

Clinical Research Methods

A step towards quality medical research

A. MATHEW, N. S. MURTHY

ABSTRACT

Only well planned and conducted research can answer questions that arise in day-to-day clinical practice. Unfortunately, most studies in India are deficient in design or execution. The results thus obtained are biased and cannot be extrapolated to the population studied. Some important questions which arise during the statistical design of investigations are selection of subjects on which observations are to be made, number of observations falling into different groups, allocation of subjects among different groups, etc. Besides planning and execution, appropriate methods need to be chosen for data analysis beforehand. If the study has been well designed and correctly analysed, the interpretation of results can be fairly simple. The present communication focuses on some issues relating to the planning of investigations in medical research with emphasis on statistical design.

Natl Med J India 1998;11:283-6

INTRODUCTION

Medical research is increasingly becoming a matter of teamwork. The design and conduct of investigations in medical research needs to bring together clinical experience, biological understanding, statistical expertise and other special skills, which vary from one study to another.

The structure of a research investigation can be broadly classified into three phases: (i) planning and design of research project, (ii) execution of the investigation, and (iii) analysis and meaningful interpretation of the results.

Careful study design forms the foundation of quality medical research. It is important to devote adequate time for careful and thorough planning of any investigation so that the objectives of the study are fully achieved. Planning the whole study should always precede execution. Statistical design of medical research means planning the study in a scientific manner so that not only are the objectives of the study fulfilled but also meaningful interpretation of the data collected during the study. The right time to seek statistical advice begins from the planning stage and continues till the evaluation and interpretation of results are completed.

TYPES OF RESEARCH INVESTIGATIONS

The various types of investigations carried out in the biomedical field are shown in Fig. 1. Research investigations are classified

into two types: observational and experimental. In observational studies, as the name suggests, the investigator is a passive observer, interfering as little as possible with the phenomenon he wishes to record; collecting information on the attributes of interest without influencing events and outcome. On the other hand, execution of experimental studies involves a planned interference with the natural course of events so that its effects on the natural course of the disease can be observed.

Observational studies

Observational studies include prevalence, incidence, case-control and cohort studies. Prevalence and incidence studies are designed to obtain estimates of certain simple characteristics of the population. These investigations focus on describing the distribution of diseases, which populations do or do not develop a disease, its geographical distribution, and how the disease frequency varies over time (trend). Prevalence studies are also known as cross-sectional studies in which individuals are observed only once at a given time. Incidence studies, on the other hand, are longitudinal in nature, investigate changes over time, and record observations on more than one occasion in the same population.¹⁻³

Case-control and cohort studies are conducted to ascertain the associations between certain variables with outcome of interest or to identify the causal factors. These studies focus on the determinants of a disease by testing the hypothesis formulated from earlier similar descriptive studies. In these studies, the researcher collects information on the attributes or measurements of interest but, as in prevalence and incidence studies, does not influence events. In a case-control study, the number of subjects with the disease in question (the cases) are identified along with some unaffected subjects (controls). The past history of these groups in relation to exposure(s) of interest is then compared. In contrast, in a cohort study, a group of subjects of interest is identified and then followed prospectively till the outcome of interest occurs or for a specified time period.^{1,4}

Experimental studies

Experimental studies include clinical trials and animal and laboratory studies. These studies are classified into prophylactic and therapeutic trials and interventional studies. These are usually carried out to assess the effects of various treatments on different experimental groups. The distinguishing feature of experimental studies is that there is planned interference by the investigator. Experimental studies often provide the most reliable evidence regarding the hypothesis being tested. The crucial question in experimental studies is how to allocate the intervention to the available units or subjects. It is necessary to arrange the groups being compared. The groups are made similar and as alike as possible except for the features of direct interest (matching). Ethical considerations form an important issue in the finalization of all studies.¹

Regional Cancer Centre, Thiruvananthapuram 695011, Kerala, India
A. MATHEW Division of Cancer Epidemiology and Cancer Research
Institute of Cytology and Preventive Oncology, Maulana Azad Medical
College Campus, Bahadur Shah Zafar Marg, New Delhi 110002, India
N. S. MURTHY Division of Biostatistics

Correspondence to N. S. MURTHY

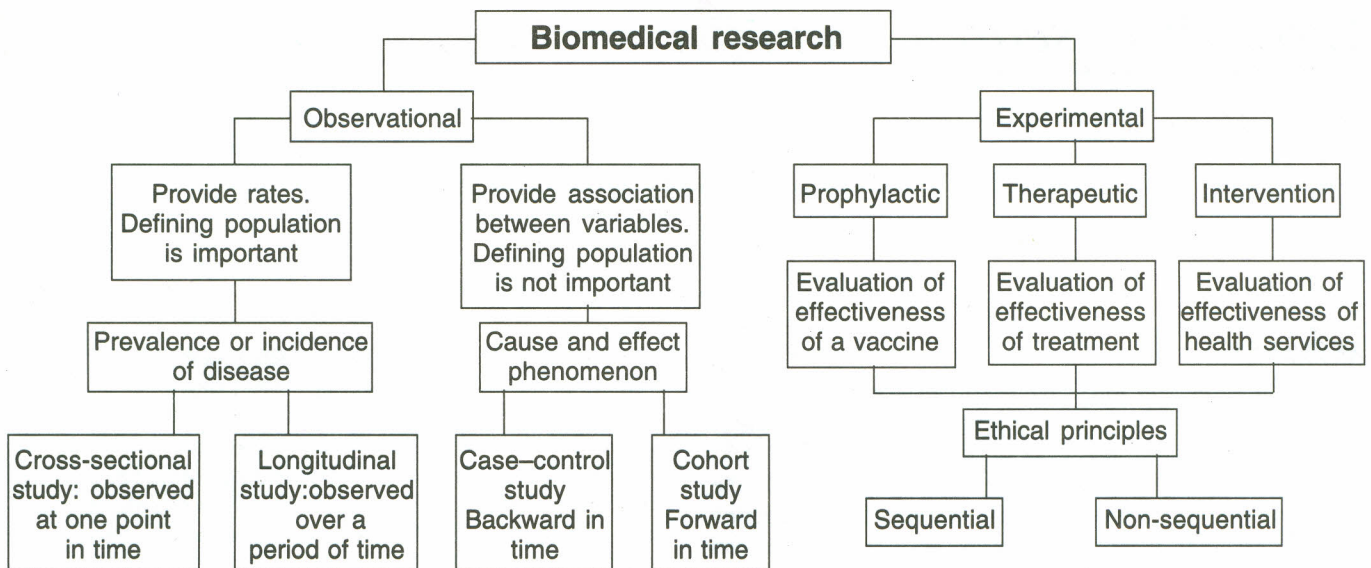


FIG 1. Types of research studies

STEPS IN PLANNING A STUDY¹⁻⁶

Selecting a problem for research

The most important step in the formulation of a research protocol is to define the problem to be studied. For this purpose, one has to study the patterns and changes which have already taken place, understand what has already been done, what has yet to be done, what are the priorities and prior knowledge regarding the problem. This can be achieved by a detailed study of the available literature and discussions with peer groups.

Formulation of objectives and hypothesis

After selecting a problem, the aims and objectives of the study need to be defined through statements in a precise and clear manner. Such statements are known as hypotheses. Ultimately, these hypotheses will be proved or disapproved on the basis of the results of the proposed research.

Area of study

After selecting a problem, the study area has to be selected. In a community-based study, the geographical area for the study could be a district, a state or a smaller unit like a *panchayat* or ward. In certain circumstances, a special group can be further selected, e.g. a religious group, or people belonging to one profession, etc. In a hospital-based study, on the other hand, the contributing departments or hospitals have to be identified prior to commencing the investigation.

Duration of the study

Temporal trends over years or seasonal trends over months have to be kept in mind while designing a study. The duration for a particular type of study has to be decided, depending upon the problem. Questions such as duration of recruitments of subjects into the study and the follow up time, the schedule of follow up, the rationale for a particular schedule, etc. need to be defined. For studies involving infective agents, a shorter period of observation usually suffices while investigations into causal factors may require years of observation.

Design of the study

The study design varies according to the type of research. In most cases, it is tailored to suit a particular study. In clinical and paraclinical fields studies are usually experimental, while observational studies are more frequently done in community- or hospital-based settings. A therapeutic study can be a two- or three-arm randomized or non-randomized, blind or double-blind trial. Similarly, observational studies must also be designed carefully. The help of a statistician or a research methodology expert is particularly valuable at the designing stage. Fallacies in study design can render extremely good data worthless.

Investigators of the study

As medical research usually involves a group of people, it is beneficial to assign specific tasks to each person at the beginning of the study. Identifying key personnel in certain areas helps in the smooth execution of the research project.

Subjects to be studied

The subjects for the study may suffer from a disease or different grades of the disease such as mild, moderate or severe. Alternatively, they may just be residing in an area chosen for the study. The inclusion and exclusion criteria of subjects for the study need to be defined and adhered to rigidly. Any violation of this ultimately leads to exclusion of that subject or set of observations from the analysis. This can lead to biased and invalid results.

Selection of variables

On the basis of the defined objectives and hypothesis to be tested, a list of relevant variables, including the co-variables pertinent to the problem, should be defined. While choosing the variables it is necessary to assess the feasibility of collecting the data in a reliable and accurate manner. The tools and equipment required for collecting the information are dependent on the variables selected for the study.

Designing tools for data collection

Experimental study designs are based on definite methodologies,

while observational studies are carried out using questionnaires or proformas for eliciting data. All the subjective questions used in the proforma such as satisfactory or unsatisfactory, exposed or unexposed need to be standardized to obtain uniform answers from all the respondents. The conditions being studied must be defined explicitly. It would be ideal to adopt standardized definitions for various terms such as severity of a disease, stage, distribution, etc. It is also best to test the tool developed in a pilot study.

Sample design

In medical research, the aim is to make general statement based on observations about a wider set of subjects. The set of all subjects (human or animal under investigation) is called the population of interest. Some examples include institutional populations, armed personnel, factory workers, school children, a population suffering from a disease.

However, it is usually not possible to cover a large population and hence sample of observations or subjects is selected. This is much more important in observational than in experimental studies. The sample chosen should permit a satisfactory estimation of the characteristics of the population with a certain degree of precision. A representative sample from a population can be obtained only by a random sampling method. Information obtained from such a sample can be generalized to the whole population.

The various methods of selecting a representative sample are:

1. *Simple random sampling*: In this technique, each unit has the same probability of being selected. Usually, random samples are selected by using random number tables. When the population is homogeneous with regard to the characteristics being studied, this method can be used.
2. *Stratified sampling*: In this method of sampling, the population is first divided into subgroups or strata, according to some characteristics, and a random sample is then taken from each subgroup. When the population is known to be heterogeneous and or when the heterogeneity is thought to have a bearing on the characteristics to be studied, this method is used.
3. *Systematic sampling*: In this method of sampling, the first unit of the sample is selected at random and the subsequent units are selected in a systemic manner. If there are 'N' units in the population and 'n' units are to be selected and if 'K' is the quotient obtained after dividing N by n, then one number is selected at random out of the remainder obtained after division. The other units are obtained subsequently by the addition of K to the previously selected number. K is known as the sampling interval.
4. *Multistage sampling*: In this type of sampling, sampling units are selected at various stages. For example, if the prevalence of a disease is to be estimated, a sample of village may be selected at random in the first stage, out of these selected villages, a random sample of houses can be selected in the second stage and out of these sampled houses, a random sample of individuals may be selected in the third stage.
5. *Cluster sampling*: A cluster of individuals (villages, localities, etc.) are treated as units and a sample of these units are drawn and observations for each member of every sample cluster are recorded.

The choice of method depends upon the type of study.

Sample size

This has to be determined to obtain a meaningful result. The

greater the sample size or the larger the experiment, the more precise will be the estimates and their differences. An increase in sample size costs more money and time.

Calculation of the sample size involves knowledge of the following factors:

1. an approximate idea of the characteristics being studied,
2. variation of the estimate within the population,
3. precision required for the sample estimate,
4. level of probability of the chance factor in bringing out the difference for the required precision,
5. the sample design to be adopted, and
6. practicability, availability and cost of investigations.

In a field survey to estimate the prevalence rate of a disease, the sample size is calculated by the formula:

$$n = \frac{4pq}{L^2}$$

where 'n' is the required sample size, 'p' the approximate prevalence rate of the disease for which the survey is being conducted, 'q'=1-p and 'L' the permissible error in the estimate of p. p is obtained from previous surveys or from a pilot study. The above formula has been worked out for a probability level of 0.05. The prevalence rate 'p' estimated from the above sample size will have an error of 'L' and is true in 95 out of 100 samples.

Randomization

An important step in experimental studies is the allocation of treatment strategies to different groups or individuals in a random manner. This avoids a possible source of bias, which may be due to a variation in features that are not part of the study design. Randomization is one of the key elements in research studies and the sheet anchor of all statistical designs. Random allocation can be done by using computer programmes, random number tables or even by tossing a coin or drawing chits.

To avoid a bias due to the observer's judgement or patient's knowledge of the type of treatment being instituted, it is desirable to keep both patients and observers ignorant of the treatment being given. This procedure is called 'blinding'. If patients alone are ignorant of the treatment being given, it is called 'single blinding' and if both the patients and observers are ignorant of the treatment, it is called 'double blinding'.

Selection of controls

An appropriate control group is required for comparison in both experimental and observational studies. It may not be possible to evaluate the effect of a new therapy or procedure without a proper control group. The presence of a control group strengthens the inferences that can be made from the results of the study. In randomized clinical trials, the controls selected are as similar as possible to the cases, except that they do not have the disease being investigated. The different types of controls that can be used are:

1. *Retrospective controls*: Patients with the same disease, who have been treated by a different method in the past by the same or some other worker (historical controls), can be used for comparison. However, a number of factors may have changed during the time of the two studies and hence these are not ideal.
2. *Patient as his own control*: All patients will not react equally to any drug. This variability amongst different patients to the drug can be overcome by recording the results in the same patients before and after treatment in certain types of investiga-

tion. In using such a control the basic assumption will be that the patient will not have variations over a period of time with regard to response to the drug, and that there is no carry over effect. By this type of control, it is possible to remove 'between subject variation' by studying changes from a baseline level.

3. *Concurrent controls*: For assessing the results of a new treatment for a group of patients, a control group receives only a placebo or another drug. Such controls are the best, as biases that may arise by choosing the previous two types are eliminated.

It is advisable to have controls in observational studies also. When cases are hospital patients with a particular condition, it is common to take controls in the same hospital(s) with different conditions. However, the use of hospital controls may lead to an underestimate of the relationship in such situations. Hence in case-control studies, selection of an appropriate control group is difficult. Healthy controls from the population under study are the best.

Investigational errors and quality control

Mainly two types of errors can occur in the course of a study. The first type of errors are response errors. These arise due to under coverage of the sample, non-cooperation of patients or due to the death of patients during the study period. Non-response can be a major problem in many experimental and cross-sectional studies. Such errors can be reduced by persistent efforts of the investigator and proper planning of the sample size which takes into consideration these eventualities. The second type of errors are measurement errors, or faulty techniques of recording. Laboratories tend to differ in their standardization of chemicals, reagents and techniques. This leads to different results. Differing methods of assessment are likely to give differing results. These can be reduced by standardization of techniques and proper training of personnel. A consistency in standards of observation is crucial to any comparative study. These standards must be monitored so that the deviations can be detected and remedied.

In case-control studies, another common source of bias is due to a differential recall by the cases and controls, especially when retrospective information is obtained by interviewing the subjects. As patients may have thought about a possible link with their past behaviour they can easily recollect information. Efforts should be made at the time of interview to minimize the recall bias.

Recording

The information collected during a study has to be recorded on predesigned cards, proforma or schedules. The design of records

depends on the facilities for analysis. Using a computer, data can be stored for easy retrieval and analysis.

Method of data processing and analysis

It is necessary to decide how the collected data would be processed, whether manually or through computers. Various tables that are required for interpretation should be worked out at the planning stage of the investigation including the type of analytical methods which would be used.

Presentation and publication

A decision to present the data at conferences and publish it in a biomedical journal should be taken jointly by all the authors. It is also essential to be clear at the beginning of the study on the authorship of the articles that may be published from the collected information will be acknowledged. All major biomedical journals follow a set of uniform guidelines according to which an author is one who (a) provides concept and design or analysis and interpretation of data, (b) drafts the article or revises it for important intellectual content, (c) gives final approval of the version to be published.⁶

CONCLUSION

Well designed and conducted investigation alone can prove or negate a hypothesis that is being tested. Statistical techniques cannot rectify mistakes due to careless or dishonest recording of data or faulty planning. The data should be collected honestly and sincerely without pre-conceived ideas about the outcome of interest.

ACKNOWLEDGEMENT

The authors wish to thank Dr Manoj Pandey, Regional Cancer Centre, Trivandrum for his critical comments and suggestions in the preparation of this manuscript.

REFERENCES

- 1 Altman DG. *Designing research: Practical statistics for medical research*. London: Chapman and Hall, 1989:74-103.
- 2 Armitage P. *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications, 1983:167-88.
- 3 Hennekens CH, Buring JE. Design strategies in epidemiologic research. In: Mayrent SL (ed). *Epidemiology in medicine*. Boston: Little, Brown and Co., 1987: 16-20.
- 4 Juneja A, Agarwal SS, Murthy NS. Designing of case-control studies. Rao V. (ed). *Biostatistics: A manual of statistical methods for use in health, nutrition and anthropology*. New Delhi: Jaypee Brothers, 1996:404-18.
- 5 Rao NSN. *Elements of health statistics*. Varanasi: Tara Book Agency, 1989:89-101.
- 6 International Committee of Medical Journal Editors. Uniform requirement for manuscript submitted to biomedical journals. *BMJ* 1988;296:401-3.