Cost Assessment of Epidemiologic Surveys in Dentistry

Juliana Gonçalves*, Janice Paula, Glaucia Ambrosano, Fábio Mialhe and Antônio Pereira

Department of Community Dentistry, Piracicaba Dental School, University of Campinas (UNICAMP) Piracicaba, SP, Brazil

Abstract: *Objective*: The aim was to assess the relationship between the variability of data, sample size (n) and costs involved in epidemiologic surveys of dental caries.

Research design and settings: In order to conduct this study, simulations of the variation in costs of hypothetic epidemiologic surveys were made and studied. Thus, all costs with reference to a survey were described and divided into two categories: fixed and variable.

Outcome measures: The following margins of sampling errors were analyzed; 5%, 10%, 12% and 15% and the coefficients of variation (CV) of sampled data evaluated were, 50%, 80%, 100% and 120%.

Results: The required sample size increased with the reduction in the margin of error. For a CV of 50%, considering an error of 5%, the sample size was 384; for the same CV and error of 10%, n was 96. Thus it was observed that the relationship of sample size between the errors of 5% and 10 % was 4 times higher. Whereas with regard to cost, when an error of 5% was adopted, this was approximately three times higher when compared with the error of 10%.

Conclusion: Thus, when planning sample calculation, it is important to consider the Coefficient of Variation and the coherent errors with the variables under study, thus avoiding overestimating the sample and, consequently, increasing the costs involved in the research. It is fundamental to consider the possibility of working with other margins of error, thereby maintaining scientific strictness and establishing adequate costs.

Keywords: Coefficient of variation, cost, dentistry, epidemiology, sample size.

INTRODUCTION

Among the instruments for evaluating the performance of health services, health surveys stand out. The main goal of the epidemiologic survey in dentistry is to characterize populations with regard to their oral diseases. Surveys are important for monitoring the population's health status in order to help plan actions and investments and evaluate health policies [1]. Nevertheless, it is frequently unfeasible to evaluate the entire study population in one survey. Consequently, samples of the population are used. Samples represent a limited number of observations selected from a population on random basis, which yield generalizations about the population [2].

It is important for the researcher to know how to calculate the ideal sample size in dental surveys in accordance with defined statistical methods. Scientific methods for evaluating sample size provide credibility, reliability and the best cost benefit for dental surveys.

It is known that the larger the sample size, the greater the precision of the estimate [3]. It is the "Law of Large Numbers" enunciated centuries ago by the Frenchman Jaccob Bernoulli (1674-1705). Small samples usually generate unstable conclusions. If some units are added to these samples, substantial changes may occur in the estimates.

To determine the ideal sample size in an epidemiologic survey, one must take precision into consideration, that is, the sampling error that can be tolerated in the study [2]. Such precision depends upon the generalization of the results for the sample, that is, inference to the population. This way, one admits an error in the estimate that will be directly proportional to the variations in the presence of a certain characteristic in the population, and inversely proportional to the sample size (the larger the sample, the less chance of error) [2]. However, if the researcher is less demanding in his/her specifications of error, the survey can be developed with feasible resources and sample sizes.

Another characteristic that influences sample size is the type of sampling, that is, the way the sampling units are selected. The manner of selecting the sampling units, whether it is random or not, should be taken into consideration [4]. The main sampling techniques are Simple Random Sampling with or without replacement, Stratified Sampling, Systematic Sampling, Cluster Sampling and Convenience Sampling [5].

Thus, sample size must be practical and affordable. In addition, sample size in research deserves ethical considerations. This number must be the smallest possible in order not to involve people unnecessarily [6].

*Address corresponding to this author at the Department of Community Dentistry, Piracicaba Dental School, P.O. BOX 52, University of Campinas – UNICAMP, 13414-903, Piracicaba, SP, Brazil; Tel: 5511 21065209; Fax: 5511 21065209; E-mail: julianarocha11@uol.com.br

E-ISSN: 1929-6029/12

In general, the sample size is determined with such precision that allows the research findings to be maximized and costs minimized [7]. Therefore, the cost involved in planning research must be assessed as it is directly related to the sample size. The authors advocate the need for providing a representative sample of the population a timely and in accordance with the foreseen budget. Articles in the literature on the number of patients, costs and efficiency of dental studies are scarce, particularly with regard to epidemiologic surveys [1].

The aim of this study was to assess the relationship between the variability of data (Coefficient of Variability), sample size, costs involved in epidemiologic surveys of dental caries, and specifically, to observe changes in the cost when the sampling error is changed.

MATERIAL AND METHODS

Sample Calculation in Epidemiologic Surveys

Dental caries is the most prevalent oral disease and the focus of various studies in dentistry and public health. In oral health surveys, the researcher generally uses the mean (μ) of the DMFT index (number of decayed, missing, filled teeth). In the majority of cases, this mean is estimated using information coming from a sample, as the populations to be studied are large and a complete survey of the entire population would render the study unfeasible due to extra costs and time.

The number that represents the most probable value of the mean (based on sampling data) is called the point estimate of μ. Typically, the estimated value is not exactly equal to the true mean, thus a confidence interval is determined, which represents an interval of probable values for the mean based on the sampling data [8].

Therefore, a confidence interval of 95% for the mean DMFT represents an interval in which the actual mean is expected by 95% of the time and it can be expressed by:

$$\bar{x} - t_{(n-1;0,05)} \frac{s}{\sqrt{n}}, \bar{x} + t_{(n-1;0,05)} \frac{s}{\sqrt{n}}$$

where, $\bar{x}$ is the mean of the sample, t is obtained from the t-distribution table, s is an estimate of the standard deviation of DMFT in the population (σ) and n is the sample size.

The level of confidence must be fixed according to the desired probability of being correct in the estimation of the mean DMFT. In a confidence interval, the following expression is called semi-amplitude (*d*):

$$d = t_{(n-1;0,05)} \frac{s}{\sqrt{n}}$$

Thus, by fixing d and the level of confidence, one can calculate the sample size through the expression [8].

$$n = \left(\frac{t_{(n-1;0,05)} s}{d}\right)^2$$

In addition, one must consider the effect of the design, which is basically the ratio of actual variance, under the sampling method used, for the variance in the case of simple random sampling [8, 9].

Cost Calculation

For this research, a detailed study was conducted on the items of an epidemiologic survey and the costs of each item. These costs were classified as fixed and variable costs. Fixed costs were divided into a) fixed investment costs and b) infrastructure costs. Fixed investment costs are those related to the equipment required for conducting a research. Fixed infrastructure costs (human resources) refer to the costs of building up the structure of resources for developing the research. This includes the minimum team required for planning the study [10]. In addition, the variable costs were described, which are those that change as a result of the number of exams performed. Chart **1** shows details of the items of each cost assessed in this study.

Estimated Sample Sizes and Costs

Using Excel Spreadsheets (Microsoft Corp., Redmond, Washington, USA), epidemiologic surveys were simulated and the costs evaluated from the calculated sample size that was calculated by the mean, error and standard deviation of the hypothetically determined DMFT [11]. The sampling error is the difference between the estimated sample and the population parameter [8], whereas the

1- Fixed/Structural Costs of the Experiment
1.1) Investments (Equipment etc)
Computer
Software
1.2) Infrastructure
Salaries/3-month contract with the planning team
Researcher in Charge/ Head of Research: Coordination, Training and Data Analysis.
2- Variable Costs Assessed
Dentist
Assistant
Tongue retractor
Gauze
Mirror
Forms

Chart 1: Description of cost items assessed in the Epidemiologic Survey.

Coefficient of Variation (CV) is the measurement of independent dispersion of the unit of measurement of the variable in relation to the mean. One can reach the estimate of CV by dividing the standard deviation by the mean [12].

To estimate the costs and samples, fixed values of sampling error (d) were determined at 5%, 10%, 12% and 15% of the mean and of the Coefficient of Variation at 50%, 80%, 100% and 120%.

RESULTS

Table **1** represents the relationship between sample size and sampling error (5%, 10%, 12% and 15%) for the different CVs (50%, 80%, 100% and 120%). It may be noticed that for the error of 5%, the sample size is much larger when compared with the other errors in all the CVs.

Table 1: Relationship Among Coefficient of Variation (DMFT Determined Hypothetically), Sampling Error (Estimated by 0.05, 0.10, 0.12, 0.15) and Sample Size Calculated According Formule CV Equal to Variability Dived by Mean

	DMFT(CV)			
Error range	**1.9 (50)**	**2.7 (80)**	**2.07 (100)**	**2.28 (120)**
0,05	384	983	903	2219
0,10	96	246	226	555
0,12	67	171	157	385
0,15	43	109	100	247

Also in Table **1**, the sample size is presented with various and sampling error. One can observe that, for a margin of error of 5%, the sample size is 384 when a 50% Coefficient of Variation was considered. For this we used the follow formula:

$$n = m*S^2/(e*m)(e*m)$$

where the mean (m) was 1.9; the error (e) was 0.05 and the variability (s) was 0.95. For an error of 10%, this value was considerably lower, 96 for the same Coefficient of Variation.

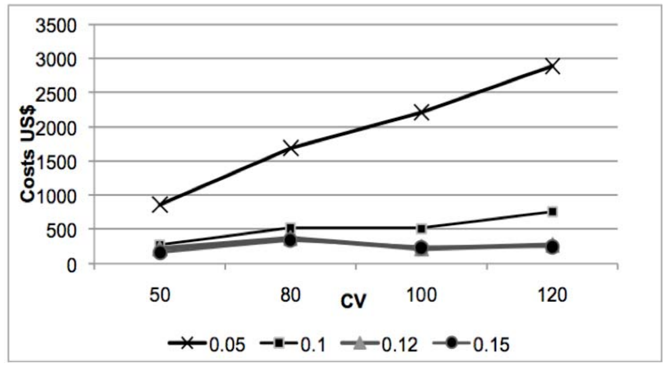

Figure 1: Relationship among costs and coefficient of Variation.

When assessing the cost (Figure **1**) when an error of 5% was adopted, this was approximately three times higher when compared with the error of 10%, and approximately four times higher when related to the error of 12% for a Coefficient of Variation of 50% (Table **2**).

Table 2: Variation in Cost (US$) as a Function of Error and Coefficient of Variation

	CV			
Error	**50 CV**	**80 CV**	**100 CV**	**120 CV**
0.05	862.2	1685.4	2212.3	2892.7
0.1	276.3	521.1	516.8	759.4
0.12	219.04	374.26	245.21	275.29
0.15	166.46	341.06	235.56	244.16

DISCUSSION

For descriptive purposes, population parameters are used as totals, means and proportions. In the majority of cases, however, the data are used requiring estimates of measurements of variability (such as variance and standard deviation) for calculating the confidence intervals and performing hypothesis tests [1].

It is important to define the extent of error that can be tolerated in the research, which depends on the variables that are being analyzed [1, 10, 12]. There is an inverse relationship between error and sample size: large samples are associated with small errors and small samples to large errors [1, 2]. In addition, when one wants to estimate the mean of a continuous variable, one perceives that the greater the variability of this variable in relation to the mean, the larger will be the sample necessary for estimating its mean with the desired precision [8].

Through this research, it was perceived that the sample size to be defined depends only on the CV and sampling error. According to the results found, it was possible to visualize the n (sample size) necessary based on the CV and sampling error.

When dealing with sampling errors, when one considers an error of 5%, the sample size frequently generates a very high and unnecessary cost. It would be possible to adopt an error of 10% to work with a smaller sample and, consequently, at a lower cost. Thus, researchers could correctly plan the costs involved in a research considering that the error value of 5% is an arbitrary parameter and can be altered [8].

One may observe that studies with regard to costs in health surveys have been in existence for some time whereas researches that evaluate costs in Epidemiologic Surveys are scarce in the literature [14, 15].

The results of the literature emphasize the concern about the reduction of costs involved in research [14, 16]. Some authors assessed this reduction in cohort and case control studies, altering the power (1-beta) and obtaining adequate sampling with an excellent cost benefit [13]. They proposed the following procedures: maximize the power for total fixed costs and minimize the total cost for a specific procedure Compared with the present study, the alteration made was in the range of sampling error, which made the Epidemiological Survey feasible with a reduction in n and in costs [14].

One of the important findings of the study was that the relationship of sample sizes was approximately 4 times higher between errors of 5% and 10 % (CV of 50%), whereas with regard to cost, when an error of 5% was adopted, it was approximately three times higher when compared with the error of 10% (CV of 50%). When the error of 10% was evaluated, the cost was approximately 2 times higher in relation to the error of 12% (CV of 100%) In this context, in a review of five case control studies, some researchers found that the cost of sampling represented approximately 75% of the total cost of the study, and that the unit of cost value ranged between 1.5 and 2. For these five studies, they showed that the strategy of sample design to minimize the cost resulted in savings of no more than 2% on the total cost of the study about an equal sample size design with identical power. Nevertheless, the amount of savings obtained is worth thousands of dollars or even more in large studies [17].

This investigation indicates that knowing how to plan a study adequately will allow for its technical and scientific feasibility, as it will indicate the cost for the sample size with a certain precision. This will allow greater reliability and credibility of the scientific research.

CONCLUSIONS

It is important to consider the Coefficient of Variation and the errors coherent with the studied variables. Thus, planning of the sample calculation could prevent the sample from being overestimated and, consequently, unnecessary increases in the costs involved in the research. Therefore, it is fundamental to consider the possibility of working with other margins of error, thereby maintaining scientific strictness and establishing adequate costs.

REFERENCES

[1] Szwarcwald CL, Damacena GN. Complex Sampling Design in Population Surveys: Planning and effects on statistical data analysis. Rev Bras Epidemiol 2008; 11(Suppl 1): 38-45.

[2] Pereira JCR. Tamanho de amostra: uma necessidade real ou um capricho cultural? Arteríola 2002; 4(1): 13-16. [in portuguese].

[3] Sedlmeier P, Gigerenzer G. Intuitions about sample size: The empirical law of large numbers. J Behav Decision Making 1997; 10: 33-51. http://dx.doi.org/10.1002/(SICI)1099-0771(199703)10:1<33::AID-BDM244>3.0.CO;2-6

[4] Rana H, Andersen RM, Nakakono TT. ICS II USA Research Design and Methodology. Adv Dent Res 1997; 11(2): 217-22. http://dx.doi.org/10.1177/08959374970110020401

[5] Hulley SB. Designing clinical research. 3rd ed. Philadelphia PA: Lippincott Williams & Wilkins 2007; pp. 63-72.

[6] Vanderpool, H. The Ethics of Research Involving Human Subjects: Facing the 21st Century; 1996. Frederick, MD: University Publishing Group.

[7] Hujoel P, DeRouen TA. Determination and Selection of the Optimum Number of Sites and Patients for Clinical Studies. J Dent Res 1992; 71(8): 1516-21. http://dx.doi.org/10.1177/00220345920710081001

[8] Cochran WG. Sampling Techniques - Wiley Series in Probability and Mathematical Statistics. 3rd edn. p51-88; 318-324. Canada: Willey 1977.

[9] US Census Bureau. Technical Paper 63: Current Population Survey - Design and Methodology, Available from: http://www.census.gov/prod/2002pubs/tp63rv.pdf > accessed 07 November, 2011; pp. 14-7.

[10] Woiler S, Mathias F. Projetos: Planejamento, elaboração, análise. 1st ed. São Paulo SP: Atlas 1986; pp. 115-118.

[11] Meza AR, Angelis M, Britt H, Milles DA, Seneta E, Webb BC. Aust J Public Health 1995; 19: 34-40. http://dx.doi.org/10.1111/j.1753-6405.1995.tb00294.x

[12] Eldridge SM, Ashby D, Kerry S. Sample Size cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol 2006; 35: 1292-300. http://dx.doi.org/10.1093/ije/dyl129

[13] Alves MCGP, Silva NN. Variance estimation methods in samples from household surveys. Rev Saúde Pública 2007; 41(6): 1-8.

[14] Morgenstern H, Winn DM. Statistics Med 1983; 2: 387-96. http://dx.doi.org/10.1002/sim.4780020311

[15] Miettinen OS. Individual Matching with Multiple Controls in the Case of All-or-None Responses. Biometrics 1969; 25(2): 339-55. http://dx.doi.org/10.2307/2528794

[16] Meydrech EF, Kupper LL. Cost considerations and sample size requirements in cohort and case- control studies. Am J Epidemiol 1978; 107: 201-5.

[17] Brittain E, Schlesselman JJ, Stadel BV. Cost of case-control studies. Am J Epidemiol 1981; 114: 234-43.

Received on 26-07-2012 Accepted on 24-08-2012 Published on 24-09-2012

http://dx.doi.org/10.6000/1929-6029.2012.01.01.05