Design, data analysis and sampling techniques for clinical research

Karthik Suresh, Sanjeev V. Thomas¹, Geetha Suresh²

Department of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, ¹Department of Neurology, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India, ²Department of Justice Administration, University of Louisville, Louisville, USA

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.

Key Words

Design, study, sampling

Problem Identification

Clinical research often starts from questions raised at the bedside in hospital wards. Is there an association between neurocysticercosis (NCC) 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 serological (or computed tomography [CT] scan) markers for NCC? What proportion of persons with multiple lesions in the brain has Multiple Sclerosis (MS) 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 et al.[1] and Scarborough et al.[2] 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 et al.[3] 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,[1] 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 (Precision) and Reliability (Consistency)

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 drive 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[4] determines the fundamentals of medical research. Peace, Parrillo and Hardy[5] 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 et al.[6] 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 nonsampling errors. Sampling errors that are induced by sampling design include selection bias and random sampling error. Nonsampling 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 et al. 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 kth 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 et al. 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 cases, 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 et al. 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 important 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 1.

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.

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.\[2\] 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.

References

1. De Gans J, de Beek VD. “Dexamethasone in Adults with Bacterial Meningitis”. N Engl J Med 2002;324:1549-56.
2. 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.
3. Holroyd KA, Cottrell CK, O’Donnell FJ, Drew JB, Carlson BW, Himawan L. “Effect of Preventive (β blocker) Treatment, behavioural Migraine Management or their combination on outcomes of optimized acute treatment in frequent migraine: Randomized controlled trial”. BMJ 2010;341:c4871.
4. Golafshani N. Understanding reliability and validity in qualitative research. Qual Rep 2003;8:597-607.
5. 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. Available from: http://faculty.mercer.edu/thomas_bm/documents/gphpa/documents/Archive/Pace,%20Statistical%20Inferences,%202008.pdf
6. 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.
7. Kamal A, Itrat A, Murtaza M, Khan M, Rasheed A. The Burden of stroke and TIA in Pakistan. BMC Neurol 2009;9:58.
8. Deng Y, Reeves M, Jacobs B, Birbeck L, Kothari R. IV tissue plasminogen activator use in acute stroke. Neurology 2006;66:306-12.
9. Dawson B, Trapp GR. “Basic and Clinical Biostatistics”. New York: Lange Medical Books, McGraw Hill; 2001.
10. Tamela FD, Ketchen JD Jr. “Organizational Configuration and Performance: The role of statistical power in extant research”. Strateg Manage J 1999;20:385-95.

How to cite this article: Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. Ann Indian Acad Neurol 2011;14:287-90.

Received: 12-08-11, Revised: 16-09-11, Accepted: 14-11-11

Source of Support: Nil, Conflict of Interest: Nil