Cartographic and population of census tracts use to design a hantavirus prevalence study.

Roberto de Melo Dusi* (University of Brasília (UnB), Brasilia, Brazil, robertodusi@hotmail.com), Raynner Augusto Moreira Parente (Superior Health Science School, Brasilia, Brazil, raynnerparente@hotmail.com), Angelika Bredt (Health Secretariat of Federal District (SES-DF), Brasilia, Brazil, angelika.bredt@gmail.com), Maria Isabel Rao Bofill (SES-DF, Brasilia, Brazil, bofillrao@ig.com.br), Mariana Fehr Nicácio (UnB, Brasilia, Brazil, mari.fehr@hotmail.com) Pedro Luiz Tauil (UnB, Brasilia, Brazil, pltauil@unb.br)

Abstract

Hantaviruses is an acute viral anthropozoonosis transmitted by aerosol inhalation or contact with wild rodents feces and urine. The human infection may be asymptomatic or manifest many clinical profiles, including severe outcomes. Fatality rate reached up to 100% of the confirmed cases, while the lowest fatality rate was never lower than 20%. Identifying infected persons, regardless they have symptoms, or even are lightly affected, allows the infectious disease prevalence measure. This measure makes possible to understand the dispersion intensity, regarding time and place. As it is not possible to test a 2.5 million population, using a sampling technique enables this purpose. Randomization and representativeness are emphasized, and logistical aspects need to be considered. This is a study with complex sampling design type, applied to health sector, for human population, specifically to obtain the population-based prevalence of hantavirus human infection, at Brazilian Federal District (DF). Among the 4,358 census tracts in DF, delimited as polygons by 2010 Brazilian national census, were selected those who possessed probable transmission locations (PTL) of hantavirus. The sample was composed of two sub-samples: the first, most important, randomic and representative; the second, conditional and non-probabilistic. The planning included high level refuse, because the study needed human blood sample. It was listed 81 (1.9%) DF polygons with PTL. The prevalence estimated was small and put the event as rare. For representative and randomic sampling the authors chose complex sampling, combining the steps for systematic sampling and simple sampling. The selection interval was calculated. As the sampling was complex and multi-step, the design effect value was high, in a conservative manner. Applying 95% confidence interval and others criteria, the sample size calculated was of 495 individuals, with 20% increase added in the sample size, resulting in 594 households to draw. Two samples were designed for this study. High value design effect was used. Feasible, representative and random sample was obtained. For non-probabilistic sampling the authors chose convenience sampling

Key words: probabilistic sampling; population-based; complex sampling; cross-sectional

1. Introduction

Hantaviruses is an acute viral anthropozoonosis which etiological agent is Hantavirus, genus of Bunyaviridae family. Hantaviruses are most commonly transmitted by aerosol inhalation or contact from secretions, urine and feces of wild rodents of the Muridae family, making it part of robovirus group. There are also indirect mode of transmission via the natural environmental (Kallio *et al.*, 2006), even person-to-person transmission (Padula *et al.*, 1998). The human infection may be asymptomatic or manifest many clinical profiles, including severe outcomes. There are two clinical forms, the Hemorrhagic Fever with Renal Syndrome (HFRS) (Sheedy *et al.*, 1954), typical in Europe and Asia,

and the hantavirus cardiopulmonary syndrome (SCPH) (Duchin *et al.*, 1994), detected mainly in the Americas (Manigold & Vial, 2014).

In 2014, Ebola disease surprised the health scenario with a sudden increase in number of cases at many places, simultaneously, unlike its previous focal distribution (Arwady *et al.* 2014). Although hantaviruses and Ebola disease have many differences, before 2014 both disease have focal distribution. New researches on hantaviruses in areas with high frequency may provide us conditions to detect epidemiological changes. The improvements can be applied for other focal diseases.

A first outbreak of SCPH form occurs in Brazilian Federal District (DF) in 2004 (Brasil, 2005). The fatality rate reached up to 100% of the confirmed cases at some places of Brazil, while the global FR was 39,3% from 1995 to 2006 (Elkhoury *et al.*, 2012). Although the American hantavirus' syndrome is very lethal, the number of cases is small. In contrast, the Eurasian's form of the disease has relatively low fatality rate, but with a very large number of cases and complications commonly occur. The hantavirus' transmission occurs mainly in rural areas (Manigold & Vial, 2014). The occurrence of hantaviruses at DF was relatively high in 2004 (Brasil, 2005). As described in other areas, as United States of America (USA), hantaviruses distribution of confirmed cases by age was asymmetric, with rare cases in people under 10 years old (Jonsson *et al.*, 2010).

In areas with high seroprevalence, a high proportion of hantavirus infections may be asymptomatic or have only mild symptoms (Manigold & Vial, 2014). This means that many infected people do not know they have been biologically 'invaded' by hantavirus. Identifying infected persons, regardless they have symptoms, or even are lightly affected, allows the infectious disease prevalence measure. This measure makes possible to understand the dispersion intensity, regarding time and place. These information can help to predict the future hantaviruses outbreaks (Pereira, 1995).

For diseases as hantaviruses, the prevalence in a given area for a certain period can be measured as antibodies remain detectable by blood test. As it is not possible to test a 2.5 million population, using a sampling technique enables this purpose. Randomization and representativeness are emphasized, and logistical aspects need to be considered. The randomization achieved through sampling techniques and the representativeness by sample size calculations (Silva, 2004). The aim of this study is to apply a random and a representative sampling methods for human hantavirus infection prevalence in the DF.

2. Section 2 – Methods

This is a study with complex sampling design type, applied to the health sector, for human population, specifically to obtain population-based prevalence of human hantaviruses. Among the 4,358 census tracts in the DF, delimited as polygons by 2010 Brazilian national census (IBGE, 2013), were selected those who possessed probable transmission locations (PTL) of hantavirus, according to the previous records of confirmed hantaviruses human cases. A confirmed case in DF means that the patient met the definition of a suspected case and had a confirmatory laboratory test. Confirmatory hantaviruses tests are the specific enzyme immunoassay (EIA) or molecular biology tests in tissue obtained from corpses.

The PTL for hantaviruses are geographic points considered exposure areas reported by confirmed cases or their 'proxies', verified by the specialized environmental health surveillance, and classified as receptive for transmission. The PTL were then marked with geographical positioning apparatus and the coordinates were recorded in UTM (Universal Transverse Mercator). These routine activities of health surveillance seeks to identify the sources of exposure to protect people with the same practices as the patients. Census tracts polygons were selected for each PTL. All cases with PTL at DF, regardless their home address, generated one or more census tracts. These sectors were included

in the study area, whether they are adjacent or not. For this step, the geographical IBGE (Brazilian Geographic and Statistics Institute) data base, available at internet, including census tracts mesh (IBGE, 2013) and public domain software Google Earth™ (Google, 2013), using coordinates in degrees, minutes and seconds were used.

3. Section 3 - Sample, Sample size and sampling technique

For the sampling, it was set that the participant selected should have been living in the sector for at least six months, whether continuously or not. The participant should be living, at least, at select census tract six month and should be older than 9 years, since there were no reported cases for children under this age at DF.

The sample was composed of two sub-samples (Bolfarine & Bussab, 2005): the first, primordial, with random and representative characteristics was composed by people drawn according to the above criteria; the second, conditional, non-probabilistic, convenient, consisted of people with the same age and presence in the polygons mentioned above, chosen by the researchers according to their exposure, recorded over the steps for the first sample. The second sample would be incorporated if the event was classified as rare in DF(frequency less than 10%) (Cochran, 1977).

The sample size was obtained using all: the population living in the study area older than 9 years, the level of confidence, the hypothetical frequency of the phenomenon (hantaviruses), the confidence limit and the design effect (Dean *et al.*, 2013). Also, a percentage as operational replacement was added, as the study includes blood test, inhibiting people participation. The population was obtained from the sum of all groups of people older than 9 years, from all selected polygons. The confidence level was set at 95%. The confidence limit was arbitrated according to the value of the hypothetical frequency. The design effect was quantified according to the sampling technique and its stages. Operating replacement was based on the opinion of experts who work on research projects with human blood sampling.

The hypothetical frequency was drawn up on the cumulative incidence of confirmed cases of hantaviruses at DF, from 2004 to 2012. Considering that the cases reported were obtained from epidemiological surveillance data, and were mainly serious or fatal cases, and other seraprevalence studies of viral infections reported a seraprevalence 17-folds the number of patients known (Vasconcelos *et al.*, 1998) and as the hantavirus infection prevalence at DF is not available, researchers arbitrarily designated the prevalence (P) 10 times higher than the incidence (I). As the list of each polygon inhabitants is not available, the simple sampling technique cannot be used. Therefore, seeking weighted territorial representation, the researchers opted for systematic sampling, selection interval was calculated (Silva, 2004) and assigned a number of inhabitants per census tract.

4. Section 4- Results

We found 81 (1.9%) DF polygons with PTL. The authors grouped these sectors as rural or urban (fig. 1). These sections were distributed in 14 (73.7%) of 19 geographic DF districts.

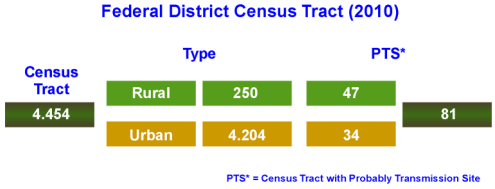

Figure 1: Distribution of census tracts in the DF according to the presence of PTL and type.

These census tracts inhabitants, according to the 2010 census, were more than 42 thousand inhabitants. Those with 10 years or more numbered 34,838 inhabitants (Fig. 2).

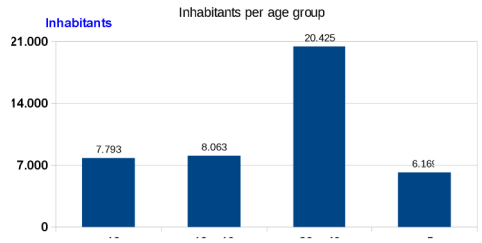

Figure 2: Age distribution of the population of 81 census tracts with cases of hantavirus.

The hypothetical frequency used the cumulative incidence of cases in the study area, from 2004 to 2012, that reached 0.25%. Applying multiple adjustment of 10, we estimated the prevalence of hantavirus infection by 2.4% in the population belonging to the 81 sectors of the DF region. For this prevalence, the confidence limit was arbitrated in 1.94%. Since it is then considered a rare event, other 240 non-probabilistic intentional samples were added to the study.

With the choice for systematic sampling technique, households random selected by weighting each polygon and raffling one resident per household, accumulated stages and conditions that led researchers to a conservative approach, and used high design effect of 2.0. Applying the 95% confidence interval, sample size calculated was 495 individuals (Bolfarine & Bussab, 2005).

Considering that the project includes the collection of human blood samples for laboratory tests, a 20% increase was added in the sample size, resulting in 594 households to draw, as a corrected sample size (css) (Fig. 3).

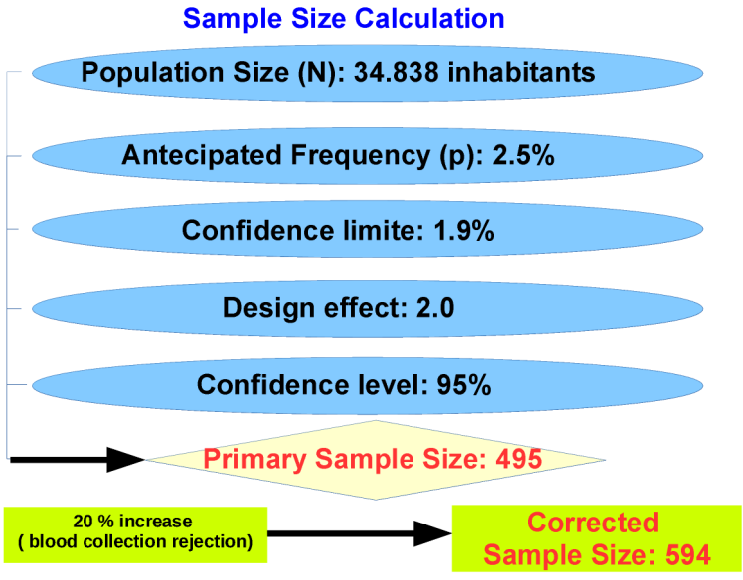

Figure 3: Sample calculation (Dean *et al.*, 2013).

Dividing the number of study population (N), by sample size (n), the researchers obtained the selection interval, that may be calculated for systematic sampling. When the researchers arrive to the drawn household, they may prepare a list of electable inhabitant and one may be selected randomly.

Limitations:

The disease transmission can affect inhabitants and visitors of polygons with PTL. For probabilistic sampling it is unfeasible to include visitors randomly, as eligible population. PTL was not identified for nine cases, despite evidence of having occurred inside DF, Brazil.

5. Conclusion

Two samples were designed for this study. For representative and randomic sampling the authors chose complex sampling, combining the steps for systematic sampling and simple sampling. High value design effect was used. Feasible, representative and random sample was obtained. The prevalence estimated was small and put the event as rare. For non-probabilistic sampling the authors chose convenience sampling.

Acknowledgments

The authors are grateful Professor Lucio Vivaldi, Professor Claudete Ruas and Professor Pedro Sadi Monteiro.

References

Arwady, M.A., Bawo, L., Hunter, J.C., Massaquoi, M., Matanock, A., Dahn, B., Ayscue, P., Nyenswah, T., Forrester, J.D., Hensley, L.E., Monroe, B., Schoepp, R.J., Chen, T., Schaecher, K.E., George, T., Rouse, E., Scafer, I.J., Pillai, S.K., & Cock K.M.D. (2014). Evolution of ebola virus disease from exotic infection to global health priority, Liberia, mid-2015. Emer. Inf. Dis. 21(4):578-584

Bolfarine, H., & Bussab, W. (2005). Elementos de Amostragem. São Paulo, Brazil, Ed. Blucher.

Brasil, Ministério da Saúde. (2005). Surto da Síndrome Cardio-Pulmonar por Hantavírus no Distrito Federal e Goiás. Bol. Elet. Epid. 05(1):1-5.

Cochran W.G. (1977). Sampling Techniques. New York, USA, John Wiley & Sons.

Dean, A.G., Sullivan, K.M. & Soe, M.M. (2013) OpenEpi: Open Source Epidemiologic Statistics for Public Health, Available in: <www.openepi.com>

Duchin, J.S., Koster, F.T., Peters, C.J., Simpson, G.L., Tempest, B., Zaki, S.R., Ksiazek, T.G., Rollin, P.E., Nichol, S., Umland, E.t., Moolenaar, R.L., Reef, S.E., Nolte, K.B., Gallaher, M.M., Butler, J.C., Breiman R.F., & Hantavirus Study Group. (1994) Hantavirus Pulmonary Syndrome: a Clinical Description of 17 Patients with a Newly Recognized Disease. N Engl J Med. 330:949-55

Elkhoury M.R., Mendes, W.S., Waldman, E.A., Dias, J.P., Carmo, E.H., & Vasconcelos P.F.C. (2012). Hantavirus pulmonary syndrome: prognostic factors for death in reported cases in Brazil. Trans. R. Soc. Trop. Med. Hyg. 106:298–302.

Google Earth. DF images 2012, Brasília - DF. Available in: <http://www.google earth.com.br/>. Access in 2012, June 25th.

IBGE :: Instituto Brasileiro de Geografia e Estatística. Available in: <http://www.ibge.gov.br/home/download/estatistica.shtm>.

Jonsson, C.B., Figueiredo, L.T.M., & Vapalahti, O. (2010) A Global Perspective on Hantavirus Ecology, Epidemiology, and Disease. Clin Microbiol Rev., 23(2): 412-441.

Kallio, E.R., Klingstrom, J., Gustafsson, E., Manni, T., Vaheri, A., Henttonen, H., Vapalahti, O. & Lundkvist A. (2006). Prolonged survival of Puumala hantavirus outside the hosts:evidence for indirect transmission via the environmental. J Gen Virol 87:2127-2134.

Manigold, T. & Vial, P. (2014) Human hantavirus infections: epidemiology, clinical features, pathogenesis and immunology. Swiss Med Wkly. 144:w13937.

Padula, P.J., Edelstein, A., Miguel, S.D.L., López, N.M., Rossi, C.M. & Rabinovith, R.D. (1998). Hantavirus pulmonary syndrome outbreak in Argentina: molecular evidence for parerson-to-person transmission of Andes virus. Virology 241:323-330.

Pereira, M.G. (1995) Epidemiologia Teoria e Prática. Rio de Janeiro, Brasil, Guanabara Koogan.

Sheedy, J.A., Froeb, H.F., Batson, H., Conley, C.C., Murphy, J.P., Hunter, R.B., Cugell, D.W., Gilles, R.B., Bershadsky, S.C., Vester, J.W., & Yoe, R.H. (1954). The clinical course of epidemic hemorrhagic fever. Am J Med.16(5):619-628.

Silva, N. N. (2004). Amostragem Probabilística – Um Curso Introdutório. São Paulo, Brasil, Edusp.

Vasconcelos, P.F.C., Lima, J.W.O., Travassos da Rosa, A.P.A., Timbó M.J., Travassos da Rosas, E.S., Lima, H.R., Rodrigues S.G. & Travassos da Rosa, J.F.S. Epidemia de Dengue em Fortaleza, Ceará: inquérito soroepidemiológico aleatório. Rev. Saúde Pública, 32(5): 447-54, 1998.