

A Study on Design, Data Analysis and Sampling Techniques for Clinical Research

R. Seenivasan

Asst. Professor, Dept of Mathematical Economics, School of Economics,
M.K.University, Madurai – 625 021, E-mail: researcher9747@rediffmail.com

Received: 27 March 2020; Revised: 10 April 2020;
Accepted: 9 June 2020; Publication: 15 July 2020

Abstract: Statistical analysis is an essential technique that enables a medical research practitioner to draw meaningful inference from their data analysis. Improper application of study design and data analysis may render insufficient and improper results and conclusion. Converting a medical problem into a statistical hypothesis with appropriate methodological and logical design and then back-translating the statistical results into relevant medical knowledge is a real challenge. This article explains various sampling methods that can be appropriately used in medical research with different scenarios and challenges.

In this paper, the basic elements related to the selection of participants for a health research are discussed. Sample representativeness, sample frame, types of sampling, as well as the impact that non-respondents may have on results of a study are described. The whole discussion is supported by practical examples to facilitate the reader's understanding.

Keywords: Design, study, Dermatology, Epidemiology and biostatistics, Epidemiologic studies, Sample size, Sampling studies.

INTRODUCTION

The essential topics related to the selection of participants for a health research are: 1) whether to work with samples or include the whole reference population in the study (census), 2) the sample basis, 3) the sampling process and 4) the potential effects nonrespondents might have on study results. We will refer to each of these aspects with theoretical and practical examples for better understanding in the sections that follow.

TO SAMPLE OR NOT TO SAMPLE

In a previous paper, we discussed the necessary parameters on which to estimate the sample size.¹ We define sample as a finite part or subset of participants drawn from the target population. In turn, the target population corresponds to the entire set of subjects whose characteristics are of interest to the research team. Based on results obtained from a sample, researchers may draw their conclusions about the target population with a certain level

of confidence, following a process called statistical inference. When the sample contains fewer individuals than the minimum necessary, but the representativeness is preserved, statistical inference may be compromised in terms of precision (prevalence studies) and/or statistical power to detect the associations of interest.¹ On the other hand, samples without representativeness may not be a reliable source to draw conclusions about the reference population (i.e., statistical inference is not deemed possible), even if the sample size reaches the required number of participants. Lack of representativeness can occur as a result of flawed selection procedures (sampling bias) or when the probability of refusal/non-participation in the study is related to the object of research (nonresponsive bias).

Although most studies are performed using samples, whether or not they represent any target population, census-based estimates should be preferred whenever possible.^{3,4} For instance, if all cases of melanoma are available on a national or regional database, and information on the potential risk factors are also available, it would be preferable to conduct a census instead of investigating a sample. However, there are several theoretical and practical reasons that prevent us from carrying out census-based surveys, including:

Ethical issues: It is unethical to include a greater number of individuals than that effectively required.

Budgetary limitations: The high costs of a census survey often limits its use as a strategy to select participants for a study.

Logistics: Censuses often impose great challenges in terms of required staff, equipment, etc. to conduct the study.

Time restrictions: The amount of time needed to plan and conduct a census-based survey may be excessive.

Unknown target population size: If the study objective is to investigate the presence of premalignant skin lesions in illicit drugs users, lack of information on all existing users makes it impossible to conduct a census-based study.

All these reasons explain why samples are more frequently used. However, researchers must be aware that sample results can be affected by the random error (or sampling error). To exemplify this concept, we will consider a research study aiming to estimate the prevalence of premalignant skin lesions (outcome) among individuals >18 years residing in a specific city (target population). The city has a total population of 4,000 adults, but the investigator decided to collect data on a representative sample of 400

participants, detecting an 8% prevalence of premalignant skin lesions. A week later, the researcher selects another sample of 400 participants from the same target population to confirm the results, but this time observes a 12% prevalence of premalignant skin lesions. Based on these findings, is it possible to assume that the prevalence of lesions increased from the first to the second week? The answer is probably not. Each time we select a new sample, it is very likely to obtain a different result. These fluctuations are attributed to the "random error." They occur because individuals composing different samples are not the same, even though they were selected from the same target population. Therefore, the parameters of interest may vary randomly from one sample to another. Despite this fluctuation, if it were possible to obtain 100 different samples of the same population, approximately 95 of them would provide prevalence estimates very close to the real estimate in the target population - the value that we would observe if we investigated all the 4,000 adults residing in the city. Thus, during the sample size estimation the investigator must specify in advance the highest or maximum acceptable random error value in the study. Most population-based studies use a random error ranging from 2 to 5 percentage points. Nevertheless, the researcher should be aware that the smaller the random error considered in the study, the larger the required sample size.

SAMPLE FRAME

The sample frame is the group of individuals that can be selected from the target population given the sampling process used in the study. For example, to identify cases of cutaneous melanoma the researcher may consider to utilize as sample frame the national cancer registry system or the anatomopathological records of skin biopsies. Given that the sample may represent only a portion of the target population, the researcher needs to examine carefully whether the selected sample frame fits the study objectives or hypotheses, and especially if there are strategies to overcome the sample frame limitations (see Chart 1 for examples and possible limitations).

DEFINITION OF SAMPLING

Sampling can be defined as the process through which individuals or sampling units are selected from the sample frame. The sampling strategy needs to be specified in advance, given that the sampling method may affect the sample size estimation. Without a rigorous sampling plan the estimates derived from the study may be biased (selection bias).

Chart 1: Examples of sample frames and potential limitations as regards representativeness

<i>Sample frames</i>	<i>Limitations</i>
Population census	<ul style="list-style-type: none"> • If the census was not conducted in recent years, areas with high migration might be outdated • Homeless or itinerant people cannot be represented
Hospital or Health Services records	<ul style="list-style-type: none"> • Usually include only data of affected people (this is a limitation, depending on the study objectives) • Depending on the service, data may be incomplete and/or outdated • If the lists are from public units, results may differ from those who seek private services
School lists	<ul style="list-style-type: none"> • School lists are currently available only in the public sector • Children/ teenagers not attending school will not be represented • Lists are quickly outdated • There will be problems in areas with high percentage of school absenteeism
List of phone numbers	<ul style="list-style-type: none"> • Several population groups are not represented: individuals with no phone line at home (low-income families, young people who use only cell phones), those who spend less time at home, etc.
Mailing lists	<ul style="list-style-type: none"> • Individuals with multiple email addresses, which increase the chance of selection compared to individuals with only one address • Individuals without an email address may be different from those who have it, according to age, education, etc.

TYPES OF SAMPLING

In figure 1, we depict a summary of the main sampling types. There are two major sampling types: probabilistic and nonprobabilistic.

PROBLEM IDENTIFICATION

Clinical research often starts from questions raised at the bedside in hospital wards. Is there an association between neurocysticercosis and epilepsy? Are magnetic resonance imaging changes good predictors of multiple sclerosis? Is there a benefit in using steroids in pyogenic meningitis? Typically, these questions lead us to set up more refined research questions. For example, do persons with epilepsy have a higher probability of having

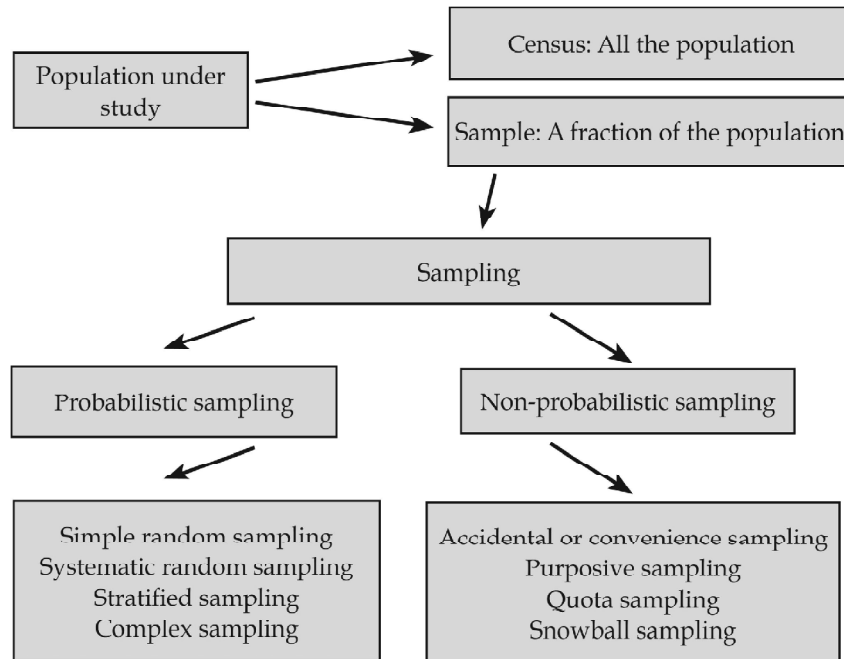


Figure 1: Sampling types used in scientific studies

serological (or Computed Tomography scan) markers for NCC? What proportion of persons with multiple lesions in the brain has Multiple Sclerosis Do children with pyogenic meningitis have a lesser risk of mortality if dexamethasone is used concomitantly with antibiotics?

Designing a clinical study involves narrowing a topic of interest into a single focused research question, with particular attention paid to the methods used to answer the research question from a cost, viability and overall effectiveness standpoint. In this paper, we focus attention on residents and younger faculty who are planning short-term research projects that could be completed in 2–3 years. Once we have a fairly well-defined research question, we need to consider the best strategy to address these questions. Further considerations in clinical research, such as the clinical setting, study design, selection criteria, data collection and analysis, are influenced by the disease characteristics, prevalence, time availability, expertise, research grants and several other factors. In the example of NCC, should we use serological markers or CT scan findings as evidence of NCC? Such a question raises further questions. How good are serologic markers compared with CT scans in terms of identifying NCC? Which test (CT or blood test) is easier, safer and acceptable for this study? Do we have the expertise to carry out these laboratory tests and make interpretations? Which

procedure is going to be more expensive? It is very important that the researcher spend adequate time considering all these aspects of his study and engage in discussion with biostatisticians before actually starting the study.

The major objective of this article is to explain these initial steps. We do not intend to provide a tailor-made design. Our aim is to familiarize the reader with different sampling methods that can be appropriately used in medical research with different scenarios and challenges.

SETTING

One of the first steps in clinical study is choosing an appropriate setting to conduct the study (i.e., hospital, population based). Some diseases, such as migraine, may have a different profile when evaluated in the population than when evaluated in the hospital. On the other hand, acute diseases such as meningitis would have a similar profile in the hospital and in the community. The observations in a study may or may not be generalizable, depending on how closely the sample represents the population at large.

Consider the following studies. Both De Gans and Scarborough looked at the effect of adjunctive Dexamethasone in bacterial meningitis. Both studies are good examples of using the hospital setting. Because the studies involved acute conditions, they utilize the fact that sicker patients will seek hospital care to concentrate their ability to find patients with meningitis. By the same logic, it would be inappropriate to study less-acute conditions in such a fashion as it would bias the study toward sicker patients.

On the other hand, consider the study by Holroyd investigating therapies in the treatment of migraine. Here, the authors intentionally chose an outpatient setting (the patients were referred to the study clinic from a network of other physician clinics as well as local advertisements) so that their population would not include patients with more severe pathology (requiring hospital admission).

If the sample was restricted to a particular age group, sex, socioeconomic background or stage of the disease, the results would be applicable to that particular group only. Hence, it is important to decide how you select your sample. After choosing an appropriate setting, attention must be turned to the inclusion and exclusion criteria. This is often locale specific. If we compare the exclusion criteria for the two meningitis studies mentioned above, we see that in the study by de Gans, patients with shunts, prior neurosurgery and active tuberculosis were specifically excluded; in the

Scarborough study, however, such considerations did not apply as the locale was considerably different (sub-saharan Africa vs. Europe).

VALIDITY AND RELIABILITY

Clinical research generally requires making use of an existing test or instrument. These instruments and investigations have usually been well validated in the past, although the populations in which such validations were conducted may be different. Many such questionnaires and patient self-rating scales (MMSE or QOLIE, for instance) were developed in another part of the world. Therefore, in order to use these tests in clinical studies locally, they require validation. Socio-demographic characteristics and language differences often influence such tests considerably. For example, consider a scale that uses the ability to drive a motor car as a Quality of Life measure. Does this measure have the same relevance in India as in the USA, where only a small minority of people drives their own vehicles? Hence, it is very important to ensure that the instruments that we use have good validity.

Validity is the degree to which the investigative goals are measured accurately. The degree to which the research truly measures what it intended to measure determines the fundamentals of medical research. Peace, Parrillo and Hardy explain that the validity of the entire research process must be critically analyzed to the greatest extent possible so that appropriate conclusions can be drawn, and recommendations for development of sound health policy and practice can be offered.

Another measurement issue is reliability. Reliability refers to the extent to which the research measure is a consistent and dependable indicator of medical investigation. In measurement, reliability is an estimate of the degree to which a scale measures a construct consistently when it is used under the same condition with the same or different subjects. Reliability (consistency) describes the extent to which a measuring technique consistently provides the same results if the measurement is repeated. The validity (accuracy) of a measuring instrument is high if it measures exactly what it is supposed to measure. Thus, the validity and reliability together determine the accuracy of the measurement, which is essential to make valid statistical inference from a medical research.

Consider the following scenario. Kasner established reliability and validity of a new National Institute of Health Stroke Scale (NIHSS) generation method. This paper provides a good example of how to test a new instrument (NIH stroke score generation via retrospective chart review) with regards to its reliability and validity. To test validity, the investigators

had multiple physicians review the same set of charts and compared the variability within the scores calculated by these physicians. To test reliability, the investigators compared the new test (NIHSS calculated by chart review) to the old test (NIHSS calculated at the bedside at the time of diagnosis). They reported that, overall, 88% of the estimated scores deviated by less than five points from the actual scores at both admission and discharge.

SAMPLING

A major purpose of doing research is to infer or generalize research objectives from a sample to a larger population. The process of inference is accomplished by using statistical methods based on probability theory. A sample is a subset of the population selected, which is an unbiased representative of the larger population. Studies that use samples are less-expensive, and study of the entire population is sometimes impossible. Thus, the goal of sampling is to ensure that the sample group is a true representative of the population without errors. The term error includes sampling and non sampling errors. Sampling errors that are induced by sampling design include selection bias and random sampling error. Non sampling errors are induced by data collection and processing problems, and include issues related to measurement, processing and data collection errors.

METHODS OF SAMPLING

To ensure reliable and valid inferences from a sample, probability sampling technique is used to obtain unbiased results. The four most commonly used probability sampling methods in medicine are simple random sampling, systematic sampling, stratified sampling and cluster sampling. In simple random sampling, every subject has an equal chance of being selected for the study. The most recommended way to select a simple random sample is to use a table of random numbers or a computer-generated list of random numbers. Consider the study by Kamal that aimed to assess the burden of stroke and transient ischemic attack in Pakistan. In this study, the investigators used a household list from census data and picked a random set of households from this list. They subsequently interviewed the members of the randomly chosen households and used this data to estimate cerebrovascular disease prevalence in a particular region of Pakistan. Prevalence studies such as this are often conducted by using random sampling to generate a sampling frame from preexisting lists (such as census lists, hospital discharge lists, etc.). A systematic random sample is one in which every k^{th} item is selected. k is determined by dividing the number of items in the sampling frame by sample size.

A stratified random sample is one in which the population is first divided into relevant strata or subgroups and then, using the simple random sample method, a sample is drawn from each strata. Deng studied IV Tissue Plasminogen Activator (TPA) usage in acute stroke among hospitals in Michigan. In order to enroll patients across a wide array of hospitals, they employ a stratified random sampling in order to construct the list of hospitals. They stratified hospitals by number of stroke discharges, and then randomly picked an equal number of hospitals within each stratum. Stratified random sampling such as this can be used to ensure that sampling adequately reflects the nature of current practice (such as practice and management trends across the range of hospital patient volumes, for instance).

A cluster sample results from a two-stage process in which the population is divided into clusters, and a subset of the clusters is randomly selected. Clusters are commonly based on geographic areas or districts and, therefore, this approach is used more often in epidemiologic research than in clinical studies.

RANDOM SAMPLES AND RANDOMIZATION

Random samples and randomization (aka, random assignment) are two different concepts. Although both involve the use of the probability sampling method, random sampling determines who will be included in the sample. Randomization, or random assignment, determines who will be in the treatment or control group. Random sampling is related to sampling and external validity (generalizability), whereas random assignment is related to design and internal validity.

In experimental studies such as randomized controlled trials, subjects are first selected for inclusion in the study on the basis of appropriate criteria; they are then assigned to different treatment modalities using random assignment. Randomized controlled trials that are considered to be the most efficient method of controlling validity issues by taking into account all the potential confounding variables (such as other outside factors that could influence the variables under study) are also considered most reliable and impartial method of determining the impact of the experiment. Any differences in the outcome of the study are more likely to be the result of difference in the treatments under consideration than due to differences because of groups.

Scarborough in a trial published in the *New England Journal of Medicine*, looked at corticosteroid therapy for bacterial meningitis in sub-saharan Africa to see whether the benefits seen with early corticosteroid

administration in bacterial meningitis in the developed world also apply to the developing world. Interestingly, they found that adjuvant Dexamethasone therapy did not improve outcomes in meningitis cases in sub-saharan Africa. In this study, they performed random assignment of therapy (Dexamethasone vs. placebo). It is useful to note that the process of random assignment usually involves multiple sub-steps, each designed to eliminate confounders. For instance, in the above-mentioned study, both steroids and placebo were packaged similarly, in opaque envelopes, and given to patients (who consented to enroll) in a randomized fashion. These measures ensure the double-blind nature of the trial. Care is taken to make sure that the administrators of the therapy in question are blinded to the type of therapy (steroid vs. placebo) that is being given.

SAMPLE SIZE

The most important question that a researcher should ask when planning a study is “How large a sample do I need?” If the sample size is too small, even a well-conducted study may fail to answer its research question, may fail to detect important effects or associations, or may estimate those effects or associations too imprecisely. Similarly, if the sample size is too large, the study will be more difficult and costly, and may even lead to a loss in accuracy. Hence, optimum sample size is an essential component of any research. Careful consideration of sample size and power analysis during the planning and design stages of clinical research is crucial.

Statistical power is the probability that an empirical test will detect a relationship when a relationship in fact exists. In other words, statistical power explains the generalizability of the study results and its inferential power to explain population variability. Sample size is directly related to power; *ceteris paribus*, the bigger a sample, the higher the statistical power. If the statistical power is low, this does not necessarily mean that an undetected relationships exist, but does indicate that the research is unlikely to find such links if they exist. Flow chart relating research question, sampling and research design and data analysis is shown in Figure 2.

The power of a study tells us how confidently we can exclude an association between two parameters. For example, regarding the prior research question of the association between NCC and epilepsy, a negative result might lead one to conclude that there is no association between NCC and epilepsy. However, the study might not have been sufficiently powered to exclude any possible association, or the sample size might have been too small to reveal an association.

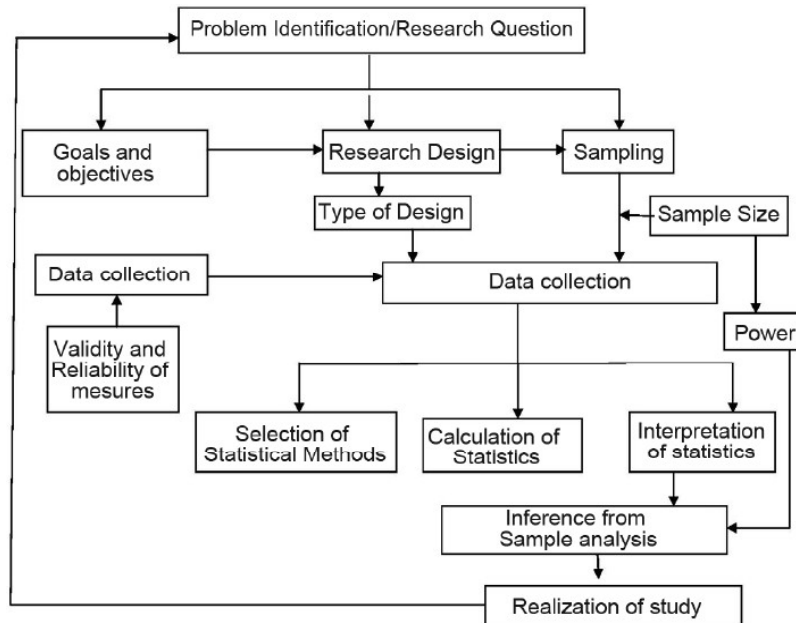


Figure 2

The sample sizes seen in the two meningitis studies mentioned earlier are calculated numbers. Using estimates of prevalence of meningitis in their respective communities, along with variables such as size of expected effect (expected rate difference between treated and untreated groups) and level of significance, the investigators in both studies would have calculated their sample numbers ahead of enrolling patients. Sample sizes are calculated based on the magnitude of effect that the researcher would like to see in his treatment population (compared with placebo). It is important to note variables such as prevalence, expected confidence level and expected treatment effect need to be predetermined in order to calculate sample size. As an example, Scarborough *et al.* state that "on the basis of a background mortality of 56% and an ability to detect a 20% or greater difference in mortality, the initial sample size of 660 patients was modified to 420 patients to detect a 30% difference after publication of the results of a European trial that showed a relative risk of death of 0.59 for corticosteroid treatment." Determining existing prevalence and effect size can be difficult in areas of research where such numbers are not readily available in the literature. Ensuring adequate sample size has impacts for the final results of a trial, particularly negative trials. An improperly powered negative trial could fail to detect an existing association simply because not enough patients were enrolled.

In other words, the result of the sample analysis would have failed to reject the null hypothesis (that there is no difference between the new treatment and the alternate treatment), when in fact it should have been rejected, which is referred to as type II error. This statistical error arises because of inadequate power to explain population variability. Careful consideration of sample size and power analysis is one of the prerequisites of medical research. Another prerequisite is appropriate and adequate research design, which will be addressed in the next issue.

CONCLUSION

Sampling methods are applied in almost all stages of a research study from the planning and design stage to the final reporting of findings. Utmost care must be taken in choosing the right sampling method at each stage of a study. Sample size estimation is very important and choosing the right formula is crucial in all types of research studies. The selection of appropriate sample size formula depends on the primary objective of the study, the type of outcome variable, study design used, the statistical analysis planned, number of groups in the study and sampling technique to be adopted. The number of subjects to be recruited in a study depends on the power of the study, the precision of the estimated value, level of significance or confidence level, clinically significant difference and also other constraints such as money, manpower, availability of subjects and time in particular and feasibility in general. There are many statistical tests and the choice of appropriate analysis depends on the objective of the study, types of the outcome variable, number of variables, sample size, number of groups in the study, whether the groups are related or not and also on distributional assumptions.

REFERENCES

- Applegate KE, Crewson PE. Statistical literacy. *Radiology* 2004; 230: 613-4.
- Barrios CH, Werutsky G, Martinez-Mesa J. The global conduct of cancer clinical trials: challenges and opportunities. *Am Soc Clin Oncol Educ Book*. 2015: e132-9.
- Boing AC, Peres KG, Boing AF, Hallal PC, Silva NN, Peres MA. EpiFloripa Health Survey: the methodological and operational aspects behind the scenes. *Rev Bras Epidemiol*. 2014; 17: 147-62.
- De Gans J, de Beek VD. "Dexamethasone in Adults with Bacterial Meningitis" *N Engl J Med*. 2002; 324: 1549-56.
- Duquia RP, Menezes AM, Almeida HL Jr, Reichert FF, Santos Ida S, Haack RL, *et al*. Prevalence of sun exposure and its associated factors in southern Brazil: a population-based study. *An Bras Dermatol*. 2013; 88: 554-61.
- Goncalves JR, Nappo SA. Factors that lead to the use of crack cocaine in combination with marijuana in Brazil: a qualitative study. *BMC Public Health*. 2015; 15: 706.

- Holroyd KA, Cottrell CK, O'Donnell FJ, Drew JB, Carlson BW, Himawan L. "Effect of Preventive (â blocker) Treatment, behavioural Migraine Management or their combination on outcomes of optimized acute treatment in frequent migraine: Randomized controlled trial" *BMJ*. 2010; 341: c4871.
- Scarborough M, Gordon S, Whitty C, French N, Njale Y, Chitani A. "Corticosteroids for Bacterial Meningitis in Adults in Sub-Saharan Africa" *N Engl J Med*. 2007; 357: 2441-50.
- Golafshani N. Understanding reliability and validity in qualitative research. *Qual Rep*. 2003; 8: 597-607.
- Peace KE, Parrillo AV, Hardy JC. "Assessing the validity of statistical inferences in public health research: An Evidence – based 'Best Practice' approach" *J Georgia Public Health Assoc*. 2008; 1: 1.
- Kasner SE, Chalela JA, Luciano JM, Cucchiara BL, Raps EC, McGarvey ML, et al. "Reliability and Validity of Estimating the NIH Stroke Scale from Medical Records" *Stroke*. 1999; 30: 1534-7.
- Kamal A, Itrat A, Murtaza M, Khan M, Rasheed A. The Burden of stroke and TIA in Pakistan. *BMC Neurol*. 2009; 9: 58.
- Deng Y, Reeves M, Jacobs B, Birbeck L, Kothari R. IV tissue plasminogen activator use in acute stroke. *Neurology*. 2006; 66: 306-12.
- Dawson B, Trapp GR. New York: Lange Medical Books, McGraw Hill; 2001. "Basic and Clinical Biostatistics".
- Martínez-Mesa J, González-Chica DA, Bastos JL, Bonamigo RR, Duquia RP. Sample size: how many participants do I need in my research? *An Bras Dermatol*. 2014; 89: 609-15.
- Röhrig B, du Prel JB, Wachtlin D, Kwiecien R, Blettner M. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2010; 107: 552-6.
- Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. *Ann Indian Acad Neurol*. 2011; 14: 287-90.
- Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013; 42: 1012-4.
- Krause M, Lutz W, Boehnke JR. The role of sampling in clinical trial design. *Psychother Res*. 2011; 21: 243-51.
- Roubille C, Richer V, Starnino T, McCourt C, McFarlane A, Fleming P, et al. Evidence-based Recommendations for the Management of Comorbidities in Rheumatoid Arthritis, Psoriasis, and Psoriatic Arthritis: Expert Opinion of the Canadian Dermatology-Rheumatology Comorbidity Initiative. *J Rheumatol*. 2015; 42: 1767-80.
- Larkin J, Ascierto PA, Dréno B, Atkinson V, Liskay G, Maio M, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med*. 2014; 371: 1867-76.
- Goncalves JR, Nappo SA. Factors that lead to the use of crack cocaine in combination with marijuana in Brazil: a qualitative study. *BMC Public Health*. 2015; 15: 706.

-
- Pimenta FB, Pinho L, Silveira MF, Botelho AC. Factors associated with chronic diseases among the elderly receiving treatment under the Family Health Strategy. *Cien Saude Colet.* 2015; 20: 2489-98.
- Kelbore AG, Alemu W, Shumye A, Getachew S. Magnitude and associated factors of Atopic dermatitis among children in Ayder referral hospital, Mekelle, Ethiopia. *BMC Dermatol.* 2015; 15: 15.
- Tamela FD, Ketchen JD., Jr "Organizational Configuration and Performance: The role of statistical power in extant research" *Strateg Manage J.* 1999; 20: 385-95.
- Victora CG, Barros FC. Cohort profile: the 1982 Pelotas (Brazil) birth cohort study. *Int J Epidemiol.* 2006; 35: 237-42.
- Zhou SJ, Skeaff M, Makrides M, Gibson R. Vitamin D status and its predictors among pre-school children in Adelaide. *J Paediatr Child Health.* 2015; 51: 614-9.