regionais do SB Brasil 2003 e, para completar as informações faltantes, o processamento de bases de dados exógenas ao projeto.

Contudo, o esforço de completar o processo de amostragem probabilística do SB Brasil 2003, calculando pesos amostrais e identificando as demais informações estruturais de sua amostra, justifica-se para ampliar o uso das informações coletadas, que, em última instância, é o que justifica o custo de qualquer pesquisa.

Além disso, inferências válidas para a população, então tecnicamente viáveis, permitirão que pesquisadores e gestores possam analisar as reais condições de saúde bucal no país e traçar metas para seu melhoramento.

Resumo

A Pesquisa sobre as Condições de Saúde Bucal da População Brasileira (SB Brasil 2003) foi a mais abrangente pesquisa sobre as condições de saúde bucal realizada no Brasil até hoje. Métodos de amostragem probabilística foram utilizados para assegurar que os dados coletados representariam a população das faixas etárias escolhidas nas cinco regiões do Brasil. No entanto, isso não foi possível, pois o processo de amostragem da pesquisa não foi finalizado. Sua conclusão pressupõe o cálculo dos pesos amostrais e a identificação das demais variáveis estruturais da amostra (estratos de seleção e unidades primárias de amostragem). Este artigo descreve o desenho de amostra desta pesquisa, formula as probabilidades de inclusão nos diversos estágios de seleção e propõe estratégias de cálculo dos pesos amostrais. A solução para determinar os pesos amostrais e identificar as demais variáveis estruturais da amostra, proposta no artigo, parte da recuperação de informações que deveriam ter sido registradas nos relatórios produzidos durante a pesquisa de campo e, em sua ausência, em informações disponíveis no Instituto Brasileiro de Geografia e Estatística e no Ministério da Educação, como aproximações válidas.

Saúde Bucal; Amostragem; Nível de Saúde; População

Colaboradores

M. C. Portela e R. C. S. Queiroz identificaram a falta de pesos amostrais no arquivo de dados do SB Brasil 2003. M. T. L. Vasconcelos orientou o estudo dos documentos existentes sobre o desenho da amostra e identificou os demais problemas do desenho. Os três autores planejaram os procedimentos de correção dos problemas identificados. R. C. S. Queiroz escreveu a primeira versão do manuscrito, que foi revisado e aprovado pelos demais autores.

Agradecimentos

Os autores agradecem à Maria Cecília Goi Porto Alves as informações relativas ao teste piloto realizado em Diadema (São Paulo) e Canela (Rio Grande do Sul), e a Angelo Giuseppe Roncalli da Costa Oliveira os esclarecimentos prestados sobre a técnica de seleção dos setores, além de sua cooperação ao fornecer a listagem dos setores selecionados.

Este trabalho recebeu apoio da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; processo E-26/171.336/2006), e os seus autores são bolsistas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

1. Divisão Nacional de Saúde Bucal, Ministério da Saúde. Levantamento epidemiológico em saúde bucal: Brasil, zona urbana. Brasília: Ministério da Saúde; 1988.
2. Oliveira AGR, Unfer B, Costa ICC, Arcieri RM, Guimaraes LOC, Saliba NA. Levantamentos epidemiológicos em saúde bucal: análise da metodologia proposta pela Organização Mundial da Saúde. *Rev Bras Epidemiol* 1998; 1:177-89.
3. Fédération Dentaire Internationale. Global goals for oral health in the year 2000. *Int Dent J* 1982; 32:74-7.
4. Costa JFR, Chagas LD, Silvestre RM, organizadores. A política nacional de saúde bucal do Brasil: registro de uma conquista histórica. Brasília: Organização Pan-Americana da Saúde/Ministério da Saúde; 2006. (Serie Técnica Desenvolvimento de Sistemas e Serviços de Saúde, 11).
5. Coordenação de Saúde Bucal, Departamento de Assistência e Promoção à Saúde, Secretaria de Assistência à Saúde, Ministério da Saúde. Levantamento epidemiológico em saúde bucal: 1ª etapa - cárie dental. Brasília: Ministério da Saúde; 1996.
6. Departamento de Atenção Básica, Secretaria de Atenção à Saúde, Ministério da Saúde. Projeto SB Brasil 2003. Condições de saúde bucal da população brasileira, 2002-2003: resultados principais. Brasília: Ministério da Saúde; 2004.
7. Antunes JLF, Peres MA, Mello TRC, Waldman EA. Multilevel assessment of determinants of dental caries experience in Brazil. *Community Dent Oral Epidemiol* 2006; 34:146-52.
8. Antunes JLF, Peres MA, Mello TRC. Determinantes individuais e contextuais da necessidade de tratamento odontológico na dentição decídua no Brasil. *Ciênc Saúde Coletiva* 2006; 11:79-87.
9. Matos DL, Lima-Costa MF. Auto-avaliação da saúde bucal entre adultos e idosos residentes na Região Sudeste: resultados do Projeto SB-Brasil, 2003. *Cad Saúde Pública* 2006; 22:1699-707.
10. Narvai P, Frazão P, Roncalli A, Antunes J. Cárie dentária no Brasil: declínio, polarização, iniquidade e exclusão social. *Rev Panam Salud Pública* 2006; 19:385-93.
11. Kiar AN. Observations et expériences concernant les dénombrements représentatifs. *Bulletin of the International Statistical Institute* 1895; 9:176-205.
12. Cochran WG. *Sampling techniques*. 3rd Ed. New York: John Wiley and Sons; 1977.
13. Pessoa DGC, Silva PLN. *Análise de dados amostrais complexos*. São Paulo: Associação Brasileira de Estatística; 1998.
14. Área Técnica de Saúde Bucal, Departamento de Atenção Básica, Secretaria Políticas de Saúde, Ministério da Saúde. Projeto SB2000. Condições de saúde bucal da população brasileira no ano 2000: manual do coordenador. Brasília: Ministério da Saúde; 2001.

Recebido em 02/Jan/2008

Versão final reapresentada em 15/Mai/2008

Aprovado em 06/Jun/2008