Maria Cecilia Goi Porto Alves[I]

Maria de Lima Salum e Morais[I]

Maria Mercedes Loureiro Escuder[I]

Moisés Goldbaum[II]

Marilisa Berti de Azevedo Barros[III]

Chester Luiz Galvão Cesar[IV]

Luana Carandina[V]

Household sampling in slums in surveys

ABSTRACT

OBJECTIVE: To identify the advantages and disadvantages of using segments compared to a complete address list, for the selection of households in a multistage cluster sampling in slums.

METHODOLOGICAL PROCEDURES: A qualitative study was performed in four slums selected by the São Paulo Municipal Health Survey of 2008, and the two selection techniques were applied. Focal groups were performed with field researchers, including the persons making the list of addresses and the interviewers. The content of the conversations were analyzed, grouped in categories and organized in themes.

ANALYSIS OF RESULTS: Use of household segments was associated with several advantages and few disadvantages. The advantages included: speed and facility in developing the sampling frame and in locating and identifying households when performing interviews, increased safety for interviewers and the population, greater access to interviewees, greater stability and coverage of the frame, and fewer errors in the identification of selected households.

CONCLUSIONS: The construction of a household registry by creation of segments is advantageous compared to the complete listing of addresses, when undertaken in slums. Due to its economy and ease, the technique is an option for simplifying the sampling process in areas characterized by high density and disorganized housing.

DESCRIPTORS: Sampling Studies. Health Surveys, methods. Data Collection. Statistical Methods and Procedures. Poverty Areas. Técnicas de pesquisa. Favelas. Investigative Techniques. Cluster sampling.

[I] Instituto de Saúde. Secretaria de Estado da Saúde de São Paulo (SES-SP). São Paulo, SP, Brasil

[II] Departamento de Medicina Preventiva e Social. Faculdade de Medicina. Universidade de São Paulo (USP). São Paulo, SP, Brasil

[III] Departamento de Medicina Social. Faculdade de Ciências Médicas. Universidade Estadual de Campinas. Campinas, SP, Brasil

[IV] Departamento de Epidemiologia. Faculdade de Saúde Pública. USP. São Paulo, SP, Brasil

[V] Departamento de Saúde Pública. Faculdade de Medicina de Botucatu. Universidade Estadual Paulista Júlio de Mesquita Filho. Botucatu, SP, Brasil

Correspondence:
Maria Cecília Goi Porto Alves
Instituto de Saúde – SES-SP
R. Santo Antônio, 590 – 5º andar
01314-000 São Paulo, SP, Brasil
E-mail: cecilia@isaude.sp.gov.br

INTRODUCTION

Household surveys are increasingly utilized in the planning and evaluation of health policy.[4] Probabilistic sampling is used because it allows for precision-based inferences.

In commonly performed multistage sampling, the sample is selected from the previously selected sampling units. This strategy presupposes the existence of complete lists for sampling units throughout all stages of selection. Therefore, in a sample where households are selected within census tracts,[a] it is necessary to have a list of all households in the tracts.

Besides making the selection feasible, the list should be organized so that the selected households are correctly identified by interviewers during data

---

[a] Census tract is a territorial unit established by the IBGE (Brazilian Institute of Geography and Statistics) in order to control listing, formed by a continuous area, situated within an urban or rural space, with the dimension and number of houses, which allows for data collection by another census worker.

collection and by field supervisors. The technique commonly used to develop this tool consists of developing a complete list of households in the census tracts while traveling through all streets (address gathering). Nonetheless, an alternative exists in the literature which consists in dividing the tracts into segments that will be subsequently selected as a group of houses.[8] It is possible to select compact segments, in which all households are included in the sample or only part of the households are included.[8] The address list instead becomes a list of segments. The choice between these two possibilities depends on factors that may conflict and that do not allow for one approach to be considered superior in all situations.

In geographic areas where housing patterns are disorganized, dense and difficult to identify, such as slums,[b] the task of creating and using an address list of households presents difficulties. Given this context, it is important to explore alternatives to simplify the process of probabilistic sampling, while meeting methodological requirements.

The World Health Organization supports the adoption of simplified methods that facilitate the performance of periodic studies in developing countries.[2] The frequency of slums in medium and large municipalities in Brazil, combined with the difficulties experienced by field researchers in surveys, reinforce the importance of evaluating sampling procedures in order to adopt the most simple method.

The objective of this article was to identify the advantages and disadvantages of the use of segments for selection of households through multistage cluster sampling in slums in comparison to selection using a complete address list.

METHODOLOGICAL PROCEDURES

The techniques for list development – complete listing of addresses and segments – were applied in the Health Survey of São Paulo Municipality (*Inquérito de Saúde do Município de São Paulo* – ISA-Capital), implemented during the second semester of 2008 and first semester of 2009. This study was performed in four slums selected in the survey, which are located in the districts of Rio Pequeno, Cidade Ademar, Sapopemba and Capão Redondo (174, 369, 328 and 128 households, respectively).

The techniques were evaluated at two moments: during the development of the address list and in the implementation of interviews at the selected households.

The two approaches for developing the list were performed in the four slums. Selection of households was performed using one listing approach in each slum: complete address lists in Sapopemba and Capão Redondo and segments lists in Rio Pequeno and Cidade Ademar. The enrollment of addresses in each slum was performed twice so that each team could experience both household listing techniques. Nonetheless, when performing the interviews, only one of the lists was used by each team in order to avoid unplanned interviews.

Multistage cluster sampling was used to obtain the sample for ISA-Capital 2008,[c] and the first stage was the census tract. The starting point used for the selection in this stage was the 2005 National Household Sample Survey (*Pesquisa Nacional por Amostra de Domicílios*, PNAD), which sampled 264 urban census tracts in the municipality that were not included in the strata of new constructions. From these, 70 tracts were selected for the sample of the 2008 ISA-Capital. The sample initially consisted of 60 census tracts, including the four slums of this study. Ten additional tracts were on reserve in case smaller than expected samples were obtained.

In the second stage, 90 households were selected in each tract with an intention to perform five interviews in children less than one year of age, since this is the less frequent group in ISA-Capital. The sample fractions are presented in the Annex.

For selection using the complete address list, all the addresses in a tract were listed in a custom spreadsheet that registered the block and side of the block. Blocks were identified in relation to the map of the area. Addresses were numbered, excluding uninhabitable structures (under construction or demolition) and non-residential structures, and the 90 households in each slum were selected by systematic sampling.

For household selection by segments, the segments were created using clear, identifiable and stable limits.[8] These limits should be the reference points for the slum. Segments should have ten households, but the size varied due to the location of reference points.

The segments were listed on a specific form with identification numbers that corresponded to the reference points established and with the number of households. The Figure 1 shows the layout of segments in Rio Pequeno slum.
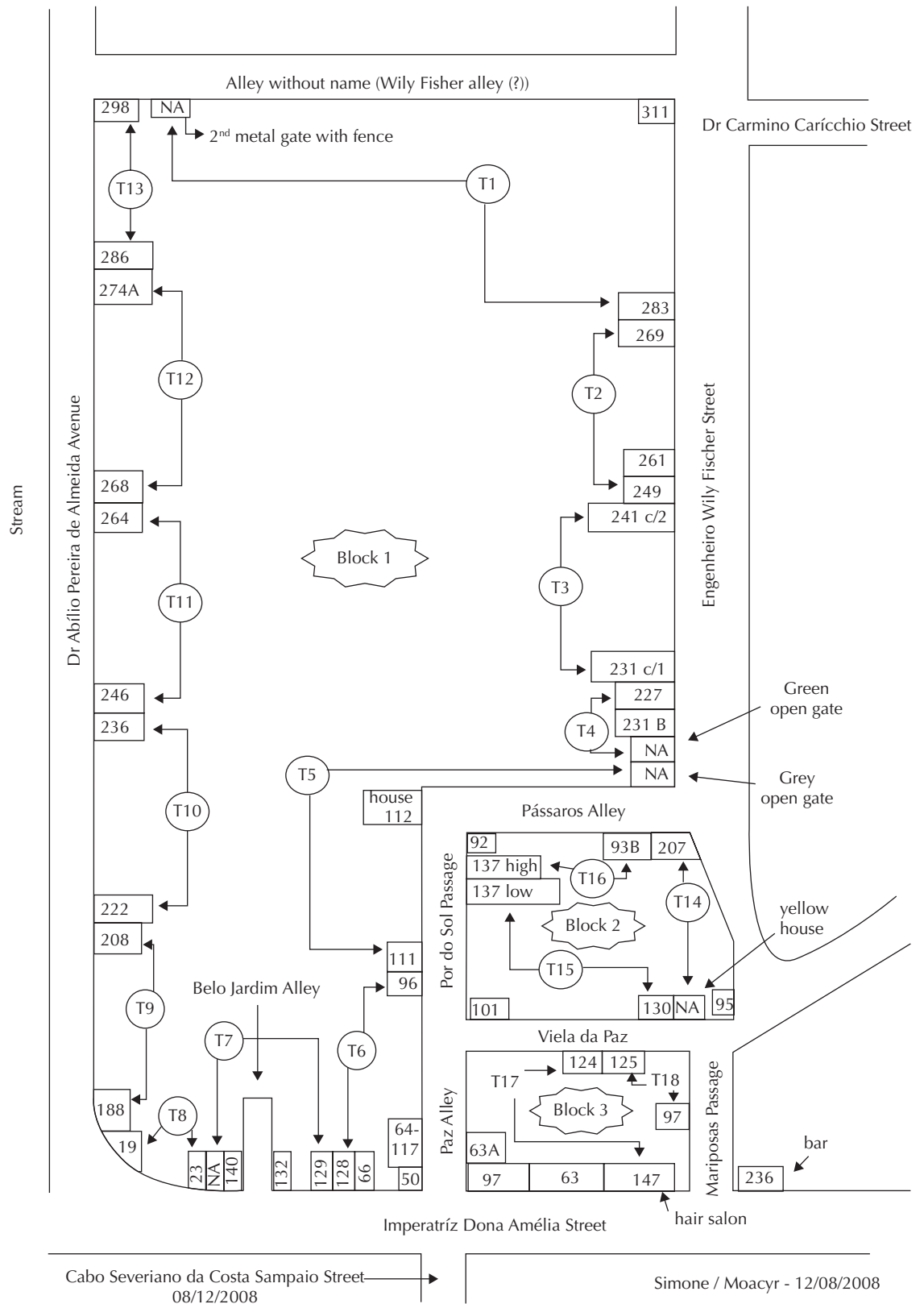

Figure 1. Segments of houses created in São Domingos favela

In Rio Pequeno and Cideade Ademar, 18 and 35 segments were respectively created. Of these 53 segments, 27 had 10 households, eight were smaller and 18 larger. Nine segments were randomly selected, for the 90 households sampled per slum.

The two sampling frame techniques, complete list of address and creation/listing of segments, are referred to in this article as traditional listing and segment listing. The techniques were evaluated through a qualitative study with focus groups,[10] including field researchers (the people making the list of addresses and interviewers) participating in the 2008 ISA-Capital.

Two focus groups sessions were conducted in September and December of 2008. The first session sought to discuss the creation of the sampling frame with the two techniques. The duration was approximately 1h30min. Two moderators and two pairs of people that make address lists participated, once the field researchers completed this stage working in pairs. The pairs used both strategies to create the sampling frame in each slum, i.e. they repeated the procedure to create the list. A succinct script sought to explore the advantages and disadvantages of the two strategies and issues that were raised during the discussion.

The second session lasted approximately 40 minutes and sought to compare the perceptions of focus group members about the context in which the interviews occurred when utilizing the two sampling techniques. The participants included two moderators and two pairs of interviewers who used one of the techniques for household interviews. The focus group comments regarding the comparison of techniques were based on previous experience, obtained by participation in other household surveys.

The focus group sessions were recorded and transcribed. The conversations underwent content analysis[3] and were grouped in categories, organized into themes that demonstrated the opinions of participants. The following themes were identified: general characteristics of field work in slums; advantages and disadvantages in implementation, map development, and performance of interviews.

Homogeneity of the segments was assessed by estimating the intraclass correlation coefficient (Rho), which allows for evaluation of intraclass homogeneity and interclass heterogeneity. The estimates obtained should approximate zero for a small loss of precision due to cluster sampling.

Kish[8] defines Rho, for a population divided into A clusters with B elements, as: $Rho = \dfrac{\sigma_a^2 - \sigma_b^2/(B-1)}{\sigma^2}$

where σ_a^2 is the variance of the means of A clusters, σ_b^2 is the variation of the elements of the cluster and σ^2 is the total variance.

The Rho estimates were obtained with the "loneway" command in Stata, which produces similar results for the intraclass correlation as the estimate proposed by Kish for the above formula.[d]

The intraclass correlation coefficients estimates were constructed considering the sampling strategy and the estimation for dichotomous variables describing morbidity (health problem in the past 15 days, allergy, spinal disease, depression and migraines) and the use of services (dentist visit in the past 12 months), in addition to sociodemographic variables (health insurance and paid employment).

The study was developed in accordance with resolution 196/96 by the National Health Council about human research ethics (process approved by the Ethics Committee of São Paulo Municipality, No. 0003.0.162.000-08). Participants signed voluntary informed consent forms, agreeing to participate in the study.

RESULTS

Due to unregulated groups of houses located on public property, slums are noticeably different than structured urban neighborhoods and involve specific issues in study implementation.
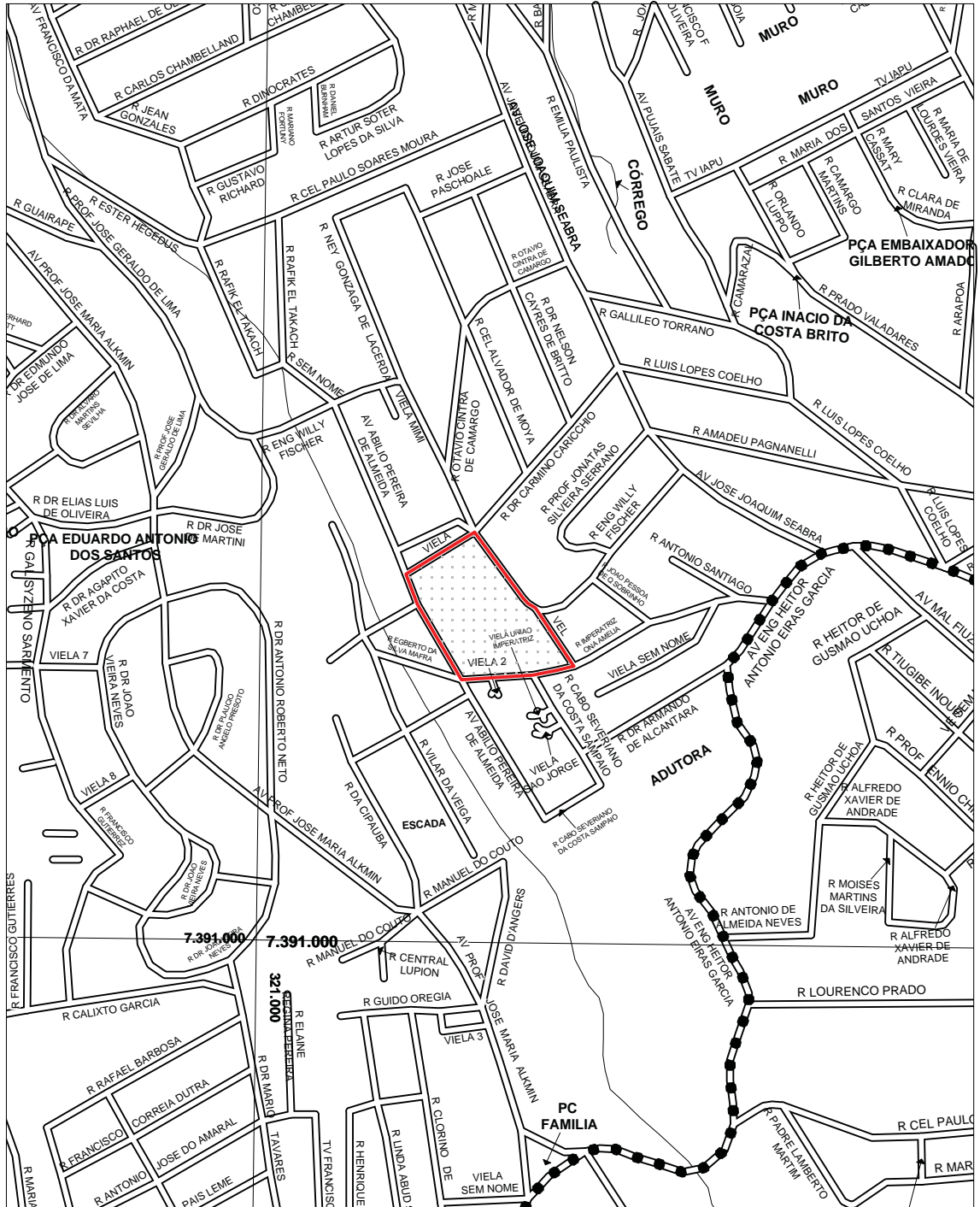
The Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, IBGE) only circumscribes an area with demarcated boundaries, as illustrated in Figure 2. Therefore, researchers encounter unknown circumstances when first entering a slum. It is necessary to map its streets, passageways, alleys, narrow paths and houses. Many slums are located in precarious areas, such as hillsides, marshlands and flood zones, which complicates localization and access of households.

The streets and alleys do not have names or have several names. Houses do not have numbers or the numbers are randomly assigned by residents. Therefore, the people who generate the address lists create a unique nomenclature to identify and locate them.

The areas are constantly changing: houses appear and disappear in days; residents constantly change the colors of their houses and doors; they build new houses behind existing ones; the cover streams with wood planks and other materials to create new passageways.

[d] Ferreira MP. Homogeneidade intraclasse: estudo empírico sobre setores censitários como unidades primárias de amostragem [Masters dissertation]. São Paulo: Schol of Public Health, USP; 2003.

Figura 2. Contour of the sector according to Instituto Brasileiro de Geografia e Estatística (Brazilian Institute of Geography and Statistics).

"A painter lived there. When we returned for the second visit, we noticed he had transformed the alley, by not only painting but also demolishing two houses." "

Houses are similar and therefore difficult to identify during interviews.

"We were creating a traditional list, and I began to include characteristics: metal fence with a blue sign, metal fence and the given color. Everything was the same. If somebody paints the fence, it becomes another house. This irreparably ended the listing."

The interviewers mentioned the lack of trust by residents. Since they occupy non-legal areas, there is a constant fear of eviction. The lack of trust and fear depends on the specific moment. There is greater resistance during tense periods following violence or if people are in hiding. Resistance is lower when conducting visits with the municipal government to improve living conditions.

"They are afraid that we are from the municipal government and will evict them…they are obviously under pressure the whole time. They are illegal and display the pressure."

"…on Saturday, the team went to conduct the interviews and the police arrived. We left. The following day, we returned and the people were very apprehensive. This difficulty exists: if something happened, if police are circulating, any suspicion…"

The interviewers also perceive a lack of safety.

"Many people are in there hiding, not even the police go…"

"A boy alerted me: 'miss, do not enter back there. There is a group, and we smelled marijuana.' I returned, waited for the other girls to finish their interviews, and we went there."

The speed of developing the address list was one of the advantages to listing by segments.

"The list creation by segments in slums is very good because it is much faster…"

According to one of the pairs, the length of time to develop the list in a large slum was 3h31min with the segment lists and was 5h28min by household; in a small slum, listing by segment took 1h55min and listing by household lasted 3h3min. The other pair did not calculate the time with such precision but reported that listing by segment required half as much time as listing by household. In a large slum, listing required two and four hours respectively, while in a small slum one and two hours were required.

The interviewers evaluated the development of maps using the two techniques. With the traditional approach,

the creation of the map is secondary, since the written notes are most important. In listing by segments, the map should be more detailed than the notes.

"In the written section, you describe the beginning and end, and the map is much clearer than the notes. While with the traditional approach, you depend much more on the notes."

"When listing by segment, what takes longest is creating the map."

The interviewers report greater ease and speed in localizing the households when performing interviews with segment lists, since it is easier to identify segments than households and the approach is less susceptible to changes in the slum.

"With the segment, no matter how much a street or house characteristic changes, you know where the previous section ends and begins."

"The probability that your segment reference is demolished or disappears is small…"

In addition, the proximity of the interviewers during the interviews provided increased safety to the team and facilitated access to residents, who perceived field team working.

"Since it was an area of increased vulnerability, it was important for the team to be closer, with one member supporting the other; segment lists also facilitate things in this sense."

"When residents see a large team, they feel safer. They are really here. Somebody from the health sector is here. If a neighbor has doubts and sees another person participating, she also participates. It calls more attention when implementing by segment. Everyone is there, and it is easier to find people. People are at their doors."

In participant discussions, it was possible to identify the implicit mention of other advantages or confirmation of the above advantages through the description of the disadvantages of the traditional strategy. The traditional strategy was considered as requiring more work, more time and more susceptible to incorrect identification of selected households. This strategy may be more affected by changes to the details of houses and the difficulty to distinguish the differences between them.

"When you will make a traditional address list, you include all the details. It is very vulnerable; things change. The first time we went to a yellow house with a brown gate. The second time it was painted blue with a brown gate, and that was the description we had. There are no address numbers."

"I think…that there is an additional difficulty in distinguishing the houses one by one, which has not been mentioned. There are slums that are very characteristic of slums, and it becomes difficult to describe the differences: all the houses are shacks, they are identical."

Another aspect mentioned as an advantage of listing by segment was the possibility to more easily deal with households invisible at first that may not be listed.

"Traditional listing involves a lot of work and can cause errors because you do not have access past the gates, and when you do see, you realize that things are different."

"You describe the first and last house, and everything is between. It is much quicker."

There were few disadvantages described in regards to listing by segment. The interviewers mainly discussed their fear that the selection of segments may concentrate on homogenous households within a larger variability. For example, the exterior sides of most slums have more comfortable houses, sidewalks and other improvements in comparison to the internal part.

"You would have to be careful when selecting the segments in order to not concentrate only on the outside or inside."

"In the slums I went to in Cubatão, there was a very big difference between the houses… Towards the end there were poor shacks… wood houses on top of sewage. If you use segments in this area, there will be differences."

Disadvantages to listing by segments were indirectly identified by researchers when they described the advantages of listing by households as greater details and possibly greater precision.

"Traditional listing has the advantage of detail. If you want something much more detailed, then you have to do the traditional listing."

"The greatest advantage that I saw is that it has the possibility of a more precise result. In the slums, there are poorer zones, less safe zones. Sampling by household has a more spread out panorama."

The Table presents prevalence estimates for variables obtained in the survey and the corresponding intraclass correlation coefficient estimates. The values were near zero, indicating little intra-segment homogeneity.

DISCUSSION

This study indicates the superiority of the sampling strategy utilizing segments of households, since the strategy was associated with many advantages and few disadvantages. The principal advantages concerned the speed and ease of the listing, the localization and identification of households during performance of interviews, increased safety for interviewers and the population,

greater access to the interviewees and greater stability of the lists. The disadvantages reported were the potential homogeneity within segments and the possibility of less detailed information on the population.

Sampling of household segments has been utilized in various health surveys.[6,9,11,14,15] Although the utilization of this strategy in slums is not unprecedented, those who used it see the strategy as an innovation that modifies the work process.[12] Subjective and personal aspects decisively influence the evaluation. According to Trindade,[13] one of the aspects that most influences the adoption of an innovation, independent of the context in which it is proposed, is the perception that the new technique or process is an improvement. This evaluation has a major subjective component. Therefore, a qualitative methodology was chosen for this study.

The focus group strategy was chosen due to the participants' vast experience in field work. This prior experience allowed for the perception of even subtle differences between the two strategies.

Simplicity was one of the most emphasized qualities of the sampling strategy with segments of households. The relevance of this aspect in sampling plans can be understood by the large efforts undertaken in various countries for alternatives that simplify the process of obtaining samples in household surveys.[1,5,7,14]

Simplification of the sampling process involves making it more agile, cheaper and easier to administer, without decreasing the accuracy of the results.[2] The present study found that simplicity was associated with sampling by segments: the length of time to perform interviews was less, reducing transportation and supervisions costs, and the selected households were located with greater ease.

Use of segments can decrease sampling errors due to problems related to the identification of selected households and allow for the inclusion of households that were not previously listed. On the other hand, the concentration of houses in segments could reduce

Table. Prevalence and intra-segment correlation coefficients for variables. São Paulo, Southeastern Brazil, 2008.

Variable	%	rho
Presented a health problem (15 days)	21.2	0.0000
Has health insurance	24.2	0.0124
Reports arterial hypertension	15.2	0.0252
Reports allergy	13.6	0.0000
Reports back pain	10.3	0.0000
Reports depression	21.2	0.0223
Reports migrain	19.7	0.0000
Had a dentist consultation (12 months)	45.5	0.0000
Employed outside the home	42.9	0.0390

Source: São Paulo Municipality Health Survey (ISA) Capital 2008.

the precision of estimates by increasing variance. Households that are individually selected from address lists and are spread throughout the tract constitute sub-samples with less homogeneity. The estimates of the intraclass correlation coefficient were near zero, indicating that the effect of clustering can be discounted.

High intraclass correlation coefficients indicate homogeneity within segments as well as heterogeneity between them. The small values obtained for this indicator reflect the similarity between the residents of the two slums in regards to the dimensions measured. Nonetheless, the preoccupation of focus group participants with this aspect demonstrates a refined understanding of sampling issues.

One of the advantages identified in listing by segments concerns coverage, also identified in the literature. According to Kish,[8] some people who conduct sampling believe that address lists tend to hurriedly prepared, possibly omitting households. Segments can be more completely covered by field workers, and completeness can be more easily verified.

The other advantage of segments identified by the literature is greater stability.[8] In slums abrupt changes can occur between the period of list creation and interview administration, and greater stability in the sampling units can avoid the introduction of errors. This study mostly based segment boundaries on reference points that were not houses, such as geographical points and street characteristics that were unlikely to change.

The complete listing of addresses provides more detailed information, which is supported by the literature.[8] Households can be classified during list creation and the variables subsequently used to assist in sampling or characteristics of residents, such as age, sex and occupation, can be ascertained and used for two stage sampling.

Social interaction (contagion) between neighbors can occur during the study period. According to Kish,[8] some researchers fear increased refusal rates and other contamination, although there is a lack of evidence to support this supposition. This study viewed the interaction as positive since it facilitated contact with the population. Greater facility in reaching the population may increase the rate of agreement to answer the questionnaire.

The sense of safety that the researchers felt from circulating in a group in an environment they consider unsafe was emphasized. Although field work was organized in groups to allow for the presence of several interviewers in the same slum, the use of segments concentrated the interviews and increased proximity, which facilitated communication between researchers and rapid localization by field supervisors.

Although the sampling of segments has been used in diverse health surveys, the results of this study demonstrate that this type of sampling is particularly advantageous when applied in slums. Sampling by segments is an economical and easy strategy to simplify the sampling process in areas characterized by disorganized and dense housing patterns.

REFERENCES

1. Alves MCGP, Silva NN. Simplificação do método de estimação da densidade larvária de *Aedes aegypti* no Estado de São Paulo. *Rev Saude Publica.* 2001;35(5):467-73. DOI: 10.1590/S0034-89102001000500010

2. Anker M. Epidemiological and statistical methods for rapid health assessment: introduction. *World Health Sta Q.* 1991;44(3):94-7.

3. Bardin L. Análise de conteúdo. Lisboa: Edições 70; 1977.

4. Barros MBA. Inquéritos domiciliares de saúde: potencialidades e desafios. *Rev Bras Epidemiol.* 2008;11(Supl 1):20-37. DOI:10.1590/S1415-790X2008000500002

5. Bennet S. A simplified general method for cluster - sample surveys of health in developing countries. *World Health Sta Q.* 1991;44(3):98-106.

6. Caraveo-Anduaga JJ, Bermúdez EC. The epidemiology of obsessive-compulsive disorder in México City. *Salud Mental.* 2004;27(2):1-6 .

7. Herderson RH & Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bull World Health Organ.* 1982;60(2):253-60.

8. Kish L. Survey sampling. New York: John Wiley & Sons; 1965.

9. Kuper H, Polack S, Limburg H. Rapid assessment of avoidable blindness. *Community Eye Health J.* 2006;19(60):68-9.

10. Morgan D. Focus group as qualitative research. London: Sage Publications; 1997. (Qualitative Research Methods Series, 16).

11. Passos MC, Lamounier JA, Silva CAM, Freitas SN, Baudson MFR. Práticas de amamentação no município de Ouro Preto, MG, Brasil. *Rev Saude Publica.* 2000;34(6):617-22. DOI:10.1590/S0034-89102000000600009

12. Rogers EM, Takegami S, Yin J. Lessons learned about technology transfer. *Technovation.* 2001;21(4):253-61 . DOI:10.1016/S0166-4972(00)00039-0

13. Trindade E. A incorporação de novas tecnologias nos serviços de saúde: o desafio da análise dos fatores em jogo. *Cad Saude Publica.* 2008;24(5):951-64. DOI:10.1590/S0102-311X2008000500002

14. Turner AG, Magnani RJ, Shuaib M. A not quite as quick but much cleaner alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design. *Int J Epidemiol.* 1996;25(1):198-203 . DOI:10.1093/ije/25.1.198

15. White LL, Golberg HI, Ballew C, Mendlein JM, Peter DG, Percy CA, et al. Rationale, design and methodology for the Navajo Health and Nutrition Survey. *J Nutr. 1997;*127(10 Suppl):2078-84S.

Annex. Sampling Fractions

The sampling fraction for the sampling of census tracts was: $f_1 = \dfrac{a_1 M_i}{M} \cdot \dfrac{70}{a_1} = \dfrac{70 M_i}{M}$, where a_1 equals the number of tracts sampled by the PNAD with a probability proportional to size, Mi equal to the number of households in tract i attributed by the 2000 Census and M the total the number of households. Simple random sampling was used for the 70 tracts for ISA-Capital 2008.

In the following stages, the sampling fractions for each of the strategies presented in the study were:

1st strategy – complete address lists

There were b households selected with a probability $f_2 = \dfrac{b}{M_i}$, where b had values of 90, 13, 30, 13 and 45 for the respective age groups of less than one year, children, adolescents, adults and older people. With this number of households, we expected to perform 5, 18, 15, 13 and 15 interviews in each tract with the population groups of interest. The overall sampling fraction was: $f = \dfrac{70 M_i}{M} \dfrac{b}{M_i}$.

2nd strategy– segments of households

For the group age less than one year, 9 segments of 10 households were selected each with a probability $f_2 = \dfrac{9}{M'_i / 10} = \dfrac{90}{M'_i}$, where M'_i equals the current number of households in tract i and M'_{ii} equals the number of households in segment ij. All households encountered in the selected samples were included in the sample. Therefore, the overall sampling fraction was: $f = \dfrac{70 M_i}{M} \cdot \dfrac{90}{M'_i} \cdot \dfrac{M_{ij}(M'_i / M_i)}{M_{ij}} = \dfrac{6300}{M}$.

For all the other households, sampling was performed in three stages and the sampling fractions for the second and third stages were: $f_2 = \dfrac{cM_{ij}}{M'_i / 10}$ and $f_3 = \dfrac{d}{M_{ij}}$, where c equals the number of segments per tract and d equals the number of households per segment. The c values for the groups of children, adolescents, adults and older people were 4, 4, 4, and 6 and for d they were 3.25, 7.5, 3.25 and 7.5. The overall sampling fraction was: $f = \dfrac{70 M_i}{M} \cdot \dfrac{cM_{ij}}{M'_i} \cdot \dfrac{d(M'_i / M_i)}{M_{ij}} = \dfrac{70cd}{M}$.