(6) Preliminary Work done

While writing the protocol it is always necessary to include a write up on the work already done by the investigator in the topic of the research project. For funding projects this is very much necessary to justify the importance & relevance of the topic based on the results obtained from the preliminary work done and to assess the expertise of the investigator in the topic of research on which the protocol is prepared. This need not be based on a planned project w.r.t. sample size requirements and other aspects of a full fledged research project. The investigator can highlight the results obtained from the preliminary study and justify its importance and relevance and his/her knowledge in the area of research and the facilities available for conducting the research and the shortcomings, if any, in the conduct of the research work done, suggesting the possible ways and methods of improving them and thus making them scientifically acceptable. One of the important questions asked in this context is the availability of budget required for carrying out the preliminary study. Normally this has to be arranged by the investigator himself/herself, either from the Institute, Organization in which he/she is working or from some other possible sources. Some Institutes like All India Institute of Medical Sciences, New Delhi has some provisions for financial assistance to the faculty / researchers in AIIMS for carrying out the preliminary work on the chosen topic. Many funding agencies / Organizations like ICMR, WHO, DST and DBT include this requirement in their Form meant to be used for requesting for financial assistance for carrying out research projects.

(7) Justification of the study

The Principal Investigator (PI) has to convince the Funding agency with proper justifications for the study he/she would like to carry out and financial assistance is requested for from them. Advantages, both specific and general w.r.t. the population of patients/subjects from which appropriate samples are to be included in the study have to be highlighted. Benefits in terms of the effect of the treatment and its implications on health, social and economic components have to be explained clearly. In other words, proper justification has to be given for the funding requested for from the Funding Institution. In this context one of the question which may be asked by the Funding agency could be that whether any other similar study was going on or not and if yes, whether the requested study could be merged with the ongoing study in any way. Though this may not be practically feasible due to a variety of reasons it might be relevant to include proper reasons for not being possible to merge the requested project with any ongoing similar studies. Appropriate reasons for convincing the Funding agency that the requested project is essential to be carried out independently will help a lot to convince them to consider for financial assistance to the requested project.

(8) Study Designs

Identifying and selecting the appropriate Study Design is very important in any research project. Many researchers confuse the study design with the sample size. Study design and sample size are entirely different. Study design is the method by which a plan is designed keeping in view the objectives of the study. It enables us to estimate the magnitude of the problem (prevalence of complete blindness), to identify / confirm its causative factors or to find out appropriate treatment modalities. Selection of the appropriate study design is one of the major components of any research study. A description of the study designs is given below:

The very basic study design is case history. For example when the first case of HIV was detected the doctor described the case and published it. That is the starting point. When a group of HIV cases were identified - Case series analysis was planned which describes the cases available in summary forms, like mean age, sex ratio, percentage of cases in different socio-economic groups etc. These two designs are planned in the very preliminary stages of the study when the first patient and later a series of patients of the same disease were spotted. The classical designs for the different types of studies are explained below:

Basically there are two types of designs:-

(A) Non-intervention / Observational/Descriptive studies: Observes, Collects and Describes

(B) Intervention / Experimental studies: Intervene with some intervention material / method - measures the outcome and compares between groups

(A) Non-intervention Studies

(a) Explorative Studies:

Such studies are Small studies with a shorter duration and are
planned when very little is known about the problem (Case study, Case series analysis, explained above)

**b) Descriptive Studies:**

This design is planned for the systematic collection of data on a disease / condition (for example, diabetic retinopathy) to get a clear picture of the problem, which will enable the researcher to estimate the magnitude of the problem, to identify the possible causative factors or to find out whether the different lab parameters are correlated or not with the clinical parameters. These types of studies fall under the following designs:

**Cross-sectional studies:**

Planned to estimate the Prevalence (total cases prevalent at a specified period of time), Incidence (new cases identified during a specified period of time) of the diseases / conditions under study or to study whether different lab parameters are correlated or not with the clinical parameters, socio-demographic and behavioural variables (habits of smoking, use of alcohol, lifestyle etc.) or to study Knowledge, Attitude and Practice (KAP) w.r.t. any problem, say, viewing TV continuously for a long time and the knowledge, attitude and practice w.r.t. its harmful implications.

Diagrammatic representation of the cross-sectional design is given below.

```
Population
     ↓
  Sample

Disease present
  ↓
Risk Factor present

Disease absent
  ↓
Risk Factor absent
```

Certain number of subjects (to be estimated based on the results observed in the past studies and the amount of confidence and precision required and using a statistical formula), Principles & methods of estimating minimum sample size for the study depending upon the design and other aspects of the study will be explained in a separate chapter, in a later issue. Appropriate statistical test can be applied for the 2 by 2 (as indicated in the diagram) table to find out whether the role of the risk factor studied is statistically significant or not.

**c) Comparative OR Analytical Studies:**

Though the cross-sectional studies enable us to identify / attempts to establish the possible causative factors for the occurrence of a disease / condition, its role in the causation of the disease cannot be confirmed from such studies. It only helps us to determine the frequency & burden of the disease under study and to generate research questions and develop appropriate research hypotheses: For confirming the role of the specific causative factors, comparative studies have to be designed. Two most important designs adopted for this purpose are:

1. Case-Control Studies
2. Cohort Studies

**1) Case-control studies**

In case-control studies, a certain number of cases and non-cases are selected from the corresponding populations and data on the presence / absence of possible causative factors among them are collected. For example, cases may be children with low vision and control may be children with normal vision. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. Principles & methods of estimating minimum sample size for the study depending upon the design of the study and other aspects of the study will be explained in a separate chapter, in a later issue. The diagrammatic representation of this design is given below.

Statistical significance of the role of the risk factor can be studied by applying appropriate statistical analysis. The advantages of case-control studies are: results can be obtained quickly, less expensive and very useful in case of rare diseases. Main disadvantage is that only odds ratio for the risk factor can be estimated and not the relative risk. In any study on the risk factors of the disease it is ideal to find out the relative risk of the causative factor. It means the risk
in terms of the number ) a subject with the risk factor gets the disease compared to a subject without the risk factor. For example if the relative risk is estimated as 5.0, it means that the risk for a subject viewing the TV for longer duration daily getting eye problem is 5 times more than a subject viewing the TV for a shorter duration. However, in case-control studies what can be estimated is only Odds, which is less a powerful statistic compared to relative risk.

(2) Cohort studies

In case of Cohort design, a certain number of subjects with the factor (say, those who view TV continuously for long duration) and a certain number of subjects without the factor (those who view TV for shorter duration) are selected from the corresponding populations and they are followed up for a certain period and see how many in each group develop the disease/condition. For example, two groups of children—one group of children viewing TV continuously for a long time daily and another group viewing TV only occasionally. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

\[ \text{(A) Population with the presence of the Risk factor} \]

\[ \begin{align*}
\text{Sample} & \\
\text{Risk Factor present} & \text{Risk Factor absent}
\end{align*} \]

\[ \text{Follow up period} \]

\[ \begin{align*}
\text{Disease occurs} & \\
\text{Disease does not occur}
\end{align*} \]

\[ \text{(B) Population with the absence of the Risk factor} \]

\[ \begin{align*}
\text{Sample} & \\
\text{Risk Factor present} & \text{Risk Factor absent}
\end{align*} \]

\[ \text{Follow up period} \]

\[ \begin{align*}
\text{Disease occurs} & \\
\text{Disease does not occur}
\end{align*} \]
The advantages of the cohort design are: It confirms the role of the risk factor with more strength (relative risk) compared to Odds ratio in case of case-control studies. However the major disadvantage is that if the end point occurs after a long time, results will be available only after a long time and depending upon the length of the follow-up cohort design may not be feasible. For example in case of studying the role of smoking in the causation of lung cancer the follow-up time will be very long and in such diseases it may not be feasible to conduct the cohort study. Also, since there is a follow-up period, the problem of drop outs may be there and exposure status, at the time of enrollment in the study, may change over a period of time. Hence depending upon the requirement, the purpose of the study, time period and finance available for the study and the knowledge already available on the confirmation of the role of the risk factor in the causation of the disease, appropriate design (cross-sectional, case-control or cohort) design can be planned.

Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

```
Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

```

```

Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

```

Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

```

Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:

```

Guidelines which may be used for selection of the appropriate design (either case-control or cohort) are:

1) If no or very less information is available on the magnitude of the disease and its possible causative factors—CROSS-SECTIONAL STUDY

2) In case of Rare diseases - Longer the interval between suspected cause and outcome - Financial constraints --- CASE CONTROL STUDY

(B) Intervention (Experimental) Studies

(a) Clinical trials

This type of design is planned to prove the efficacy of an intervention or to establish a statistically and clinically better intervention method (treatment with a drug, surgery, health advice like exercise or changing the life style such as following a particular diet pattern). Essential features of an intervention design are: Component of subjects (say, patients), intervention material as indicated above and the response such as cured / not cured or improved / not improved or alive / expired or a decrease or increase in the value of a parameter. Another important part of the intervention design is that there will be a control group with no intervention to anybody in that group. For example, a clinical trial may be planned to compare the efficacy of a specific eye drops compared to washing the eyes with clear water for testing the efficacy of the eye drops for curing conjunctivitis. Sample size in each group has to be estimated based on the results observed in the past studies and the amount of confidence and power required and using a statistical formula. The diagrammatic representation of this design is given below:
(b) Prevention (Prophylaxis) trials

This design is adopted to study the efficacy of a method in the prevention of the occurrence of the disease. Classical example of prevention is through vaccination, immunization or health education. This design can also be used in assessing community health programmes such as nutrition, use of contraception and use of alcohol, & use of hard drugs and smoking.

The basic difference between clinical trials and prophylaxis trials is that while the former one is normally done in a hospital set-up, the latter one is done in a community set-up, usually on a much larger number of subjects.

(9) Estimation of Minimum Sample size & the method of selection

What should be the sample size in my study - a common question asked by any researcher. The answer for this is not very simple, for any researcher. Many researchers think that the sample size for their study can be readily got from a statistician just by asking for it – just like a commodity from a shop. But, the researcher should know that it is not a magic number and also not a universal figure. Sample size varies from study to study depending upon many criteria. Some researchers think that 30 will be adequate sample size for any study. This number may be ok theoretically based on a statistical theorem (Central limit theorem), but, not alright for a research study. Sample size of 10 may be adequate for some studies, but, 3000 may not be adequate for another study. The pertinent question is how much should be the small portion is determined by computing the Minimum Sample size. Sample size has to be estimated based on several information such as, the design of the study, type of study variables (whether measurable (pulse rate, weight, BP, eye pressure etc.) or categorical (presence/absence of cataract, blind/not blind, improved/not improved etc.), whether the study aims at only for estimation of a parameter (prevalence of blindness, mean value of eye pressure) or it aims at testing the statistical significance of a research hypothesis (Prevalence of cataract is significantly higher in diabetics than in non-diabetics) and the required precision, confidence & power of the test (The concept of precision, confidence & power of the test and the method of estimating the minimum sample size for any particular study, depending upon the study design, type of study variable, precision, confidence & power required will be explained in a later chapter).

Once the minimum sample size is estimated the next question is how to select the required number of study subjects from the population of subjects. Population for a research study is defined as the total sampling elements (units) in the defined area at a particular period. For example, for the estimation of the prevalence rate of visual problems in school children (5 to 15 years of age) in Kerala, all the school children in Kerala studying in the schools at that time period form the population. A small portion of the population which truly represents the population with respect to the study characteristics is called ‘sample’. Most of the research studies are concentrated on samples than to the complete population due to the reasons such as much higher expenses, the time it might take to complete the study covering all the children and the feasibility of the study. The required number of study subjects have to be selected from the list of subjects in the population called sampling frame by random sampling method. There are many types of sampling methods which are commonly used for the selection of sample from the population, some of them for increasing the precision of the estimate (reducing the error in estimation) and some others for convenience in the community studies. These will be explained in detail in a later chapter.

(10) Research Tools

The tools to be used in the study should be explained clearly in the protocol. They could be a set of questionnaires, proformae, lab tests, clinical examination, investigations like X-ray, MRI, Snellen chart for measuring visual acuity, stereopsis, slit lamp, tonometry etc.

(11) Standardization of research tools

The tools which are going to be used in the study should be defined and explained clearly without any ambiguity. If more than one researcher / technician is going to be included in data collection / recording (for example, collaborative studies encouraged by ICMM), training of them should be held, preferably in a common place with the support of written guidelines. For example, assume that one item in the proforma in a community study is ‘income’. If it is not defined clearly, it could be taken as (a) income of only the head of the household, or (b) income of all earning members in the household, or (c) income of all earning members in the household + income from agriculture, property etc. None of them will give the correct indication of the economic status of the family. The correct method of getting the required information will be defining per-capita income, since the economic status of the family depends upon, not only on the total income, but also on the number of members in the family. Similarly all the equipments & instruments which are going to be used in the study should be checked for any defect and the same should be rectified and calibrated so as to get the correct values on the test. Hence standardization of
each & every aspect of research tools is very important in any research study,

(12) Data Collection Methods

Basically there are two types of data collection methods; Primary Methods / Secondary Methods

(1) Primary Methods:
   (1) Questionnaires
   (2) Proforma
   (3) Clinical Examination
   (4) Recording of Laboratory, Social, Demographic and Behavioural Parameters

(2) Secondary Methods

Secondary data (Published reports and papers, Annual reports Hospital, Census data, Doctors' Clinic Records, school health records & reports, Communications from the concerned experts etc.)

It is always preferable to collect the data by primary method. i.e., the investigator plans the study and collects /records the data directly based on a defined plan & format, which will be more scientific since the investigator has full control over the methods and correctness of information collected. But, in case of secondary data, the reliability of information recorded in the report can be questioned since they are collected by another person / organization for a different purpose. Also, utmost care should be taken while collecting data on sensitive information like, use of drugs which are abused, use of alcohol, sexual & contraception habits etc. It is more difficult to establish the reliability of information on such aspects.

(13) Data Analysis Plan & Methods of statistical analysis

A paragraph indicating how the data is going to be analysed has to be given in the protocol. Due to the easy access to the computers and statistical softwares, data analysis is normally done using the computers. First of all, the data collected / recorded in the proforma / questionnaires should be carefully checked for any mistake in entering the information. Also, care should be taken while entering the data in a specific format consistent with the analysis in the computer. Data checking for consistency, abnormal values and wrong entries should be checked very carefully. For example, the age of the child in a studying a paediatric problem is 6 years and if it is wrongly entered as 60 years and if this is not checked & identified, the analysis will give wrong results. Methods of checking these aspects have to be clearly stated under data analysis plan.

Some investigators simply write that data analysis will be done using standard statistical methods. This is not sufficient. What specific methods are going to be used to achieve the various objectives of the study should be stated here. For example, it may be written as: (1) percentage prevalence rate of complete blindness will be computed (2) To test the statistical significance of the difference in the prevalence rates between poor and high socio-economic classes, chi-squared test will be done (3) To test the statistical significance of the difference in mean visual acuity between children from public schools & Govt., schools, student's 't' test will be done (4) To study the correlation of eye pressure and age, Pearson's correlation coefficient will be computed etc. The help of a statistician may be sought for writing these aspects.

(14) Consent form

An appropriate consent form is very essential for any research study, especially in experimental studies. The consent form should clearly give all the relevant details indicating the aim of the study, the method and material of intervention, possible side effects of intervention, the structure of intervention material and all other relevant details. This is a legal and ethical requirement in any intervention study. Even in observational studies, consent form is required to indicate that the study subjects do not have any objection in participating in the study by giving the required information and by subjecting themselves for any lab test. The consent form need to be got approved by the Institutional Ethical Committee.

(15) Expected outcome

Expected outcome of the study in terms of the benefits and, both direct and indirect, to the patients / common people, both immediate and in the long run should be given. For example, how the results on the magnitude of the problem w.r.t. the factors studied and identification / confirmation of the risk factors for the occurrence of the problem or how the confirmation of the efficacy of a new treatment modality in comparison to the existing standard methods is going to be beneficial to the patients specifically and to the Government, in general w.r.t. social and economic aspects should be spelt out.

(16) Logistics - Resources and Facilities available and required - Budget details

Detailed yearly budget for various components has to be given with proper justifications. Facilities available in the place of research study, in terms of expertise and infrastructure, should be given. Personnel, like research and
other staff required, non-consumable items like equipments & instruments and consumable items like pharmaceutical & chemical material, required with the estimated budget for the same should be clearly stated with full justifications. Budget for printing proformae / Questionnaires, travel and contingencies for stationery, local travel etc. should be given with proper justifications. Depending upon the funding agency certain amount may be included as overhead costs as per rules. Both Yearly and total budget should be given.

(17) Time Schedule

Time plan w.r.t. the different components of the study should be a part of the protocol. This should indicate how the total time period indicated for the study is going to be utilized for the different aspects of the study. This is very important to start the study at the correct time and to complete it within the time period indicated. The time period indicated for the various components should be strictly adhered to so that the study can be completed within the time indicated. For example, if the duration of the study is 2 years, the time plan could be as follows:

(A) --- 3 months
(1) Preparation of Study Tools (Proformae / Questionnaires)
(2) Procurement of Equipments/instruments
(3) Standardization of Study Tools / Equipments / Instruments

(B) Data collection / recording --- 12 months
(C) Data entry in the computer and data editing & cleaning --- 3 months
(D) Statistical analysis of data and interpretation of results --- 3 months
(E) Report writing --- 3 months

(18) References

Books / Publications referred in the protocol should be given in the standard format like Vancouver style or any other stipulated style of the funding agency. It would be ideal to include the latest publications / books rather than very old ones.

(19) Annexures

Appendices / Enclosures / Attachments such as Proforma(e) Questionnaires, Consent form Important Documents related to the study, Bio-data of the Investigators and any other relevant papers should be included in the protocol.

(20) Summary

A summary of the protocol highlighting the relevance and importance of the study and indicating the objectives and brief write-up on material & methods (study design, sample size, study population, study tools) and the expected outcome of the study should be given in the protocol.

Books for further reading:

(1) Practical guide for health researchers, WHO Regional publications Eastern Mediterranean Series-30, Mahmoud F Fathalla, WHO, Regional office, Cairo, 2004
(2) Medical writing - a guide to clinicians, educators & researchers, Taylor, Robert B, 2011
(4) Medical Statistics - Principles & Methods, Sundaram KR, Dwivedi SN, Sreenivas V, BL publications, Delhi, 2009
(8) Clinical Epidemiology - The Essentials: Robert W. Fletcher, Suzanne W. Fletcher, Lippin cott Williams, 2005
(9) Statistics for Epidemiology: Nicholas P. Jewell; Chapman & Hall (CRC), 2004

Prof Sundaram was previously the Head of Biostatistics at All India Institute of Medical Sciences. Currently he heads the Department of Biostatistics at Amrita Institute of Medical Sciences, Kochi.