

GUIDELINES FOR COMPLETING A RESEARCH PROTOCOL FOR OBSERVATIONAL STUDIES

The aim of this guide is to help researchers write a research study protocol for an [observational](#) study. The guide will take you through each section of the protocol giving advice and examples of the information required in that section. This is a guide only and for those requiring more information on particular topics; some useful references are given at the end of this document. We also recommend the St. George's on line handbook (now hosted at York University) at <http://www-users.york.ac.uk/~mb55/guide/describ.htm> and there are some links to that document in this guide. Many of the methodological aspects of designing a research study and writing the protocol can benefit from the advice of a statistician. Such advice should be sought at an early stage and is available for UCL/UCLH/RFH researchers through the [Biostatistics group](#) at the Joint UCLH/UCL/RFH Biomedical Research Unit.

1. Title Page

1.1 Title

It is useful to specify both a full title and short title

- The full title should include summary study design, exposure and outcome of interest, patient population and setting.
- The short title is a summary of this
- The titles specified must be consistent across all documents relevant to the study.

1.2 Names (titles), roles and contact details of:

- Authors, investigators, experts and advisors involved in the study
- Sponsor & monitor – as agreed with Chief / principal investigator's employer and the host Trust
- Study site(s), clinical laboratory(s), technical departments and institutions involved in the study

1.3 Protocol details

- Version number
- final / draft
- Date

2. Signature Page

Signatures of all healthcare professionals involved in the study

3. Contents Page

4. List of Abbreviations

All abbreviations used should be listed and defined

5. Summary

In this section the problem being addressed should be introduced to "set the scene".

- Aim(s) and reason for the study
- A brief description of the study groups, design as well as the exposure / outcomes of interest.
- Primary and secondary objectives

- Brief description of methods

6. Background

The background should include a **clear explanation of the main research question** along with a full literature review, a detailed justification for the study and discussion of its feasibility.

6.1 *Literature Search and Review*

Precise details of the literature search completed should be given here.

- The nature of any unpublished work should also be documented here.
- Help with this aspect your intended research can often be sought from a clinical librarian. A good literature search at the beginning of the research process is invaluable and will steer the researcher in the most appropriate direction.
- Your review should make reference to relevant papers, unpublished works as well as clinical experience.

6.2 *Justification*

- A statement indicating the size of the problem (and effect on the health service) and why the study is appropriate
- Explain what the potential benefits are – to patients and the health service
- Explain what your study will add to the body of evidence already available.
- Discuss the feasibility of the study in terms of subject and data availability as well as length.

7. Specific aim(s) of the study

- Describe the primary research question
- Clearly defined objectives in terms of measurable endpoints
- Distinguish primary and secondary objectives

8. Study Design

The two most commonly used designs for observational studies are (A) case-control studies (including nested case-control studies) and (B) cohort studies. In the former, the study groups are chosen on the basis of their disease or outcome of interest. In a cohort study the comparison groups are identified according to an exposure of interest. A description of the design should be given, along with details of any matching and blinding used. Cross-sectional studies are also discussed in (C).

(A) Case-control studies

Definition of study groups:

- Describe an objective definition for the disease or outcome of interest and use this to define the **cases**.
- Describe the appropriate control group for the particular group of cases. In general the controls should be sources from a similar population as the cases, the only difference being that they do not have the outcome of interest.

Matching (if relevant):

- Describe the matching factor(s) and how the matching will be done. The main purpose of matching is to control for confounding. By matching on age, for example, we are attempting to ensure the cases and controls have the same (or similar) age distribution.

Blinding and avoiding bias:

- List the measurements that are to be blinded and how this will be done. Blinding is a technique used to minimise bias. In observational studies, the minimisation of ascertainment bias can be helped by blinding. e.g. If subjects in a case control study are to be interviewed to assess exposure status, then the interviewer should, if possible, be blind to the outcome of the subjects.
- Methods used to minimise other types of bias should be described. E.g. using multiple sources of information can help minimise verification bias.

(B) Cohort Studies

Definition of study groups:

- The exposure of interest should be clearly described and used to define the “exposed” group. Groups like doctors, civil servants, surgery patients are often chosen as the source of the groups because they are easy to define or monitor. In these situations an internal comparison group without the exposure of interest would be used. E.g. doctors who smoke and doctors who do not smoke. Sometimes groups with special exposures are chosen (e.g. brewery workers, workers in nuclear power stations) and in these cases an external comparison group would need to be found.

Blinding and avoiding bias

- List the measurements that are to be blinded and how this will be done. Blinding is a technique used to minimise bias. E.g. in a prospective cohort study where subjects are being followed up in the future, the researcher who assess any outcome measures should, if possible, be blinded to exposure status.

(C) Cross-sectional studies

In cross-sectional studies the exposure status and disease status are assessed at a single point in time (e.g. a sample survey; 2001 census). It is therefore **not possible** to establish whether the exposure preceded or resulted from the disease. However, data from these studies are useful in providing information about the health status and needs of a population. They are sometimes called prevalence studies. Selecting the study groups has the same problems/ features as with cohort studies. Response rates are also often a problem and the strategy used to recruit a representative sample and for maximising the response rate should be given.

(D) Other types of studies

Other types of observational studies may involve only one group of subjects. For example, in a [method comparison study](#), measurements will be taken on a group of subjects using two (or more) methods. The guidelines set out in this document still apply, the main difference being that a second group of subjects is not used.

9. Study group(s)

Include details of:

- Source of subjects (where they come from and why this group is appropriate). For example, are they a random sample from a larger population or are they obtained from a disease register? They may be patients attending a clinic over a particular period.
- Number of centres involved or regional / national limits on possible recruitment
- Subject inclusion and exclusion criteria (with justification if necessary)
- Expected no of eligible participants available per year and proportion of these expected to agree to take part.

10. Recruitment or choosing participants

Details of recruitment process including

- method of recruitment (e.g. via adverts, clinics, GP referral, entry in established database (e.g. Society of Cardiothoracic Surgeons National Database)
- payment of participants
- details of procedures, tests, screenings carried out to assess study suitability
- Provision of patient information sheet (include as appendix)
- gaining patient consent (how consent will be obtained, who will gain consent, whether a witness will be present, how long the subject will have to decide, the arrangements for non English speakers and special groups (e.g. mentally ill, children, those suffering from dementia.)
- detail of enrolment procedure

11. Data

11.1 Data to be collected

- if using an established database permission to use the data must be sought and a detailed description of how (and by whom) the data will be collected for the study should be given. Also, an example of the data extraction form used to obtain the data for your study should be included.
- provide a detailed list of all data (outcome variables, explanatory variables, potential confounding variables etc) to be collected, with each description including :
 - source of the data (e.g. patient questionnaires, patient notes, electronic data, procedure)
 - time point for collection (baseline, during treatment, at follow up point)
 - who will collect the data
 - why the data is being collected (e.g. baseline comparison data, primary outcome, important prognostic / explanatory /confounding variable)
 - whether the data is from a standardised tool (e.g. McGill pain score) involves a procedure (in which case full details should be supplied). If a non-standard tool is to be used, detail on reliability and validity should be given.
 - what form the data will take (e.g. binary, continuous (numeric), time to event)

- useful to include table / diagram describing schedule for data collection.
- describe methods used to maximise completeness of data (e.g. telephoning patients who have not returned postal questionnaires)
- include data collection forms and questionnaires as appendices

11.2 Data handling and record keeping

- describe procedures for data collection and recording (software to be used, location of the data etc)
- detail methods implemented to ensure validity and quality of data (e.g. double entry, cross validation etc)
- Security / storage of data
- Records retention – duration and location
- Adherence to Data Protection Act 1998 and Caldicott

12. Statistical Considerations

12.1 Sample size calculation

Details of the precision or power calculation used to estimate the required sample size based on primary outcome

- Assumptions made (statistical assumptions regarding distribution)
- Estimates of difference to be detected along with appropriate justification
- Chosen levels of significance and power.
- Reference or details of method/formula used for the calculation

12.2 Analysis

- Describe which variables will be used to assess groups comparability and how they will be reported (e.g. means, proportions)
- Description of primary and secondary analyses including summary measures used, methods of analysis (e.g. t-test , logistic regression) and how the results will be reported (e.g. odds ratios with 95% confidence intervals)
- Details of adjustments for pre – defined confounders.
- Approach used to deal with missing data and loss to follow-up.
- When will the analysis be done and by whom.
 - Details of how any planned subgroup analyses will be done

13. Compliance

13.1 *Subject compliance*

- procedures for monitoring (e.g. exercise diary for subjects on a rehabilitation programme)
- recording of patient compliance information (what will be recorded and where)
- detail of follow-up of non compliant subjects

13.2 *Withdrawal of subjects*

- describe under what circumstances and how subjects will be withdrawn from the study
- give details of documentation to be completed on subject withdrawal (including recording reasons for withdrawal and any follow-up information collected)

14. Ethical Considerations

Description of ethical issues related to the study. For example consider:

- Approvals from relevant groups (e.g. MREC, LREC, MHRA, Trust(s))
- Informed consent (subject information and informed consent form appended)
- Allowances for special groups (e.g. non English speakers, children, mentally ill)
- Patient withdrawal / discontinuation

15. Finance and Insurance

- Finance and insurance details (if not addressed in separate agreement)
- Cover for non negligent and negligent harm

16. Reporting and Dissemination

Details of how, where and when the results of the study will be reported / presented.

Useful reading

Websites

- Declaration of Helsinki (<http://www.wma.net/en/30publications/10policies/b3/index.html>)
Provides ethical principles for medical research involving human subjects
- COREC guidelines (www.corec.org.uk)
Includes patient information sheet and consent form guidelines
- Martin Bland et al, Statistical Guide for Research Grant Applications,
<http://www-users.york.ac.uk/~mb55/guide/guide.htm>
Includes detailed information and definitions of many aspects required for a research protocol as well as information about randomisation software and services
- ICH Harmonised Tripartite Guidelines for Good Clinical Practice
(<http://www.cgmh.org.tw/intr/intr1/c0040/web/C/ICH%20GCP%20E6.pdf>)

Books

- Altman, DG. (1991) *Practical Statistics for Medical Research*. London: Chapman and Hall, Chapter 15.
- Bland M, (1995) *An Introduction to Medical Statistics*, Oxford University Press
- Betty R Kirkwood and Jonathan AC Sterne (Blackwell 0-86542-871-9)
- Ed Douglas Altman, David Machin, Trevor N Bryant, Martin J Gardner (BMJ Books ISBN 0 7279 1375 1) *Statistics with Confidence*
- Rothman K.J, Greenland S, *Modern Epidemiology* (2nd Edition). Lippincot-Raven
- Hennekens C.H., Buring J.E.(1987) *Epidemiology in Medicine*. Lippincot-Raven (excellent chapters on design and bias)
- Machin D, Campbell MJ, Fayers PM and Pinol APY (1997) *Sample Size Tables for Clinical Studies*. Second Edition. Oxford: Blackwell Science.

Papers

BMJ statistics notes provide some brief but useful information on various topics including:

- J Martin Bland & Douglas G Altman, Matching BMJ 1994;309:1128 (29 October)
- J Martin Bland & Douglas G Altman Multiple significance tests: the Bonferroni method BMJ 1995; 310:170
- J Martin Bland & Douglas G Altman *Presentation of numerical data* BMJ 1996; 312:572 (2 March)
- J Martin Bland & Douglas G Altman Survival Probabilities (the Kaplan-Meier method) BMJ 1998; 317:1572-1580
- J Martin Bland & Douglas G Altman The logrank test BMJ 2007; 328:1073
- J Martin Bland & Douglas G Altman Analysis of continuous data from small sample BMJ 2009; 338:a3166
- DG Altman, JM Bland BMJ One and two sided tests of significance 1995; 311:485
- D G Altman, JM Bland Comparing several groups using analysis of variance BMJ 1996; 312:1472-1473
- DG Altman, JM Bland Bland Units of analysis BMJ 1997; 314: 1874
- DG Altman, JM Bland Time to event(survival) data BMJ 1998; 317:468-469
- DG Altman, JM Bland Standard deviation and standard errors BMJ 2005; 331:903
- Day JD, Altman DG. *Blinding in clinical trials and other studies* BMJ 2000; 321: 504
- David A Grimes, Kenneth F Schulz Descriptive studies what they can and cannot do Lancet 2002; 359:145-49
- David A Grimes, Kenneth F Schluz Bias and causal association in observational research Lancet 2002; 359:248-52
- David A Grimes, Kenneth F Schulz Case-control studies: research in reverse Lancet 2002; 359:431:34
- David A Grimes, Kenneth F Schulz Compare to what? Finding controls for case-control Lancet 2005; 365:1429-33

Other useful articles

Statistics Notes in the [*British Medical Journal*](#)

This series is edited by Doug Altman, Cancer Research UK, and Martin Bland, University of York. All are available on the [BMJ](#) website.

For an extended list of publications on statistics and research methods use the BMJ link to topic collections www.bmj.com/cgi/collection and go to non-clinical - statistics and research methods.