Probability sampling design in ethnobotanical surveys of medicinal plants

Mariano Martinez Espinosa,[1] Isanete G. C. Bieski,[2] Domingos Tabajara de Oliveira Martins,[2]*

[1]*Departamento de Estatística, Instituto de Ciências Exatas e da Terra, Universidade Federal do Mato Grosso, Brazil,*
[2]*Departamento de Ciências Básicas em Saúde, Área de Farmacologia, Faculdade de Medicina, Universidade Federal de Mato Grosso, Brazil.*

Article

Abstract: Non-probability sampling design can be used in ethnobotanical surveys of medicinal plants. However, this method does not allow statistical inferences to be made from the data generated. The aim of this paper is to present a probability sampling design that is applicable in ethnobotanical studies of medicinal plants. The sampling design employed in the research titled "Ethnobotanical knowledge of medicinal plants used by traditional communities of Nossa Senhora Aparecida do Chumbo district (NSACD), Poconé, Mato Grosso, Brazil" was used as a case study. Probability sampling methods (simple random and stratified sampling) were used in this study. In order to determine the sample size, the following data were considered: population size (N) of 1179 families; confidence coefficient, 95%; sample error (d), 0.05; and a proportion (p), 0.5. The application of this sampling method resulted in a sample size (n) of at least 290 families in the district. The present study concludes that probability sampling methods necessarily have to be employed in ethnobotanical studies of medicinal plants, particularly where statistical inferences have to be made using data obtained. This can be achieved by applying different existing probability sampling methods, or better still, a combination of such methods.

Introduction

Since time immemorial, man has been using plants in different forms for the prevention and treatment of diseases. In addition, medicinal plants have been the main and most accessible form of therapy for the less privileged in the global population up to the present day. The World Health Organization, through its recommendations to developing nations, has been a formidable driving force in the popularisation of herbal medicine, particularly concerning scientific validation and preservation of ethnobotanical knowledge of medicinal plants (Ministério da Saúde, 2005).

Selection of medicinal plants with a view to discovering new pharmaceutical agents can be achieved by various means, but plant selection based on its popular use is by far the most promising by virtue of decreased time and cheaper cost incurred in the collection of this information (Albuquerque & Hanazaki, 2006). Towards this end, probability (Almeida & Albuquerque, 2002; Andrade, 2002; Pechansky et al., 2004; Poss et al., 2005; Cavéchia & Proença, 2007) and non-probability (Tomazzoni, 2004; Tomazzoni et al., 2006; Viganó et al., 2007; Jesus et al., 2009; Silva et al., 2010a) sampling methods are often used.

A non-probability sampling method implies the absence of any probabilistic mechanism in the sample selection, while probability sampling denotes the involvement of a probabilistic procedure in the sample selection (Bolfarine & Bussab, 2005), the latter being more appropriate when one wishes to ensure representativeness of the sample and greater robustness (Albuquerque et al., 2010b).

In ethnobotanical surveys of medicinal plants, the most commonly employed non-probability sampling methods are sampling by convenience and snowball sampling (Andrade, 2002; Pechansky et al., 2004; Poss et al., 2005; Cavéchia & Proença, 2007).

Convenience sampling is a sampling method in which units are selected based on easy access or availability. This method of sampling is generally less time-consuming and of lower cost. For this reason it is widely used despite its serious limitations, such as sources of bias in the selection of sample units. Thus, samples obtained by convenience are not suitable for research involving inferences about populations, but may be used in exploratory research (Mattar, 1996).

Snowball sampling is a technique that is often

utilised for intentional selection of expert informants. In this method, the first contact with the community may be through a well-known expert, who then indicates another expert, and so on, until all the specialists in the community are covered. It may happen that an expert may not indicate another name for various reasons; if the researcher is aware of such a situation, the process may need to be repeated all over again (Albuquerque et al., 2010). It is noteworthy that convenience and snowball samples may not be representative for the purposes of making statistical inferences about the population.

In ethnobotanical surveys of medicinal plants, the most commonly used probability sampling methods are simple random, stratified, cluster and systematic probability sampling methods, or a combination of various complex sampling methods (Bolfarine & Bussab, 2005; Scheaffer et al., 2006; Silva et al., 2010).

Simple random sampling is a process whereby a sample size n of a random variable X, with a given distribution is selected from a finite population of N units, in such a way that each possible sample is given the same probability of being selected and all units of the population have the same probability of being included in the sample (Scheaffer et al., 2006; Levy & Lemeshow, 2008).

Simple random sampling is more representative provided that the elements of the population are homogenous, otherwise the population under study may have to be divided into groups or strata of elements that are more homogenous. In this way, variability within groups is smaller than within the total population, and in addition present higher representativeness at a lower cost. This process is defined as stratified random sampling (Bolfarine & Bussab, 2005; Scheaffer et al., 2006).

Stratified random sampling involves division of elements of the population under study into groups called strata, based on certain characteristics of interest in the population under study. Random samples are then selected from each stratum proportional to the stratum's size (Bolfarine & Bussab, 2005; Scheaffer et al., 2006).

When the unit of the population consists of a cluster or group of elements, then cluster sampling can be used. Cluster sampling is an efficient sampling technique designed to obtain information from a given population at low cost, either when there is no frame of reference or it is too difficult to have one, or when there is no register that lists the elements of the population, as well as when the cost to obtain samples increases with the distance separating the sampling units (Scheaffer et al., 2006; Levy & Lemeshow, 2008).

On the other hand, systematic sampling is a method of selecting sample members from a larger population according to a random starting point and a fixed, periodic interval. It is conducted by randomly selecting one unit from the first k units in the sampling frame and then every k^{th} unit. Where, $k = N / n$, N is the population size and n is the desired sample size. This technique is generally used for populations which have elements arranged according to a reference system (Scheaffer et al., 2006).

Considering that the use of a non-probability sampling method does not guarantee that the sample is representative and does not enable generalisations regarding the population under study to be made (Kish, 1965), it is, therefore, not suitable for research involving statistical inferences on the population (Bolfarine & Bussab, 2005). An alternative to this limitation of non-probability sampling in studies involving statistical inference is the use of probability sampling.

The present study aims to present a probabilistic sampling design for ethnobotanical studies of medicinal plants, using the Pantanal region of Nossa Senhora Aparecida do Chumbo district (NSACD), in Poconé municipal, Mato Grosso, Brazil as a case study.

Materials and Methods

Area under study

The study was conducted in the micro-areas of the communities of NSACD. This district is 30 km from the Poconé headquarters and 94.8 km from Cuiabá. NSACD consists of 37 communities, distributed in 16 micro-areas, totalling 3652 individuals belonging to 1179 families, as presented in Table 1. Micro-areas were considered instead of communities, because of official reasons: micro-area is an official term employed by the Ministry of Health to improve logistics in order to facilitate provision of health-related services to the families in these communities who will be attending local Family Health Units in the district (Poconé, 2008).

The population studied consisted of an individual (informant) per household, aged at least 40-years-old that had resided in NSACD for at least five years. We considered the family and not an individual in the sampling design for logistical reasons, and also considering that the district is a rural area. Moreover, in this setting, it is easier to identify an informant within the family.

Sampling design

A probabilistic approach was employed considering simple and stratified random sampling methods (Levy & Lemeshow, 2008) using as a case study the ethnobotanical survey of medicinal plants conducted at NSACD, in 2009 (Bieski et al., 2012).

Given the demographic details of the families

Table 1. Distribution of 37 communities of Nossa Senhora Aparecida do Chumbo district (NSACD), by micro-areas, with total number of individuals and families in each micro-area, including sample fraction and sample size, Poconé, Mato Grosso, Brazil, 2009.

Communities	MA	Total of individuals	Total of families	Sample fractions	Expected Sample sizes	Sample sizes obtained
Chumbo	1	311	71	0.0602	17	16
	2	397	78	0.0662	19	18
	3	238	67	0.0568	16	16
Canto do Agostinho, Santa Helena, Os Cagados, Várzea bonita	4	179	52	0.0441	15	10
Furnas II, Salobra, Zé Alves	5	165	59	0.0500	15	10
Campina II, Furnas I, Mundo Novo, Rodeio	6	279	81	0.0687	20	11
Campina de Pedra, Imbé	7	188	67	0.0568	16	12
Barreirinho, Coetinho, Figueira	8	253	95	0.0806	23	23
Bahia de Campo	9	257	74	0.0628	18	13
Agrovila, São Benedito	10	184	66	0.0560	16	16
Agroana	11	186	81	0.0687	20	19
	12	186	97	0.0823	24	19
Bandeira, Minadouro	13	248	82	0.0696	20	22
Carretão, Deus Ajuda, Sangradouro, Pesqueiro, Varzearia	14	216	77	0.0653	19	23
Chafariz, Ramos, Sete Porcos, Urubamba	15	208	67	0.0568	16	16
Céu Azul, Capão Verde, Morro Cortado, Passagem de Carro, Varal	16	157	65	0.0551	16	18
N		3652	1179	1	290	262

MA-Micro-area, Bieski (2010).

studied, we were able to determine the sample size using a simple random sampling method. Towards this end, given a specified proportion (*p*), statistical expression (1) was employed to determine the number of families. Generally, most of the variables in this type of study are qualitative in nature, thus statistical tests of association such as chi-square test (x^2), prevalence and tests of proportions are frequently used.

In order to determine approximate sample size (*n*) for each micro area, in this case the number of families randomly selected, as well as to estimate the value of *p* with confidence interval of , expression (1) was used according to Bolfarine & Bussab (2005) and Levy & Lemeshow (2008):

$$n = \frac{Np(1-p)}{(N-1)(d/z_{\alpha/2})^2 + p(1-p)}, \qquad \textbf{(1)}$$

Where:
n = approximate sample size;
N = population size of all the 37 communities of NSACD;
p = proportion of population considered, to which is assigned the value of 0.5;
d = limit to error estimate;
α = level of significance considered;
$Z_{\alpha/2}$ = value obtained from the table of Standard Normal Distribution (Scheaffer et al., 2006).

It should be emphasised that if the population size is not known or if it is too large, equation (2) can be used instead of equation (1) to estimate the value of *p* with level of significance 1-α (Levy & Lemeshow, 2008):

$$n = \frac{(z_{\alpha/2})^2 \, p(1-p)}{d^2} \qquad (2)$$

Where:
n = approximate sample size;
p = proportion of population considered;
d = limit to error estimate;
α = level of significance considered;
$Z_{\alpha/2}$ = value obtained from the table of Standard Normal Distribution (Scheaffer et al., 2006).

Once the approximate sample size was obtained, we proceeded with the determination of the number of families in each micro-area, which in this study is referred to as the stratum. It should be noted that instead of micro-areas, the strata could also be regions. Thus, a random stratified sample was obtained by separation of the elements of the population into groups, called strata, and then subsequent selection of a random sample from each layer.

Consequently, if the population is divided into *k* strata, a stratified random sample of size (*n*) will consist of *K* simple random samples of size *nk*. The

sample size (*n*) was obtained by adding the sample sizes in each stratum given by the expression (3) (Bolfarine & Bussab, 2005):

$$n = n_1 + n_2 + \ldots + n_K = \sum_{i=1}^{K} n_k \qquad (3)$$

Where:
n = approximate total sample size;
nk = approximate sample size in each stratum.

If the population in each stratum (micro-area) is given by N_k, then the population size (*N*) is given by the sum of the size from each stratum, which can be obtained by the expression (4):

$$N = N_1 + N_2 + \ldots + N_K = \sum_{i=1}^{K} N_k \qquad (4)$$

The sample fractions of the elements (w_k) contained in each stratum are obtained by the expression (5):

$$w_k = \frac{n_k}{n} \qquad (5)$$

Where:
n = approximate sample size;
n_k = approximate sample size in each stratum;
w_k = sample fractions of the elements.

Thus, the sample size (n_k) in each stratum is given by the expression (6):

$$n_k = w_k \times n, \qquad (6)$$

For *k* = 1,2,...,K.
Where:
n = approximate sample size;
n_k = approximate sample size in each stratum;
w_k = sample fractions of the elements.

Results

To illustrate this sampling procedure, we considered the data obtained in the ethnobotanical study that took place in NSACD (Bieski et al., 2012), presented in Table 1, of 1179 families distributed in 37 communities (Figure 1).

It is worth mentioning that, using the expression (1), a sample size of 290 or more families ($n \geq 290$) was obtained for the 16 micro-areas. Therefore, with a sample of at least 290 individuals from 290 families randomly selected, it is expected that 95% confidence

intervals estimated with a semi-amplitude of 0.05, will contain the true frequencies of the determined percentages. The value of *p*=0.5 was assigned due to lack of information of this value from ethnobotanical surveys of medicinal plants conducted in Mato Grosso. However, only 262 individuals were interviewed.

To determine the approximate sample size in each micro-area, we used expression (5) and (6), corresponding to the number of families and sample fraction, respectively.

Population fraction was first determined for each micro-area, by dividing the number of families in each micro-area with the total number of families of all the micro-areas (*N*=1179) as shown in Table 1. Thereafter, the sample size in each micro-area was obtained using expression (6), by multiplying the population fraction of each micro-area by 290, resulting in the sample size for each micro-area. However, during the application process of the data capture instrument, twenty eight (28) individuals were not present at their dwellings resulting in a sample loss of 9.7%.

For selection of families in each micro-area, a systematic random sampling was carried out to obtain a more representative sample using MINITAB, version 15.0 (Minitab, 2007) statistical package. Other statistical packages that have routines for random selection of numbers may also be employed.

Finally, it is worth pointing out that the techniques presented in this study can also be applied to other methods of probability sampling.
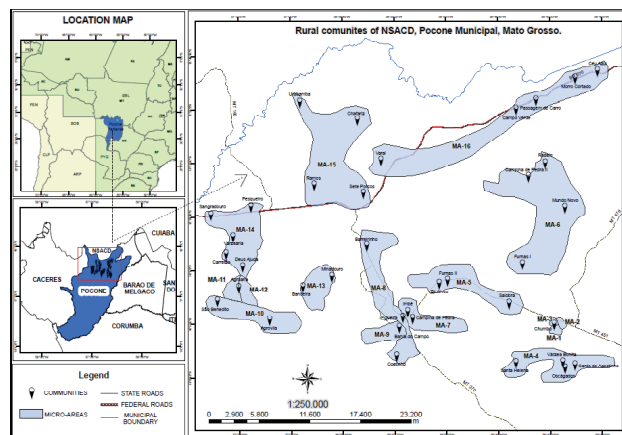
Figure 1. Location of the study area, Nossa Senhora Aparecida do Chumbo district (NSACD), Poconé-MT, Brazil, Bieski (2010).

Discussion

Sampling by convenience is less time-consuming and least expensive when compared with other sampling techniques. Despite these advantages, this sampling method has serious limitations. It presents many potential sources of bias in the selection

process, which includes among others, self-selection of respondents (Malhotra, 2004). According to Churchill (1998) another problem of sampling by convenience is that there is no way of knowing if all people in the sample are representative of the population and that the sampling method is less reliable because the researcher selects the sample due to convenience, with little rigor in the selection process.

The main advantage of snowball sampling is that it increases the possibility of getting the desired information in the population (Malhotra, 2004). Although, this technique is used in ethnobotanical surveys of medicinal plants (Andrade, 2002; Cavéchia & Proença, 2007), its use does not guarantee that the sample will be representative of the population being studied (Kish, 1965).

It should also be noted that although non-probability methods are commonly used in ethnobotanical surveys of medicinal plants, the results of any survey based on this type of sampling do not allow for generalisations of the population under study to be made. This is because the selection of each element depends on the judgment of the researcher, the sampling is therefore not randomised (Kish, 1965), thus not appropriate for research involving inferences about populations. However, they may be used in exploratory research to generate ideas and working hypotheses.

According to Malhotra (2004), probabilistic sampling techniques vary in terms of efficiencies of sampling. Sampling efficiency is a concept that reflects a trade-off between cost and accuracy of the sampling method. Precision refers to the level of uncertainty in the characteristic being measured or counted. The greater the precision, the higher may be the cost. The efficiency of a probabilistic sampling technique can be assessed by comparing it with simple random sampling (Silva, 2001).

In simple random sampling, all units have an equal chance of being sampled and all samples have an equal probability of being selected. Therefore, simple random sampling tends to produce representative samples and this procedure has greater validity when elements of the population are homogeneous (Scheaffer et al., 2006; Levy & Lemeshow, 2008).

If the population is not homogeneous then it is better to opt for stratified random sampling, which is done by separating the elements of the population into groups, called strata, and then subsequent selection of random sample from each stratum (Silva, 2001; Bolfarine & Bussab, 2005; Scheaffer et al., 2006).

The main reasons for using stratified random sampling instead of simple random sampling is to increase the accuracy and representativeness at a lower cost and to obtain estimates of population parameters with the population subgroups (strata).

The choice between the types of non-probability and probability sampling should be based on considerations such as the nature of the study, the variability within the population and also by statistical considerations. For example, in exploratory research, non-probability sampling can be used. On the other hand, in conclusive research where the researcher intends to use the results to estimate parameters, probability sampling is preferable (Malhotra, 2004).

Non-probability sampling has been widely used in ethnobotanical surveys of medicinal plants, however, due to its limitations it should be used only in situations where a statistical inference of the population is not intended.

Despite the high acceptability of the instrument and the sampling methods utilised, there was a loss of 9.7% of the samples due to the absence of a few of the respondents in their homes at the time of application of the questionnaire. According to Kish (1965) and Silva (2001) a loss of ≥15% of sample units compromises inferences about the data obtained. However, since the percentage loss recorded in this study was only 9.7%, it may be affirmed that it has no effect on the data analysis. This clearly demonstrates the applicability of using probability sampling.

In this study, simple random and stratified random probability sampling methods were employed, which allowed us to determine the precise size of the samples in each micro-area considered in the research, as well as helping to reduce the variance of the estimators.

Consequently, it is concluded that probability sampling needs to be employed if statistical inferences are part of the intended goals of the research.

Acknowledgments

References

Albuquerque UP, Hanazaki N 2006. As pesquisas etnodirigidas

na descoberta de novos fármacos de interesse médico e farmacêutico: fragilidades e perspectivas. *Rev Bras Farmacogn 16*: 678-689.

Albuquerque UP, Lucena RFP, Cunha LVFC 2010a. Métodos e Técnicas na Pesquisa Etnobiológica e Etnoecológica. In: Albuquerque UP, Lucena RFP, Alencar NL (org.). *Métodos e técnicas para coleta de dados etnobiológicos*. Recife: NUPEEA, p. 41-64.

Albuquerque UP, Lucena RFP, Cunha LVFC 2010b. Métodos e Técnicas na Pesquisa Etnobiológica e Etnoecológica. In: Albuquerque UP, Lucena RFP, Lins Neto EF (org.). *Seleção dos Participantes da Pesquisa*. Recife: NUPEEA, p. 21-38.

Almeida CFCBR, Albuquerque UP 2002. Uso e conservação de plantas e animais medicinais no estado de Pernambuco (nordeste do Brasil): um estudo de caso. *Interciencia 27*: 276-285.

Andrade CT 2002. *Um estudo etnobotância da conexão homem/Cactaceae do semi-árido Baiano*. Bahia, 102 p. Dissertação de mestrado, Programa de Pós-graduação em Botânica, Universidade Estadual de Feira de Santana.

Bieski IGC, Rios Santos F, Oliveira RM, Espinosa MM, Macedo M, Albuquerque UP, Martins DTO 2012. Ethnopharmacology of medicinal plants of the Pantanal region (Mato Grosso, Brazil). *Evid Based Complement Alternat Med 2012*: 1-36.

Bieski IGC 2010. *Conhecimento etnofarmacobotânico de plantas medicinais utilizadas por comunidades tradicionais do Distrito Nossa Senhora Aparecida Chumbo, Poconé, Mato Grosso, Brasil*. Cuiabá, 269 p. Dissertação de Mestrado, Programa de Pós-graduação em Ciências da Saúde, Universidade Federal de Mato Grosso.

Bolfarine H, Bussab WO 2005. *Elementos de amostragem*. São Paulo: Editora Edgar Blücher.

Cavéchia LA, Proença CEB 2007. Resgate cultural de plantas de usos de plantas nativas do Cerrado pela população tradicional da região do atual Distrito Federal. Brasília. *Heringeriana 1*: 11-24.

Churchill G 1998. *Marketing research: methodological foundations*. Orlando: The Dryden Press.

Jesus NZT, Lima JCS, Silva RM, Espinosa MM, Martins DTO 2009. Levantamento etnobotânico de plantas popularmente utilizadas como antiúlceras e antiinflamatórias pela comunidade de Pirizal, Nossa Senhora do Livramento-MT, Brasil. *Rev Bras Farmacogn 19*: 130-139.

Kish L 1965. *Survey sampling*. New York: Wiley.

Levy PS, Lemeshow S 2008. *Sampling of populations. Methods and Applications*. New York: John Wiley & Sons.

Malhotra NK 2004. *Pesquisa de Marketing: Uma orientação aplicada*. Porto Alegre: Artmed Editora.

Mattar FN 1996. *Pesquisa de marketing*. São Paulo: Atlas.

Ministério da Saúde 2005. Produzir e aplicar conhecimento na busca da universalidade e equidade, com qualidade da assistência à saúde da população. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. *Anais da 2a. Conferência Nacional de Ciência, Tecnologia e Inovação em Saúde*. SCTIE/DECIT, CNS. Brasilia.

Minitab 2007. Statistical Software-Powerful Statistical Software for data analysis, versão 15.0.

Pechansky F, Diemen LV, Inciardi JA, Surratt H, de Boni R 2004. Fatores de risco para transmissão do HIV em usuários de drogas de Porto Alegre, Rio Grande do Sul. Rio de Janeiro, Brasil. *Cad Saude Publica 20*: 1651-1660.

Poss J, Pierce R, Prieto V 2005. Herbal remedies used by selected migrant farmworkers in El Paso, Texas. *J Rural Health 21*: 187-191.

Poconé 2008. *Plano Municipal de Saúde de Poconé*. Prefeitura Municipal de Poconé. Secretaria municipal de Saúde. Poconé-MT.

Scheaffer RL, Mendenhall W, Ott L 2006. *Elementary Survey Sampling*. Belmont: Thomson.

Silva MAB, Melo LVL, Ribeiro RV, de Souza JPM, Lima JCS, Martins DT, Silva RM 2010. Levantamento etnobotânico de plantas utilizadas como anti-hiperlipidêmicas e anorexígenas pela população de Nova Xavantina-MT, Brasil. *Rev Bras Farmacogn 20*: 549-562.

Silva NN 2001. *Amostragem probabilística: um curso introdutório*. São Paulo: EDUSP.

Tomazzoni MI 2004. *Subsídios para a introdução do uso de fitoterápicos na rede básica de saúde do município de Cascavel/PR*. Cascavel, 113 p. Dissertação de Mestrado, Programa de Pós-graduação em Enfermagem, Universidade Federal do Paraná.

Tomazzoni MI, Negrelle RRB, Centa ML 2006. Fitoterapia popular: a busca instrumental enquanto prática terapeuta. *Texto Contexto Enferm 15*: 115-121.

Viganó J, Viganó JA, Silva DS, CTA 2007. Utilização de plantas medicinais pela população da região urbana de Três Barras do Paraná. *Acta Sci Health Sci 29*: 51-58.

*Correspondence

Domingos Tabajara de Oliveira Martins
Departamento de Ciências Básicas em Saúde, Área de Farmacologia, Faculdade de Medicina, Universidade Federal de Mato Grosso
Av. Fernando Correa da Costa, n. 2367, Campus Universitário
78060-900, Cuiabá-MT, Brasil
domingos.tabajara@pq.cnpq.br
Tel. 55 65 3615 8862
Fax: 55 65 3615 8863