



**St. George's Hospital Medical School**

## **Statistics Guide for Research Grant Applicants**

[Back to Brief Table of Contents](#)

### **A. Describing the study design.**

- [Introduction](#)
- [A-1 Type of study](#)
  - [A-1.1 Type of study: Observational or experimental](#)
  - [A-1.2 Combinations and sequences of studies](#)
  - [A-1.3 Cohort studies](#)
  - [A-1.4 Case-control studies](#)
  - [A-1.5 Cross-sectional studies](#)
    - [A-1.5a Prevalence studies](#)
    - [A-1.5b The estimation of sensitivity and specificity](#)
    - [A-1.5c When to calculate sensitivity and specificity](#)
    - [A-1.5d Cross-sectional ecological studies](#)
    - [A-1.5e Studies of measurement validity, reliability and agreement](#)
  - [A-1.6 Confounding](#)
    - [A-1.6a Confounding or interaction](#)
  - [A-1.7 Experiments and trials](#)
  - [A-1.8 Randomised controlled clinical trials](#)
  - [A-1.9 Pilot and exploratory studies](#)
- [A-2 Follow-up](#)
- [A-3 Study subjects](#)
- [A-4 Types of variables](#)
  - [A-4.1 Scales of measurement](#)
  - [A-4.2 Types of data](#)
  - [A-4.3 Methods of data collection](#)
  - [A-4.4 Validity and reliability](#)

- [A-4.4a Validity](#)
- [A-4.4b Repeatability \(test-re-test reliability\)](#)
- [A-4.4c Inter-rater reliability \(inter-rater agreement\)](#)

## Introduction

The section of an application, which is called Plan of Investigation (sometimes called Subjects & Methods), is where applicants describe what they propose to do. The purpose of the investigation and the background are described elsewhere, and presumably establish that the research or development is a worthwhile idea in a deserving cause. Here, however, the application needs to provide details of their proposed methods and establish that the practical issues have been thought through. The reviewers want to know that the study is both methodologically sound and feasible, and that the applicants are capable of doing it (see [F-1](#)). Failure to discuss relevant practicalities will reduce the plausibility of the proposal.

The precise topics that should be covered depend on the type of study. So it makes sense to provide fairly early on an overall description of the study or experiment, preferably using some standard terms for study design. This is one area where there is no need to be wary of jargon, since all reviewers will be familiar with, and looking for an indication that the study is a cross-sectional study, cohort study, double-blind randomised controlled trial, or whatever.

[Back to top.](#)

## A-1 Type of study

### A-1.1 Type of Study: Observational or experimental

The most obvious distinction that can be made between studies is whether they are experimental or observational. Experimental studies are what their name suggests, experiments. They are studies in which the applicant has some control over the experimental conditions and the way in which groups of subjects for comparison are constructed. They also involve some sort of treatment or other intervention. Observational studies on the other hand are studies in which subjects are observed in their natural state. The groups of subjects that are compared are self-selected e.g. manual workers versus non-manual workers or subjects with and without disease. Subjects may be measured and tested (e.g. total cholesterol measured, disease status ascertained) but there is no intervention or treatment (e.g. patients allocated to different exercise programs, patients allocated to new drug or placebo). Observational studies include cohort studies, case-control studies, ecological studies, cross-sectional studies, prevalence studies and studies of sensitivity and specificity.

To illustrate the difference we will consider the following scenario:

Scenario A-1.1: One grant proposal containing in effect three different studies, two observational and one experimental. The applicants are interested in the aetiology and treatment of a disease affecting the knee. Briefly they plan to:

- a) Compare leg measurements between subjects with and without disease (Observational)
- b) Compare leg measurements between the symptomatic and asymptomatic leg of diseased individuals (Observational)
- c) Randomly allocate subjects with disease to treatment or no treatment and compare change in leg measurements over a period of 6 months between the two groups (Experimental)

For further information on types of study see [Bland \(2000\)](#), [Altman \(1991\)](#) p74-106, and also sections [B](#) and [C](#) of this handbook.

[Back to top.](#)

### **A-1.2 Combinations and sequences of studies**

Sometimes the proposed research and development is a programme consisting of a sequence or combination of overlapping studies with different study designs. Efficiencies in time and cost can be achieved in this way. To keep the plan of investigation clear, it will help if the design of the various studies can be separately described using in each case the appropriate jargon for study design. This can be quite hard to do in some cases. The promised savings have to be weighed against the decreased likelihood of success from added complexity. The applicant should give careful thought to whether his powers of description can make a complex programme seem simple. If not, there may be a better chance of a simpler study with intermediate objective being funded - on the grounds that the study is clearly feasible - rather than a programme of studies that is in principle more resource efficient, but less clearly feasible.

[Back to top.](#)

### **A-1.3 Cohort studies**

In a cohort study a population of subjects is identified by a common link (e.g. living in the same geographical area, working in the same factory, attending the same clinic) and information collected on the study subjects concerning exposure to possible causative factors. The population is then followed forward in time to see whether they develop the outcomes of interest. Cohort studies often occur where the exposures are potential risk factors for disease (e.g. smoking, high blood pressure) and the outcomes are the development of those diseases (e.g. lung cancer, IHD). For further information on cohort studies see [Breslow & Day \(1987\)](#).

[Back to top.](#)

### **A-1.4 Case-control studies**

A case-control study is one in which all the subjects with a given disease (or condition) in a given population (or a representative sample) are identified and are compared to a control group of subjects without the disease (or condition). They are compared in terms of information on potential risk factors, which is collected retrospectively. One of the problems inherent in case-control studies is how to select a comparable control group (see [C-1.1](#)). For example you might choose to take a random sample of those without disease from the population which gave

rise to the cases. However, this assumes that a list of subjects in the population (i.e. a sampling frame) exists. Sometimes one or more controls are matched to each case so that cases and controls are comparable in terms of variables such as age and sex. The variables used for matching are those that might influence the disease, provided they are not part of the pathway by which the risk factors under study are thought to influence the disease. However the use of matching tends to complicate the subsequent statistical analysis (see [C-1.2](#), [E-5](#)). Another problem inherent in case-control studies is bias due to the retrospective nature of the risk factor information (see [C-2](#) and [C-3](#)). These issues and more are discussed in [Breslow & Day \(1980\)](#).

[Back to top.](#)

### **A-1.5 Cross-sectional studies**

A cross-sectional study occurs where a population or sample of subjects is studied at a single point in time e.g. the 2001 census. A sample survey is an example of a cross-sectional study. One problem with a cross-sectional study is that it tells you little about the order of events e.g. which came first, disease or exposure? Special types of cross-sectional study include prevalence studies (see [A-1.5a](#)), cross-sectional ecological studies (see [A-1.5d](#)) and studies of sensitivity and specificity (see [A-1.5b](#), [A-1.5c](#)). For further information see [Altman 1991](#) p99-101.

[Back to top.](#)

#### **A-1.5a Prevalence studies**

A prevalence study is designed to estimate the prevalence of a particular disease / condition / characteristic in a population of interest. Prevalence studies are sample surveys where the primary aim is estimation. Clearly of major importance in this type of study is obtaining a sample which is representative of the population of interest (see [C-4](#) and [C-5](#)) and in making sure that results are not biased by a poor response rate (see [C-6](#)).

[Back to top.](#)

#### **A-1.5b The estimation of sensitivity and specificity**

These studies often arise when the aim is to evaluate the usefulness of a new screening technique. It is often the case that a new ELISA assay has been produced to detect disease and the applicants wish to compare the accuracy of this often quicker method of ascertainment with that of cell culture (i.e. the 'gold standard'). Two groups are selected; those with and without disease according to cell culture. The subjects are then tested using the ELISA assay to determine which are test positive and test negative. Sensitivity and specificity are then calculated as the proportion of those with disease that test positive and the proportion of those without disease that test negative respectively. These studies often require more subjects with disease than the applicant envisages (see section on sample size calculation) and the need to do the ELISA test 'blind' to the results of cell culture is often overlooked.

[Back to top.](#)

### **A-1.5c When to calculate sensitivity and specificity**

If the diagnostic or screening test being assessed is intended to become the first available prospective screening tool then determining the sensitivity and specificity against a gold standard will be a constructive contribution. On the other hand, if the test is a candidate to replace an established test, then both of these tests should be compared against the gold standard. The new test will be preferable if both sensitivity and specificity turn out to be superior to those of the established test. If one could be larger and the other smaller (either as estimated or within confidence interval), it is then necessary to weigh the costs (financial and other) of false positives and false negatives, before there is the basis for a practical recommendation to adopt the new test.

It is sometimes important to calculate the positive predictive value (PPV) as well as sensitivity and specificity. If the objective of a test is to identify a high risk group for whom special and rather "expensive" treatment will be offered, i.e. high cost to supplier or substantial downside to recipient, while the test negative group would continue to be offered the standard treatment, then the positive predictive value (PPV) is more relevant. This is the proportion of people testing positive, who are actually positive. If the PPV is low, then a substantial number of false positives may be unnecessarily worried by a potential diagnosis, or given expensive, unpleasant or time-consuming treatment they do not need. There is a tendency for the PPV to be low when the prevalence of the condition is low in the population being screened.

[Back to top.](#)

### **A-1.5d Cross-sectional ecological studies**

A cross-sectional ecological study is one in which we are looking at correlations between variables measured at a level higher than the one on which we want to make conclusions. For example investigating the relationship between leukaemia and radon by correlating the rate of leukaemia registration per million per year for several countries with their estimated average level of radon exposure over the same period ([Henshaw \*et al.\*, 1990](#)) i.e. the unit of analysis is the country and not the individual. This type of study is particularly prone to the effects of confounding (see [A-1.6](#) and [Lilienfeld & Lilienfeld 1980](#) p13-15).

[Back to top.](#)

### **A-1.5e Studies of measurement validity, reliability and agreement**

Some studies investigate the properties of measurement methods. This can include numerical measurements such as blood pressure, categorical observations such as health status, and questionnaire based measurement scales such as those for anxiety. A study of validity investigates the extent to which the measurement measures what we want it to measure ([Bland & Altman 2002](#)). Here the issues are whether there is a genuine gold standard or criterion by which the measurement method can be judged and if not how validity can be investigated. Reliability concerns the extent to which repeated measurements by the same method on the same subject produce the same result ([Bland &](#)

[Altman 1996](#), [1996a](#), [1996b](#)). These may be by the same observer or different observers (observer variation) and may investigate reliability over time or the effect on measurements of different parts of the measurement process. Particularly important here are the selection of measurement subjects and the number and selection of observers. A third type of study is of agreement between two methods of measuring the same quantity. Here we are concerned with whether we can replace measurements by one method with measurements using another method ([Bland & Altman 1986](#)). Several topics related to the design and analysis of such studies are discussed by Bland (<http://www.mbland.sghms.ac.uk/commeth.htm>).

[Back to top.](#)

## **A-1.6 Confounding**

There are many drawbacks associated with the different types of observational study but one that they all share is the potential for spurious associations to be detected or real associations masked due to the effects of confounding factors. Confounders are generally variables that are causally associated with the outcome variable under investigation and non-causally associated with the explanatory variable of interest. Thus an observed association between disease and a potential risk factor may simply be due to that factor acting as a marker for one or more real causes of disease. That is why you cannot conclude causality from an observational study. Confounding arises because in observational studies we are not always comparing 'comparable' groups. For more information see [Breslow & Day 1980](#) p93-108.

[Back to top.](#)

### **A-1.6a Confounding or interaction**

The term confounding (see [A-1.6](#)) should not be confused with interaction. An interaction occurs if the nature (i.e. magnitude and direction) of the association between two variables differs dependent on the value of some third variable. For example, the association observed between gender and current asthma may differ with age since asthma tends to be more common among males than females in childhood but not in later life ([Burr 1993](#)). We say that there is an interaction with age. In general we are interested in describing such interactions. A confounding variable by contrast, is a nuisance variable. In adjusting for or designing out a confounding variable we assume that it does not influence the association that it confounds. In other words, for any given value of the confounding variable we assume that the magnitude and direction of the association of interest is the same.

[Back to top.](#)

## **A-1.7 Experiments and trials**

Experimental studies where the aim is to evaluate the effectiveness of a new treatment or intervention are referred to as trials. If the study subjects are human with the same medical condition the term clinical trial can be used (Pocock 1983). However, whether the study 'subjects' are humans, mice or even administrative groups (e.g. general practices, clinics) the same design considerations apply (see [A-1.8](#) and [Section](#)

[B](#)).

In trials (e.g. clinical trials) we have the ability to manipulate the situation to ensure that groups are comparable. Uncontrolled trials i.e. those with a single treatment group and no other group to act as a control are to be avoided. Without a comparison group there is no way of knowing whether an overall improvement in outcome is due to the new treatment or would have happened anyway in the absence of the new treatment. A further discussion of why trials should be controlled and why subjects should be randomly allocated to groups is given in [A-1.8](#), [B-3](#), [B-5](#) and in [Pocock \(1983\)](#). For information on cross over trials and other similar designs see [B-5.10](#).

[Back to top.](#)

### **A-1.8 Randomised controlled clinical trials**

Randomised controlled trials are designed to compare different treatments or interventions. Subjects are randomly allocated to groups so that groups are comparable at the beginning of the study in terms of their distribution of potential confounding factors e.g. age and sex (see [B-5](#)). The treatments/interventions are then administered and the outcomes compared at the end of the follow-up period. There may be two groups or several groups. There may be one treatment group and one control group, or two treatment groups, or two treatment groups and 1 control group etc. The control group may receive a placebo treatment to aid blinding of treatment allocation from both the study subjects and those assessing outcome; although it may be considered unethical to have a control group receiving placebo, or an untreated control group, if a proven treatment is already in standard use ([Rothman 2000 et al.](#), [F-3.3](#)). If both the assessor and study subject are blind to allocation then this is known as double-blind. Single-blind means that one of the parties (i.e. study subject or assessor) is privy to information on allocation (see [B-4](#)). In [Scenario A-1.1](#), part (c), there is one intervention group and one control group. Since the intervention group consists of some sort of training for which a placebo is not easily constructed, the patient will be aware of the treatment allocation. However, the person making the leg measurements can be kept in the dark provided they are not told accidentally by the patient; a possibility which could be reduced by telling the patient to keep that information quiet.

[Back to top.](#)

### **A-1.9 Pilot and exploratory studies**

The term "pilot study" is often misused. A pilot is someone or something which leads the way. A pilot study tests on a small scale something which will be used on a larger scale in a larger study. Hence a pilot study cannot exist on its own, but only in relation to a larger study. The aim of the pilot study is to facilitate the larger study; it is not a study in its own right. Pilot studies may be used to test data collection methods, collect information for sample size calculations, etc.

A pilot study should always have a main study to which it leads. This does not mean that funding cannot be sought for a genuine pilot study apart from the main study for which it leads. It may be that full funding cannot be sought until some pilot information is obtained, perhaps relating to sample size or feasibility of data collection.

A pilot study does not mean a study which is too small to produce a clear answer to the question. Funding organisations are unlikely to fund such studies, and rightly so. They are poor research, conducted for the benefit of the researcher rather than society at large (which pays for the grant).

Not all small studies are unjustified. It may sometimes be that an idea is at too preliminary a stage for a full-scale definitive study. Perhaps a definitive study would require a multi-centre investigation with many collaborators and it would be impossible to recruit them without some preliminary findings. It may be that where no study has ever been done, there may be insufficient information to design a definitive study. A smaller study must be done to show that the idea is worth developing. What should we call such a study? A pilot leads the way for others, someone who boldly goes where no-one has gone before is an explorer. We need to explore the territory. Such a study is an exploratory study, rather than a pilot, because we do not at this stage know what the definitive study would look like.

[Back to top.](#)

## **A-2 Follow-up**

Many studies including most cohort studies and randomised controlled trials are prospective i.e. they have a period of follow-up. Surprisingly the length of proposed follow-up is often a piece of information that grant applicants leave out of their proposal. It may be stated that measurements will be repeated every 3 months but without information on the total length of follow-up, this tells us nothing about the number of measurements made per patient. Information on length of follow-up is often crucial in assessing the viability of a project. For example, let us suppose that when describing a proposed randomised controlled trial of a treatment for a particular cancer, an 80% recurrence rate is assumed for the untreated group and this figure is used in the sample size calculation (see [D-8.2](#)). If the figure of 80% relates to recurrence over 5 years the calculation will yield the appropriate sample size for a 5-year study. However, if the proposed length of follow-up is only 2 years the resulting study will be hopelessly under-powered. Length of follow-up is also important in trials where the effects of the intervention such as an educational intervention, are likely to wear off over time. In this situation, assessing outcome only in the immediate post-intervention period will not be very informative.

[Back to top.](#)

## **A-3 Study subjects**

It is important to know where the study subjects come from and whether they are an appropriate group to study to address the research question of interest. For example if a disease is most prevalent in the young why is the study based on the middle aged? It is of interest to know how the study subjects will be selected. For example are they a random sample from some larger population or are they all patients attending a clinic between certain dates? It is also important to specify any exclusion/inclusion criteria. The applicants should also state how many subjects will be asked to participate and how many are expected to agree. Remember, the sample size is the number that agree to participate and not the number that are approached.

[Back to top.](#)



## A-4 Types of variables

It is important to describe both the outcome and explanatory variables that will be investigated in the proposed study by specifying the type of data and the scale of measurement (see [A-4.1](#), [A-4.2](#) and [Bland 2000](#)). It is this sort of information that will help determine the nature of any statistical analysis as well as the appropriate method of sample size calculation.

[Back to top.](#)

### A-4.1 Scales of measurement

- i) Interval scale: data have a natural order and the interval between values has meaning e.g. weight, height, number of children
- ii) Ordinal scale: data have natural order but the interval between values does not necessarily have meaning e.g. many psychological scores.
- iii) Nominal scale: categorical data where the categories do not have any natural order e.g. gender (male / female)

[Back to top.](#)

### A-4.2 Types of data

Quantitative data: data measured on an interval scale

- i) Continuous data: variable can take all possible values in a given range e.g. weight, height
- ii) Discrete data: variable can take only a finite number of values in a given range e.g. number of children

Qualitative data: Categories, which may or may not have a natural, order (i.e. measurements on nominal and ordinal scales).

[Back to top.](#)

### A-4.3 Methods of data collection.

The quality of a study depends to a large extent on the quality of its data. The reviewer is therefore interested in how the applicants plan to collect their information. If they propose to use a questionnaire, how will the questionnaire be administered, by post (or otherwise delivered for self-completion) or by interview? The use of an interviewer may aid the completeness of data collection but bias may be introduced in some studies if the interviewer is not appropriately blinded e.g. to case / control status in a case-control study (see [C-2](#)) or treatment group in a clinical trial (see [B-4](#)). If the applicants propose to extract information from records e.g. GP notes or hospital notes, how will this be done, by hand or by searching databases or both? Again bias may arise if the extractor is not appropriately blinded (see [C-2](#)). The applicants also need to ask themselves whether the mode of collection chosen will yield sufficiently complete and accurate information. For example, by searching GP databases you are unlikely to pick up complete information on casualty attendance. A single blow into a spirometer does not produce a very reliable (see [A-4.4b](#)) assessment of lung function and most studies use the maximum from 3 consecutive blows.

[Back to top.](#)

#### **A-4.4 Validity and reliability.**

Where possible information should be provided on the validity and reliability of proposed methods of measurement. This is particularly important if a method is relatively new or is not in common usage outside the applicant's particular discipline. For further information see [A-4.4a](#), [A-4.4b](#), [A-4.4c](#), [Altman \(1991\)](#) and [Bland 2000](#).

[Back to top.](#)

#### **A-4.4a Validity**

By validity we mean does the method actually measure what the applicant assumes? For example does a questionnaire designed to measure self-efficacy (i.e. belief in ones ability to cope) on some ordinal scale actually measure self-efficacy? Even if validity has been demonstrated previously, has it been demonstrated in an appropriate setting? For example a method which has been validated for use among adults may not be valid if used among children and validity does not always cross countries. Sometimes applicants reference a questionnaire score that has been previously validated but they plan to use a modified version. Of interest to the reviewer is whether these modifications have affected validity. This may well be the case if the number of questions used to produce the score have been reduced for ease of administration. For further information see [Bland & Altman \(2002\)](#).

[Back to top.](#)

#### **A-4.4b Repeatability (test-retest reliability)**

By repeatability we mean how accurately does a single measurement on a subject estimate the average (or underlying) value for that subject? The repeatability of a measurement therefore depends on the within-subject standard deviation, which can be calculated using a sample of closely repeated (in time) measurements on the same subjects. The repeatability coefficient is simply the within-subject standard deviation multiplied by 2.83 and is an estimate of the maximum difference likely to occur between two successive measurements on the same subject (see [Bland & Altman 1986](#), [1996](#), [1996a](#) and [1996b](#)).

[Back to top.](#)

#### **A-4.4c Inter-rater reliability (inter-rater agreement)**

For methods of measurement in which the role of an observer is key, inter-rater reliability should also be considered ([Altman 1991](#) p403-409). In other words what sort of differences in measurement are likely where you have the same subject measured by different observers/raters? How closely do they agree? Substantial and unacceptable biases can arise in a study if the same observer is not used throughout. Sometimes,

however, the use of more than one observer is the only practical option as for example in multi-centre clinical trials. In such cases it is important to:

- \* try and improve agreement between observers (e.g. by training)
- \* use the same observer when making 'before' and 'after' measurements on the same subject.
- \* in a clinical trial, to balance groups with respect to assessor and to make all assessments blind (see [Pocock, 1983](#) p45-48).
- \* in an observational study, to note down the observer used for each subject so that observer can be adjusted for as a potential confounder in the analysis.

## References for this chapter

Altman DG. (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London.

Bland JM and Altman DG. (1986) [Statistical methods for assessing agreement between two methods of clinical measurement](#). *Lancet*; i 307-10.

Bland JM, Altman DG. (1996) [Measurement error](#). *British Medical Journal* **313** 744.

Bland JM, Altman DG. (1996a) [Measurement error and correlation coefficients](#). *British Medical Journal* **313** 41-2.

Bland JM, Altman DG. (1996b) [Measurement error proportional to the mean](#). *British Medical Journal* **313** 106.

Bland JM, Altman DG. (2002) Validating scales and indexes *British Medical Journal* **324** 606-607.

Bland M. (2000) [An Introduction to Medical Statistics, 3rd. ed.](#) Oxford University Press, Oxford.

Breslow NE and Day NE. (1980) *Statistical Methods in Cancer Research: Volume 1 - The analysis of case-control studies*. IARC Scientific Publications No. 32, Lyon.

Breslow NE and Day NE. (1987) *Statistical Methods in Cancer Research: Volume 11 - The design and analysis of cohort studies*. IARC Scientific Publications No. 82, Lyon.

Burr ML. (1993) Epidemiology of Asthma: in Burr ML (ed): *Epidemiology of Clinical Allergy*. Monogr Allergy Vol 31, Basel, Karger. p 80-102.

Henshaw DL, Eatough JP, Richardson RB. (1990) Radon as a causative factor in induction of myeloid leukaemia and other cancers. *Lancet* **335**, 1008-12.

Lilienfeld AM, Lilienfeld DE. (1980) *Foundations of Epidemiology, 2nd ed.*. Oxford University Press, Oxford.

Pocock SJ. (1983) *Clinical Trials: A Practical Approach*. John Wiley and Sons, Chichester.

Rothman KJ, Michels KB, Baum M. (2000) [For and against. Declaration of Helsinki should be strengthened.](#) *British Medical Journal* **321** 442-445.

[Back to top](#)

---

[Back to Brief Table of Contents](#)

[Public Health Sciences Home Page](#)

[St. George's Hospital Medical School Home Page](#)

This page is maintained by [Martin Bland](#)

Last updated: [September 11, 2002]

**Copyright St. George's Hospital Medical School**

[Back to top](#)