

Amostra mestra e geoprocessamento: tecnologias para inquéritos domiciliares

Master sample and geoprocessing: technologies for household surveys

Nilza Nunes da Silva^a, Tarcisio Neves da Cunha^b e José Alberto Quintanilha^b

^aDepartamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, SP, Brasil. ^bEscola Politécnica da Universidade de São Paulo. São Paulo, SP, Brasil

Descritores

Amostras. Amostragem.
Epidemiologia descritiva. Censos.
Geografia. Sistemas de informação.

Resumo

Objetivo

Reduzir as despesas e o tempo associados ao processo de amostragem de domicílios e mostrar a viabilidade de uso compartilhado dos cadastros de endereços entre vários inquéritos epidemiológicos, a partir da reciclagem de material produzido pela Pesquisa Nacional de Amostras de Domicílios (PNAD).

Métodos

Cadastros permanentes de endereços foram construídos e atualizados para 72 setores censitários, e retidos como unidades primárias da amostra mestra do município de São Paulo, SP. No período 1995-2000, sortearam-se três diferentes amostras pelo processo de conglomerados, em dois estágios. Operações de sorteio e mapeamento das amostras foram automatizadas por meio da aplicação de tecnologia de geoprocessamento.

Resultados

Cerca de 25.000 domicílios compõem o cadastro permanente de endereços da amostra mestra. A produção ágil e barata de cada amostra, aliada ao conhecimento acumulado sobre os perfis demográficos e topográficos dos setores censitários, pode ser considerada a principal contribuição dos resultados alcançados. Verificou-se que o sistema proposto contrasta com a seleção de amostras independentes para cada investigação "ad hoc", o que ainda é prática comum nas agências ou grupos que realizam inquéritos domiciliares.

Conclusões

A reciclagem das listagens anuais de endereços da PNAD, apesar de restringir a abrangência de replicação da experiência aos municípios mais populosos do País, evita a repetição do processo de amostragem para cada inquérito; além disso, significa importante contribuição ao planejamento de investigações descritivas na área da saúde pública. Questões conceituais ou teóricas, decorrentes do procedimento operacional, foram facilmente superadas pela introdução de técnicas estatísticas compensatórias.

Keywords

Samples. Sampling studies.
Epidemiology, descriptive. Censuses.
Geography. Information systems.

Abstract

Objective

To reduce cost and time associated with household sampling process and to assess the feasibility of shared use of address data file of census enumeration areas in

Correspondência para/ Correspondence to:

Nilza Nunes da Silva
Departamento de Epidemiologia/FSP/USP
Av. Dr. Arnaldo, 715
01246-904 São Paulo, SP, Brasil
E-mail: nndsilva@usp.br

Baseado em tese de livre-docência apresentada ao Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo, em 2001.

Recebido em 17/5/2002. Reapresentado em 24/2/2003. Aprovado em 28/2/2003.

several epidemiological surveys using updated information from the National Survey of Households (PNAD).

Methods

Address data file comprising 72 census enumeration areas was kept as primary sampling units for the city of São Paulo. During the period 1995-2000, three distinct household samples were drawn using the two-stage cluster sampling procedure. Geographic Information System (GIS) technology allowed delimiting boundaries, blocks and streets for any primary sampling unit and printing updated maps for selected sub-samples.

Results

Twenty-five thousand dwellings made up the permanent address data file of the master sample. A cheaper and quicker selection of each sample, plus gathering information on demographic and topographical profiles of census enumeration areas were the main contribution of the study results.

Conclusions

The master sample concept, integrated with GIS technology, is an advantageous alternative sampling design for household surveys in urban areas. Using the list of addresses from the PNAD updated yearly, although limiting its application to the most populated Brazilian cities, avoids the need of creating an independent sampling procedure for each individual survey carried out in the period between demographic censuses, and it is an important contribution for planning sampling surveys in public health.

INTRODUÇÃO

Os inquéritos epidemiológicos realizados em populações brasileiras encontram, mesmo atualmente, sérias dificuldades quanto ao processo de obtenção das amostras de domicílios. O sorteio por conglomerados em vários estágios, em que o setor censitário é a unidade primária de amostragem, é o método probabilístico usualmente adotado (Kish,⁶ 1965). Informações demográficas sobre os setores censitários, sorteados no primeiro estágio, podem ser adquiridas na Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), que também produz cópias dos mapas construídos para os censos demográficos, realizados a cada 10 anos.

Entretanto, a identificação dos domicílios atualizada para a época em que se realiza o inquérito, o sorteio das sub-amostras de endereços e a elaboração de roteiros para o percurso dos entrevistadores no campo são tarefas que demandam, no segundo estágio do processo, a montagem de infra-estrutura técnica especializada. Isto impõe gastos e tempo em geral inaceitáveis aos protocolos bastante limitados pela realidade financeira da investigação científica no Brasil. Ainda assim, o esforço empreendido para assegurar as despesas referentes ao processo de amostragem, bem como a capacitação de pessoal auxiliar para a composição das listagens e sorteio dos endereços, podem significar real prejuízo, considerando-se que a maioria dos inquéritos ainda são realizados isoladamente dentro de uma mesma ins-

tuição ou área de pesquisa, de modo que todo o material e informações produzidas raramente serão reaproveitados em outras pesquisas.

Por outro lado, a retenção dos setores censitários como unidades primárias de amostragem durante o período intercensitário e o seu uso em diversos inquéritos epidemiológicos têm sido práticas usuais nos institutos de pesquisa dos países desenvolvidos (King,⁵ 1945; Sindini,¹¹ 1999). Além de se dividir as despesas do processo de amostragem entre vários orçamentos de pesquisa, o tempo que seria dedicado a essa etapa da investigação científica é drasticamente reduzido. Também os marcos de amostragem, produzidos pelos censos demográficos nacionais, têm sido amplamente utilizados para a obtenção de sub-amostras de domicílios. Uma experiência recente no Brasil foi o trabalho desenvolvido pela Sociedade Civil Bem-Estar Familiar (BEMFAM,¹ 1997), que utilizou os setores censitários da Pesquisa Nacional por Amostra de Domicílios (PNAD, 1995) para definir uma amostra de 13.283 domicílios e realizar a Pesquisa Nacional sobre Demografia e Saúde.

A PNAD é uma das pesquisas anuais da Fundação (FIBGE,⁴ 1991), que produz, desde 1967, informações socioeconômicas para o Brasil, estimadas através de uma amostra de domicílios obtida pelo processo de conglomerados em três estágios. Com infra-estrutura fixa em todo território nacional, essa pesquisa produz anualmente a listagem dos endereços dos domicílios particulares permanentes de cada se-

tor censitário, que compõem a amostra dos municípios sorteados no primeiro estágio. É possível, portanto, imaginar esse material como um cadastro permanente de endereços a ser utilizado para a obtenção consecutiva de amostras de domicílios particulares nesses municípios.

Assim, o presente trabalho apresenta uma amostra de 72 setores censitários do Município de São Paulo que, obtida para a realização de um inquérito nutricional em 1995, passou a ser considerada um cadastro permanente de endereços para a definição de outras amostras de domicílios a serem usadas pelos demais inquéritos epidemiológicos que se sucederem nessa área. Um sistema digital, construído com base nesse cadastro e dotado dos recursos do geoprocessamento, sorteia endereços, localiza-os em suas respectivas quadras e produz saídas impressas, contendo os mapas e as listagens das amostras de domicílios de cada setor censitário.

MÉTODOS

A amostra da PNAD-1995 para o Município de São Paulo era composta por 263 setores censitários, sorteados do total de 10.096 definidos pela malha territorial do censo demográfico de 1991. Setenta e dois setores censitários, sorteados dessa amostra, formam o conjunto de unidades primárias do cadastro permanente, elaborado para o período 1995-2000.

Cópias das “cadernetas das áreas de listagens” contendo os mapas, descrição dos perímetros, nomes dos logradouros, localização das quadras e respectivas faces, foram obtidas no Centro de Documentação e Disseminação de Informações da Fundação IBGE. As listagens dos endereços de domicílios particulares permanentes de cada setor censitário para os anos de 1995, 1998 e 1999, cedidas pela Fundação IBGE, também foram providenciadas por aquele Centro.

Para cada setor, construíram-se arquivos digitais da base cartográfica e da listagem de endereços dos domicílios particulares permanentes. A seguir, os endereços em formato DBase foram geocodificados e lançados como camadas independentes sobre a base cartográfica do município, gerando-se os 72 cadastros georreferenciados de domicílios particulares permanentes. O Map-Info desktop versão 4.1 foi o software utilizado nessa operação. A geocodificação por endereços foi implementada posicionando-se cada número de domicílio com a métrica de cada rua em que estavam situados, possibilitando, assim, a geração automática de suas coordenadas geográficas. Esse procedimento define coordenadas posicionalmente aproximadas, mas sua topologia é supostamente cor-

reta em termos de lado da rua, segmento e posição sequencial dos endereços.

A sub-amostra de 72 unidades primárias (setores censitários) foi obtida por sorteio sistemático aplicado ao arquivo da PNAD-1995 (Cochran,² 1977), ordenado segundo distrito e grau de escolaridade do chefe da família. Como a amostra de setores censitários para a PNAD-95 no Município de São Paulo foi sorteada sob o critério de probabilidade proporcional ao tamanho, a fração de amostragem correspondente a esse estágio do sorteio é calculada pela expressão

$$f_1 = \left[\frac{263 \cdot D_j}{\sum_j D_j} \times \frac{72}{263} \right]$$

onde o primeiro termo refere-se à probabilidade dos setores censitários serem sorteados no estágio do processo de amostragem desenvolvido pela Fundação IBGE, e o segundo é o inverso do intervalo adotado para o sorteio sistemático da sub-amostra tomada para o cadastro permanente. O tamanho do *j*-ésimo setor, representado por D_j , corresponde ao número de domicílios particulares permanentes contados no censo demográfico brasileiro de 1991.

Diversas amostras de domicílios poderão ser sorteadas do cadastro permanente, até que todos os endereços de grande parte das unidades primárias estejam já sorteados e entrevistas tenham sido realizadas. O sistema gerou, no período 1995-2000, amostras para três levantamentos realizados na área da saúde pública. Procurando atender às características demográficas das populações alvo de cada um desses inquéritos epidemiológicos, o número de endereços sorteados (*d*) para um inquérito é definido em função do número desejado de entrevistas (*n*) e da densidade demográfica por domicílio (RD). Assim, o uso da expressão [$d_j = (n / RD) / 72$] permitiu calcular o número de endereços a serem sorteados na amostra do *j*-ésimo setor censitário, a fim de que, no final do trabalho de campo, seja alcançado o número mínimo de entrevistas pré-fixado pelo critério de precisão de cada pesquisa.

Do arquivo de endereços de cada setor sortearam-se os d_j endereços dentre o total disponível (D_j), resultante da atualização realizada para o correspondente inquérito. O tipo de sorteio foi adaptado às condições apontadas pela logística do trabalho de campo de cada inquérito, levando-se em consideração aspectos de redução de custos e de tempo para o acesso aos domicílios em que se realizariam

as entrevistas. A probabilidade de um particular domicílio ser sorteado neste estágio foi calculada, para cada inquérito, pela expressão

$$\left[f_2 = \frac{d_j}{D_j} \right]$$

cujos termos foram apresentados no início deste parágrafo.

As amostras sorteadas foram marcadas nas respectivas tabelas de endereços georreferenciados de cada setor censitário. A partir daí, foram construídos os mapas que indicam a localização, por quadra, dos domicílios a serem visitados para busca do respondente elegível. O sistema produz, ainda, listagens impressas dos endereços sorteados em cada setor censitário.

O procedimento de sorteios sucessivos de endereços, oriundos das mesmas unidades primárias utilizadas durante o período inter-censitário, resulta em alterações nas características iniciais dos planos de amostragem.

A primeira alteração refere-se ao rompimento do critério de equi-probabilidade, tornando a amostra de elementos não auto-ponderada. Ou seja, a probabilidade de um domicílio e do seu residente não será mais constante (igual entre os setores censitários), pois depende das alterações ocorridas e do número de endereços disponíveis no arquivo do setor censitário na ocasião do sorteio de uma amostra de domicílios durante esse período. Essa variação impede a aplicação dos estimadores simples, usualmente utilizados para calcular erros padrão ou intervalos de confiança (Cochran,² 1977), e obriga o uso de fatores de correção definidos como pesos amostrais (Mohadjer L,⁷ 1999).

No primeiro estágio, sorteou-se uma sub-amostra da amostra de setores censitários da PNAD-95, aceita, convencionalmente, como a amostra definida na primeira etapa desse estágio. Finalmente, no segundo estágio, são sorteados os endereços dentro de cada um dos 72 setores censitários resultantes desta sub-amostra. Como essas amostras são realizadas em diferentes épocas, o processo de estimação deverá considerar o desvio introduzido na suposta equi-probabilidade do modelo pelas alterações nos resultados obtidos para o número de endereços D_j , resultantes da operação de atualização implementada para cada inquérito. Assim, utilizando-se a expressão

$$p_j = \left(\frac{72 \cdot D_j \cdot d}{\sum_j D_j \cdot D_j} \right)^{-1}$$

calcularam-se os pesos (p_j) que compensam esse aspecto do plano de amostragem. Para cada amostra gerou-se, então, o arquivo de pesos amostrais que deverão ser considerados no momento da definição dos estimadores adotados para o processamento dos resultados em cada inquérito (United Nations,¹³ 1993; Dean et al,³ 1994).

A repetição de alguns domicílios entre as amostras de diferentes inquéritos, embora não muito freqüente, pode ser considerada a segunda alteração introduzida nos planos de amostragem. A participação de um mesmo respondente elegível em mais de um inquérito, dentro de um curto espaço de tempo, pode estimular a sua resistência ou recusa em ser entrevistado. Setores com pequenas alterações nos seus tamanhos, durante o período de estudo, e cuja documentação do trabalho de campo estava disponível, foram designados para avaliar a ocorrência de endereços repetidos entre os inquéritos. O endereço sorteado com respondente elegível de um inquérito foi procurado entre os endereços com respondente elegível dos demais inquéritos. A partir da construção da distribuição dos setores segundo o número de ocorrências de repetição, calcularam-se a média e desvio-padrão do número de endereços repetidos entre cada dois inquéritos. Estimou-se, ainda, o número de endereços repetidos para cada mil endereços sorteados em cada inquérito.

As bases de dados dos setores censitários correspondentes às favelas ou áreas invadidas continham informações espaciais e demográficas pouco precisas, o que impediu a elaboração das tabelas de endereços georreferenciados e o sorteio das amostras de domicílios. Visitas a essas áreas possibilitaram a composição de roteiros sobre a base digital cartográfica do Município de São Paulo. Mapas orientaram o percurso dos entrevistadores que, ao mesmo tempo em que registravam a localização do domicílio em formulário próprio, identificavam aqueles sorteados para a composição da amostra e realização das entrevistas. Esse método simplificado de amostragem (Silva & Sundefeld,¹² 1999), testado em áreas urbanas com ocupação desordenada, é eficiente para o sorteio de unidades domiciliares sem referência de endereços, localizados em comunidades com intensa dinâmica migratória.

RESULTADOS

O mapa da Figura 1 apresenta os 72 setores cen-

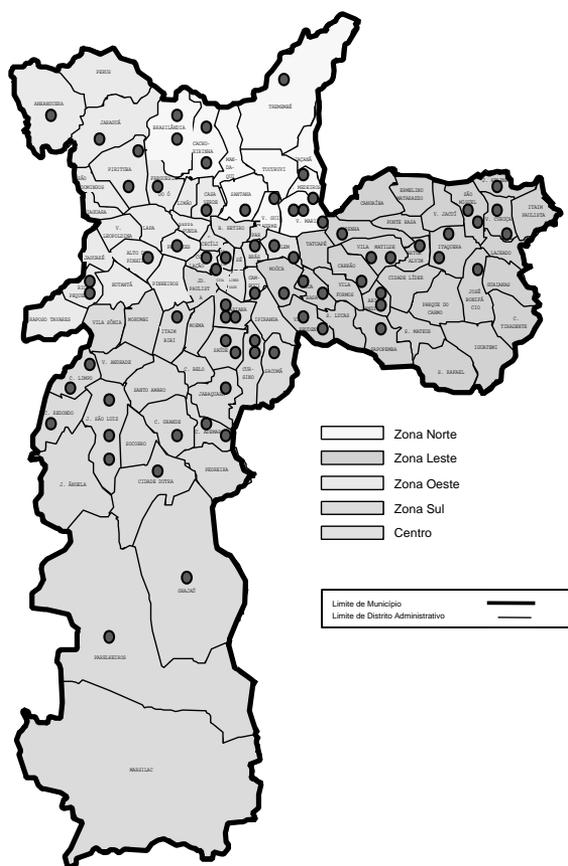


Figura 1 - Amostra mestra, Município de São Paulo (1995-2000). Setores censitários distribuídos entre zonas geográficas.

sitários, distribuídos segundo zonas geográficas do Município de São Paulo. O critério de sorteio, sob probabilidade proporcional ao número de domicílios particulares de cada setor, resultou em maior número de unidades para as áreas mais populosas. Pode-se notar que os distritos com maiores populações possuem maior número de setores na amostra (por exemplo, Vila Mariana e Jardim São Luis na zona sul e Vila Maria na zona norte).

Três inquéritos epidemiológicos (Monteiro,⁸ 1999; Watts et al,¹⁵ 1998; Palloni,⁹ 1997), realizados em domicílios particulares permanentes do Município de São Paulo, utilizaram os cadastros de endereços da amostra mestra durante o período 1995-1999. As características demográficas dos grupos populacionais estudados e as especificidades referentes à execução dos trabalhos de campo induziram à execução de diferentes estratégias de sorteio dos domicílios. Em decorrência da densidade de respondentes elegíveis por domicílio, nem todos os endereços selecionados em cada inquérito foram aproveitados para a realização de entrevistas. Quase 70% dos domicílios sorteados nos dois primeiros inquéritos foram repos-

tos no cadastro, continuando disponíveis para os inquéritos posteriores. O estudo realizado na população de mulheres em idade fértil, que sorteou um número menor de domicílios, disponibilizou cerca de 30% dos endereços sorteados.

As principais saídas impressas do sistema são as listagens das amostras de endereços e seus respectivos mapas (Figuras 2, 3 e 4), que indicam a localização de cada uma dessas sub-amostras nas quadras, facilitando o acesso do entrevistador aos domicílios sorteados em um inquérito específico.

As três últimas colunas da Tabela contêm os valores dos pesos amostrais calculados para as amostras produzidas pelo sistema. A representatividade das entrevistas realizadas em cada inquérito deverá ser corrigida por esses valores quando do cálculo das estimativas e seus erros padrão. Como esperado, a variabilidade destes resultados é menor na zona centro, cujos tamanhos dos setores censitários permaneceram quase inalterados ao longo do período. Inversamente, como conseqüência das alterações nos seus tamanhos, as maiores diferenças ocorreram entre os setores da zona sul. Os valores atípicos obtidos para os pesos do setor 75 do distrito de Artur Alvim, nos dois últimos inquéritos, decorrem de alterações introduzidas pela administração da PNAD-98, que reduziram drasticamente o número de domicílios particulares permanentes registrados no censo demográfico de 1991.

Os principais problemas encontrados durante o processo de geocodificação dos endereços referiram-se às diferentes grafias para os nomes das ruas,

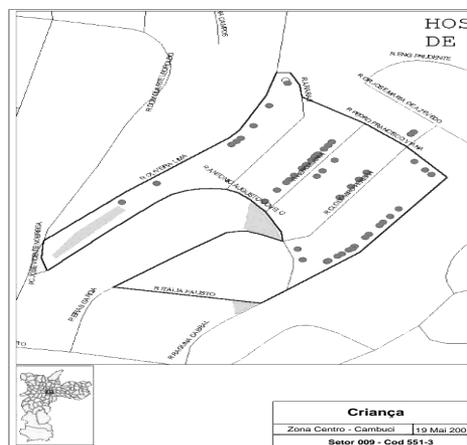


Figura 2 - Município de São Paulo. Cambuci (Setor 11) amostra de domicílios. Crianças.

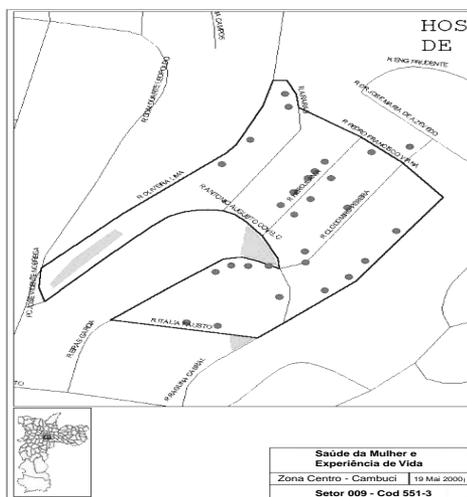


Figura 3 - Município de São Paulo. Cambuci (Setor 11) Amostra de domicílios. Saúde da mulher e experiência de vida.

inconsistências topológicas, e/ou domicílios sem endereços, localizados em favelas ou áreas invadidas. Os domicílios concentrados em um mesmo ponto de endereço (apartamentos, condomínios, vilas) são omitidos no mapa, mas aparecem detalhados na coluna complemento da listagem de endereços das amostras. Dois setores censitários (Campo Limpo-53 e Pirituba-116) eram, respectivamente, favela e área invadida. Visitas *in loco* possibilitaram a construção de mapas atualizados que, embora aproximados, facilitaram a criação do processo diferenciado para o sorteio das sub-amostras de domicílios. O percurso na área, antes impossível de ser realizado com o material disponível, tornou-se orientado pelo mapa construído com auxílio das funções do sistema de informações geográficas.

DISCUSSÃO

Os resultados alcançados mostram a utilidade do conceito de amostra mestra na facilitação da tarefa de obtenção de amostras de domicílios. Três diferentes inquéritos usaram o banco de endereços, sucessivamente atualizado, dos mesmos 72 setores censitários retidos na amostra de unidades primárias do município de São Paulo, durante o período 1995-2000.

A reciclagem das listas de endereços produzidas anualmente pela PNAD é a principal característica dessa idéia, reduzindo os custos referentes à operação de contagem e identificação dos domicílios, necessária à construção dos cadastros sobre os quais

foram sorteadas sub-amostras dentro de cada setor censitário. Visitas *in loco*, visando à recontagem dos domicílios e reconstrução das listagens, foram realizadas em apenas 2 setores. Nos demais, as dificuldades de localização dos domicílios sorteados foram facilmente resolvidas durante a realização das entrevistas. Infelizmente, o uso desse material é possível apenas para os municípios incluídos na amostra da PNAD. Essa é a importante restrição da aplicação do conceito de amostra mestra, como apresentado neste trabalho.

Apesar de usar listagens já atualizadas pela PNAD, o sistema proposto demandou a construção dos arquivos digitais e cartográficos de cada setor censitário. Este é, entretanto, o aspecto de eficiência (Kish,⁶ 1965; Reg,¹⁰ 1999) que fundamenta o conceito de amostra mestra: “a divisão dos custos de tarefas básicas entre os diversos usuários do sistema e a disponibilização ágil de cadastros de endereços geocodificados para o sucessivo sorteio das sub-amostras“. O aperfeiçoamento desses cadastros, pela consecutiva introdução de correções topológicas, de erros de digitação ou de grafia dos endereços durante a operação de listagem, é também outro aspecto da eficiência alcançada pelo sistema proposto.

Os dois principais problemas introduzidos, durante a implementação dos planos de amostragem, foram o afastamento do princípio de equi-probabilidade das amostras e a repetição de domicílios selecionados para entrevistas entre os diversos inquéritos usuários do sistema. A produção de amostras não pro-

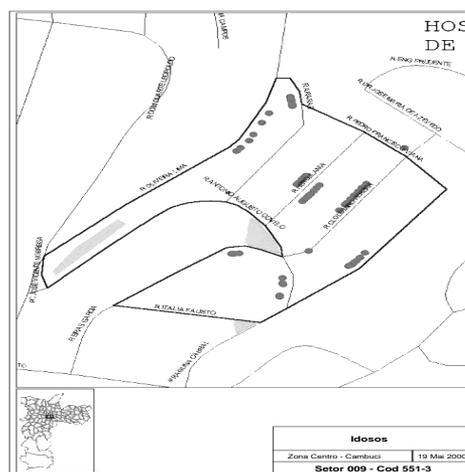


Figura 4 - Município de São Paulo. Cambuci (Setor 11) Amostra de domicílios. Idosos.

Tabela - Amostra mestra: pesos amostrais calculados para os três inquéritos.

Zona	N	Distrito	Setor	Menor 5	Maior 60	Mulheres
Centro	1	Bela Vista	102	637,02	420,18	1260,56
	2	Belém	1	620,34	399,89	1199,71
	3	Belém	7	620,11	417,68	1253,01
	4	Brás	23	706,53	440,94	1322,87
	5	Cambuci	11	618,76	450,39	1351,21
	6	Consolação	94	747,03	499,89	1499,64
	7	República	49	730,49	486,98	1460,98
Leste	8	Água Rasa	42	716,43	445,35	1336,04
	9	Aricanduva	42	615,51	410,34	1231,01
	10	Aricanduva	78	644,70	369,49	1108,44
	11	Artur Alvim	75	750,36	59,712	179,13
	12	Carrão	49	610,12	414,65	1243,90
	13	Itaquera	57	662,86	488,99	1466,98
	14	Itaquera	117	639,92	402,65	1207,94
	15	Jardim Helena	18	681,02	465,59	1396,86
	16	José Bonifácio	37	610,97	407,33	1221,93
	17	Mooca	34	663,61	436,51	1309,47
	18	Penha	99	635,68	423,77	1271,35
	19	São Miguel	29	619,11	348,95	1046,85
	20	São Miguel	94	753,80	728,33	2185,09
	21	Sapopemba	15	700,77	595,09	1785,31
	22	Vila Curuçá	2	749,47	595,99	1787,98
23	Vila Curuçá	60	601,83	403,80	1211,47	
24	Vila Matilde	48	639,57	418,86	1256,56	
25	Vila Matilde	81	653,50	464,21	1392,63	
26	Vila Prudente	8	598,54	387,97	1163,83	
27	Vila Prudente	56	639,52	405,81	1217,39	
Norte	28	Brasilândia	10	589,46	378,15	1134,48
	29	Brasilândia	41	607,07	333,22	999,70
	30	Cachoeirinha	4	601,47	409,87	1229,57
	31	Cachoeirinha	89	690,44	514,28	1542,83
	32	Santana	116	623,58	403,12	1209,37
	33	Tremembé	70	696,07	474,80	1424,51
	34	Vila Guilherme	42	597,16	362,07	1086,21
	35	Vila Maria	8	832,33	560,58	1681,64
	36	Vila Maria	72	689,84	477,31	1431,90
	37	Vila Maria	107	611,88	411,31	1233,93
38	Vila Medeiros	43	596,21	440,43	1321,32	
39	Vila Medeiros	137	627,36	315,53	946,64	
Oeste	40	Alto de Pinheiros	2	628,43	486,07	1458,29
	41	Anhanguera	1	675,97	702,54	2107,76
	42	Casa Verde	32	647,96	416,45	1249,40
	43	Freguesia do Ó	103	608,30	381,12	1143,32
	44	Jaraguá	22	578,69	224,41	673,25
	45	Perdizes	99	735,19	487,67	1463,02
	46	Pirituba	89	609,77	374,79	1124,40
	47	Pirituba	116	769,52	532,02	1596,04
Sul	48	Rio Pequeno	14	660,17	390,48	1171,50
	49	Rio Pequeno	73	677,32	478,08	1434,33
	50	Campo Grande	39	781,97	524,64	1573,97
	51	Campo Grande	63	624,53	382,67	1148,00
	52	Campo Limpo	41	594,16	387,28	1161,87
	53	Campo Limpo	53	830,88	506,18	1518,61
	54	Capão Redondo	101	685,12	500,32	1500,94
	55	Cidade Ademar	38	687,28	458,19	1374,55
	56	Cidade Ademar	147	1104,44	798,33	2395,18
	57	Cidade Dutra	149	651,64	394,31	1182,98
	58	Cursino	13	683,37	491,69	1475,04
	59	Cursino	30	629,11	410,67	1232,00
	60	Grajaú	165	840,64	454,46	1363,44
	61	Itaim Bibi	57	729,10	504,62	1513,88
	62	Jabaquara	88	619,90	406,57	1219,74
63	Jardim São Luís	6	695,29	387,54	1162,62	
64	Jardim São Luís	72	710,22	476,46	1429,32	
65	Jardim São Luís	167	735,19	531,72	1595,26	
66	Parelheiros	32	652,04	582,48	1747,54	
67	Sacomã	159	708,05	503,38	1510,07	
68	Saúde	4	645,90	439,14	1317,46	
69	Saúde	22	602,70	382,42	1147,20	
70	Vila Mariana	36	742,93	484,19	1452,52	
71	Vila Mariana	125	949,93	466,21	1398,62	
72	Vila Mariana	161	766,60	475,86	1427,58	

porcionais, prevista na elaboração teórica do método por conglomerados sob partilha proporcional ao tamanho (PPT), ocorre mesmo em planos "ad hoc" de pesquisas domiciliares, principalmente em países com fortes correntes migratórias. Entre os procedimentos compensatórios previstos, os avanços da tecnologia da informática estimularam o uso corrente da ponderação por pesos amostrais no processamento de dados levantados sob tais planos de amostragem. Reciclar as listagens da PNAD para compor os cadastros atualizados de endereços dos 72 setores censitários, sorteados sob informações geradas no censo demográfico de 1991, é uma idéia apoiada nesta técnica. Assim, as diferenças introduzidas nas frações de amostragem, entre os domicílios de diferentes unidades primárias, foram facilmente superadas pela disponibilização de arquivos contendo os pesos amostrais referentes ao plano operacional de cada inquérito.

Apesar de ser uma questão intrínseca ao conceito de amostra mestra, a ocorrência de domicílios repetidos entre as três amostras pode ser considerada dentro dos níveis aceitáveis. Os resultados mais desfavoráveis registram a ocorrência média de um domicílio repetido por setor censitário, ou três repetições para cada mil endereços sorteados entre os dois últimos inquéritos. Mesmo assim, é importante registrar relatos dos entrevistadores quanto ao desconforto ocorrido na abordagem em um domicílio já entrevistado em inquéritos anteriores, e considerar a importância do estudo de mecanismos para impedir essa ocorrência, especialmente quando houver coincidência na época do trabalho de campo de diferentes inquéritos. Restrições teóricas e operacionais, apresentadas por consultor técnico sobre o assunto (Verma,¹⁴ 2000), desaconselham a

simples eliminação dos endereços anteriormente sorteados e sugere que seja preservada a natureza probabilística de todas as amostras geradas a partir dos cadastros da amostra mestra.

A produção de bases digitais para os arquivos de endereços e a aplicação de tecnologia de geoprocessamento, para a geração automática e mapeamento de amostras de domicílios, significam o aspecto inovador desta experiência, em acréscimo ao conceito de amostra mestra. Além de superar as dificuldades geradas pela quase inexistência de bases cartográficas digitais disponíveis ao usuário brasileiro, destaca-se a sua importância na criação de soluções cartográficas para os setores censitários cujo material disponível era inaproveitável.

Os cadastros dos 72 setores censitários da amostra mestra reuniam cerca de 25.000 domicílios. Durante o período de desenvolvimento deste estudo, apenas três inquéritos usaram esses cadastros, e o número de domicílios ainda disponíveis (cerca de 70%) sugere que o sistema tem capacidade de cobrir um número bem maior de usuários. Entretanto, o controle dos tamanhos dos setores e das respectivas frações de amostragem é um importante aspecto a ser considerado em cada etapa de decisão.

Os resultados alcançados, em especial a produção do material informativo para a realização ágil e econômica do trabalho de campo, valida a relevância de investimentos na construção de infraestrutura técnica e capacitação de pessoal, visando ao funcionamento contínuo do sistema como recurso de apoio aos diversos inquéritos domiciliares que venham a ser realizados no Município de São Paulo durante os períodos inter-censitários.

REFERÊNCIAS

1. [BEMFAM] Sociedade Civil Bem-Estar Familiar no Brasil. *Pesquisa nacional sobre demografia e saúde, 1996*. Rio de Janeiro; 1997.
2. Cochram WG. *Sampling techniques*. New York: John Wiley & Sons; 1977.
3. Dean AG, Dean JA, Columbier D, Brendel KA, Smith DC, Burton AH et al. CSAMPLE: Análise de dados provenientes de planos complexos de amostragem. In: *Epi-Info 6.04*. Atlanta: CDC; 1994. cap. 14.
4. Fundação Instituto Brasileiro de Geografia e Estatística (IBGE). *Para compreender a PNAD (um texto simplificado)*. Rio de Janeiro; 1991.
5. King AJ, Jessen RJ. The master sample of agriculture. *JASA* 1945;40:38-56.
6. Kish L. *Survey sampling*. New York: John Wiley & Sons; 1965.

7. Mohadjer L. *Sample weights for households with multiphase data collection approaches*. Proceedings of the 52nd ISI Session, August 10-18, 1999; Helsinki; Finland.
8. Monteiro CA. Saúde infantil: tendências e determinantes na cidade de São Paulo na segunda metade do século XX. *Rev Saúde Pública* 2000;34(6 Supl):1-4.
9. Palloni A. *Health conditions of the elderly in Latin America and the Caribbean*. Washington (DC): Pan American Health Organization;1987.
10. Reg G. The master for household surveys. *Stats News, [online] available from* <http://196.25.65.2/statsnews/May99/mastersample.htm> [15 oct 2000].
11. Sidindi M, Laldaparsad S. GIS a major boost in planning census 2001. *Stats News, [online]*. Available from <http://196.25.65.2/statsnews/statsnews/gis.htm> [15 oct 2000].
12. Silva NN, Sundefeld MLM. Simplified sampling plans for epidemiological surveys in metropolitan areas in Brazil. *Surv Sta* 1999;40:13-8.
13. United Nations. Statistical Division. *Sampling errors in household surveys*. New York; 1993.
14. Verma V. Question/answer. *Surv Stat* 2000;42:17-8.
15. Watts C, Heise L, Ellberg M, Willians L, Moreno CG. *WHO multi-contry study of women's health and domestic violence, core protocole*. Geneva: WHO; 1998.