

How to Design a (Good) Epidemiological Observational Study: Epidemiological Research Protocol at a Glance



Como Desenhar um (Bom) Estudo Epidemiológico Observacional: Um Olhar sobre o Protocolo de Investigação

Inês FRONTEIRA¹

Acta Med Port 2013 Nov-Dec;26(6):731-736

ABSTRACT

In this article, we propose a general structure for designing a research protocol of an observational epidemiological study. We start by highlighting the importance of the research protocol, namely in accounting for some bias and guaranteeing methodologic rigor and study reproductibility. Next, we reflect on some of the essential elements of a research protocol no matter its objective. We further present some specific issues to be included according to the type of study: cross-sectional, case-control and cohort.

Keywords: Biomedical Research; Epidemiologic Methods; Epidemiology; Epidemiologic Studies; Research Design.

RESUMO

Neste artigo, apresentamos uma proposta de estrutura geral para o desenho de um protocolo de um estudo epidemiológico observacional. Começamos por realçar a importância do protocolo de investigação, nomeadamente no controlo de alguns vieses e na garantia do rigor metodológico e reproductibilidade do estudo. De seguida, reflectimos sobre os elementos essenciais que devem constar num protocolo de investigação, independentemente do objectivo do mesmo. Apresentamos, ainda, alguns aspectos específicos a constar em protocolos de estudos transversais, de caso-controlo e de coorte.

Palavras-chave: Epidemiologia; Estudos Epidemiológicos; Métodos Epidemiológicos Investigação Biomédica; Estudo Observacional; Desenho do Estudo.

INTRODUCTION

Designing an epidemiological study, and the research protocol that should go with it, is a very challenging task. The majority of researchers usually start in overexcitement but, despite enthusiasm and interest being essential for success, designing and planning how to conduct the study is also fundamental. People often want to go directly from the idea that they have worked out for their research to collect data, sometimes without having read any material on the subject. But, if you want to have a study done, you have to sit for a while and start thinking about what you want, can and will do to have a good epidemiological study. The difficulty is that, most of the times, excitement does not survive planning. Believe me that, when planning an epidemiological study, Murphy's Law does apply and if anything can go wrong it will for certain.

After having an idea about the problem you want to study you should always spend a little bit of time searching for literature, reading what has been done in that area, what the current knowledge is and what have been the methodological approaches to the subject. After you have been through this process, sometimes, after you talked to a few colleagues, you will have a clearer idea about the objectives and hypothesis of your research in the back of your mind and you will be more informed to start designing a (good) epidemiological study.

There is no perfect epidemiological study and those who seek for it might end in a "dead end", meaning not doing any study at all. However we can always design a good epidemiological study which will be the one that tries to control

as much error and bias as possible, allowing for the most precise and accurate estimates, is adapted to context, and has the right balance between internal and external validity.

The secret to a good epidemiological study relies basically on planning, and planning and planning again. And planning refers to thinking ahead, to ensure that the study is probability capable of answering to the research question and predicting all things that can happen to bias the study – we must consider all issues from conceptual framework in which the study is based on, objectives, data collection and analysis and how those elements contribute to achieve the objectives of the study. The place to start doing this is the research protocol.

Epidemiological research protocol

A milestone of any epidemiological study is the writing of the research protocol.¹ It is a guide for researchers and collaborators to follow and consult during the implementation of the study. On the other hand, when we write the research protocol, we are guaranteeing rigor of the research, i.e., the implementation of the study will not be affected by the results of the study itself or by any other issue that might arise.

The format of the research protocol varies widely according not only with the area of knowledge (e.g., epidemiology vs sociology), type of study (randomized controlled trial versus cross-sectional study), or the goal of writing it (e.g., apply for funding or academic purposes). Nevertheless, there are several common features to all research protocols.

1. Unidade de Saúde Pública Internacional e Bioestatística. Instituto de Higiene e Medicina Tropical. Universidade Nova de Lisboa. Lisboa. Portugal.

Recebido: 08 de Julho de 2013 - Aceite: 08 de Julho de 2013 | Copyright © Ordem dos Médicos 2013

Basically, a research protocol is divided in 6 big sections: title, introduction, materials and methods, expected results, references and annexes. Although all chapters are important and complementary, the methods and materials section is by far the one that should worry us most and where we should invest more. In this section we will try to address error and avoid bias, we will compromise between internal and external validity and we will present a guide that ultimately will allow the study to be replicated.

What are we writing the research protocol for?

The format and structure of the research protocol might vary according to its objective. Generally, research protocols are written to (1) apply for funding; (2) present in academic context for evaluation/ degree or (3) guide the researcher throughout the research without any other objective.² An example of this last case is when we propose ourselves to conduct an epidemiological study for producing a paper and we are not applying for funding or doing an academic exam. Sometimes, when this is the case, we often forget to write the protocol postponing it to when we will be preparing the methods section of a paper for submission to an international journal. This will inevitably lead to a series of constraints: research might be biased and deviated from initial objective, we might be 'fishing for data', among others. So, whatever the case, a research protocol is always needed and must be wrote before the implementation of the study.

If a research protocol is going to be developed to apply for funding we should look carefully at the demands made by the funding agency and write the research protocol according to them. Most of the times, the funding agencies ask for information that has to be organized in a different way that it would be in a more traditional, academic like research protocol. An example: in the 2012 call application form of Fundação para a Ciência e Tecnologia (FCT), included, in the project technical description: literature review, plan and methods, tasks (beginning/ end, duration, description and expected results, members involved and time allocation – in person* month), project timeline (management structure, milestones, timeline – gantt chart), bibliographic references, budget and budget rationale and possible conflicts of interest.

In this paper we will present the traditional, more 'academic' format of a research protocol for an observational epidemiological study since many of the information presented here does not change in content (only in format) to that required by funding agencies. Please note that it is a general structure and that, depending on the type of epidemiological study, items can be present, absent or more or less developed. Due to the specificity of experimental research protocols these are not addressed in this paper.

Language and writing style

The language and writing style of a research protocol should be plain, direct to the point and in future tense. We must use short sentences and change paragraph whenever there is a variation in subject. Since the research protocol

might be assessed or simply read by individuals that might not have the same background as we have, we should write it in a doubtless way, always defining concepts even if we are very familiar with them. It might be helpful to think of the study protocol as being a recipe that either us or anyone else can follow and arrive at the same result. Thus, we should write for others to read and understand.

Title

The title of the study must be comprehensive and clear. Ideally it should state the epidemiological triad – persons, time and place – and include the type of study. This helps readers (either teachers or evaluators from funding agencies) to contextualize the study and it tells straight forward what type of evidence we might be looking at as well as the major limitations (i.e., in terms of study design). If we think about upcoming outcomes of the study, as a published paper, a good, self-explanatory title might help prospective readers to find exactly what they want and enhance the probability of the study being cited. An example of a clear title: a cross sectional survey on job satisfaction and intention to leave in medical doctors working in the ICU of an University Hospital in Oporto, Portugal.

INTRODUCTION

The introduction of a research protocol should comprise (1) literature review and (2) objectives.

Literature review

In an epidemiological research protocol, the objective of the literature review is to set the scenario, to generally describe what is known about the subject, i.e., we have to answer the what, who, when, how, and why. The literature review will 'relate a study to the larger, ongoing dialogue in the literature, filling the gaps and extending prior studies'.³ In this case it is presented at the beginning to provide direction for the research questions or hypothesis.³

In order to conduct a literature review we should: (1) define key words and databases to search; (2) search databases using keywords; (3) select articles and books of interest trying to understand if the article/ book will contribute to the understanding of the literature; (4) go through the articles/books, summarize and place them in groups according to the subject feature they address; (5) structure the literature thematically or by important concepts, summarizing, in the end, important themes and (6) explain how our study will add further to the literature, what is its innovation and expected impact.

There are several steps to write a literature review. Usually, and in a quantitative approach based on deduction we prefer going from general to specific (e.g., consider a study on the determinants of cross border reproductive care – we should analyze cross border health care in general and its determinants and then the specific case of reproductive care). While writing the literature review, and despite being more or less developed in the research protocol, we should also bear in mind that it will set the basis for the discussion.

OBJECTIVES

After we have written the literature review, we must state the objectives and/or underlying hypothesis of our research. These relate to the question that we want to see answered by our study.⁴ Defining objectives will help focus the research avoiding extra-work and deviations from initial intention, thus avoiding bias.

We should state the overall objective of the research, also called in English literature aim (e.g., to understand the determinants of condom use in adolescents) and then present the specific objectives (which are at the level of operational tasks and can often be translated in to a statistical hypothesis).¹ The specific objectives must be in line with the overall objective, contributing to its achievement, should be specific, attainable and measurable.¹ An example: to test the association between number of cigarettes smoked per day and having asthma – the underlying statistical hypothesis is H₀: there are no differences in the number of cigarettes smoked per day among those with and without asthma. In this example, by defining the objective, we are also clarifying the type of analysis and results we are looking for and, at the same time, indirectly defining the type of data that we are collecting. Thus, writing a research protocol is not a linear process – all is intertwined.

MATERIAL AND METHODS

After defining the research objectives/ hypothesis we must, by now, have a clearer idea on the methods we are going to use in order to achieve them. Part is already determined by the objectives themselves (e.g., if our objective is to determine the prevalence than we should consider running a cross sectional study).

Usually, the methods pertain to a specific approach: quantitative, qualitative and mixed methods.⁵⁻⁷ This relates to the way the researcher understands reality. In epidemiology we test objective theories by examining the relationship among variables - qualitative (or positivist approach). Creswell explains that ‘...those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings’.³ Despite that, in the last years, epidemiologists have started to use mixed methods approach (complementary combination of quantitative and qualitative approach).

The recognition of the type of approach is very important because it will set the tone for the methods section and, at the same time, help us to choose the appropriate methodological references to be use while maintaining a theoretical coherence throughout the research protocol. It is not good practice to use, for instance, ‘qualitative language’ or references while working in a positivist paradigm.

Study design

We should define the type of study that we are conducting, preferably using a taxonomy that is in line with the paradigm⁸. By doing this we are informing not only about

the study design but indicating potential bias, as well the expected type of results or their value in terms of strength of evidence.

Setting

The description of the study setting depends on the type of study and on the importance it can have on the results. It can also shed light on some of the features of the problem that we are studying and later help frame findings. This section can include information on the setting (e.g., community, institutional) and locations where data is going to be collected, relevant events/ dates, periods of recruitment, exposure, follow-up, and data collection.⁹

Population and sample

This is one of the most critical parts of the methods section. Usually we have to make a series of decisions concerning the selection of subjects to be studied and this conditions the external validity of the study. We should be extremely cautious and describe all the steps and decisions made in order for others to repeat flawless our study. Thus, we should (1) explain why we chose to conduct a census or use a sample; (2) define the studied population (even if we chose to use a sample); (3) mention eligibility criteria; (4) explain the rationale behind the chosen type of sample (probabilistic/ not probabilistic) and within each type of design chosen (e.g., simple random sample, quota, etc.); (5) describe how the minimum sample size was achieved (present mathematical forms and values used for computation, e.g., mean, proportion, confidence level, precision or stopping rules); (6) explain how sample units were selected (if necessary we should present a step by step guide).¹⁰

Depending on the type of study, further information should be added (Table 1).

Data collection instruments

All data collection instruments should be mentioned and explained. Three situations might occur: (1) the data collection instrument is already developed and validated; (2) the data collection instrument is already developed but not validated for the target population of the study; or (3) the data collection instrument is going to be developed for the purpose of the study.

In the first case, we should briefly explain the instrument (e.g., objective, how it was developed, coding and scoring) and the validation procedure using the original references for both cases.

In the second case, we might have to translate the instrument and/or adapt it culturally and then validate it. If this is the case, we should describe how we are going to do this and later, in the results section, present the results for the validation of the instrument.

When we are going to develop the instrument, we should mention the type of instrument (e.g., form, questionnaire, checklist), the type of questions (e.g., open, close), the layout and the rational for it (e.g., organized by sections, which sections come first/ last), the volume, complexity and nature

Table 1 – Information that should be presented according to type of study and item of the research protocol (partially adapted from STROBE statement)⁹

Item	Type of study		
	Cross-sectional	Case-control	Cohort
Population and sample	Eligibility criteria Sources and methods of selection of participants	Eligibility criteria Definition of case and of control Sources and methods of case ascertainment and control selection Rationale for choice of cases and controls Ratio of controls per case Matched studies – matching criteria and number of controls per case	Eligibility criteria Sources and methods of selection of participants Methods of follow-up Matched studies – matching criteria, number of exposed and unexposed
Data analysis	Description of all statistical methods, including those to control for confounders Methods to examine subgroups and interaction (effect modification) Handling of missing data Mention how sampling strategy was handled in analytic methods (e.g., design effect, weights) Sensitivity analysis	Description of all statistical methods, including those to control for confounders Methods to examine subgroups and interaction (effect modification) Handling of missing data Handling of matching of cases and controls Sensitivity analysis	Description of all statistical methods, including those to control for confounders Methods to examine subgroups and interaction (effect modification) Handling of missing data Handling of loss to follow-up Sensitivity analysis

of the data collection (e.g., face-to-face interview, registries, etc.). It is also good practice to briefly analyze the implications of each of the data collection instrument and method chosen on the quality of data (e.g., expected response rate, response bias).¹¹

No matter the situation, we should always explain how, by whom and when are data collection instruments going to be applied.

Variables

In this chapter we must clearly identify the dependent and independent variables as well as potential confounders and effect modifiers. In each case we should present a table with the variable name (e.g., gender), its definition (e.g., refers to the gender of the respondent), type (quantitative/qualitative) (e.g., qualitative), measurement scale (nominal, ordinal or numeric) (e.g., nominal) and domain (e.g., male and female). If useful, we can add information on computer notation for the variable in the database.

When variables have to be calculated (e.g., scores of SF-36) or recoded (e.g., age into age groups), this is the place to explain their computation, rationale for coding and interpretation.⁹

Data analysis

In data analysis we should mention how data is going to be introduced in the database, what program(s) for data handling and analysis we are going to use and thoroughly describe all data analysis to be performed. Ideally we should prepare a data analysis plan, referring which measures are going to be computed for each variable, which variables are going to be crossed with, the statistics to be computed and the tests to be applied. As an alternative, we can generally mention the analysis to run according to the type of variable (e.g., compute central tendency measures for numerical scale variables), and the type of analysis (e.g., for studying the relation between two numerical scale variables we are going to use Pearson correlation coefficient). If we expect to use multivariate analysis we should mention the type and the criteria for choosing the variables to enter the model. Other statistical analysis should also be mentioned (e.g., cluster analysis, factorial analysis).

Additionally we should set the confidence level for statistical inference as well how missing data are going to be handled. We should not forget that data analysis should reflect the study objectives – the results from the previous are going to enable us to respond to the later ones.

Statistical analysis might change according to the type of study (Table 2).

Implementation

In the implementation section we should thoroughly describe how the study is going to be implemented. This should include all steps from preparing the protocol, procurement arrangements, submissions to ethical committee to getting approvals (e.g., hospital management board), among others.

Table 2 – Ethical principles involved in epidemiologic research, general recommendations on how to address them in the research protocol and illustrative example^{12,14,15}

Ethical principle	How to address in the research protocol	Example (study to screen for HIV using rapid tests)
Beneficence	Clarify the direct and indirect benefits of the research and balance benefits against risks	Early detection of HIV infection enhances treatment effectiveness, allows for a better follow-up and improves quality of living in infected patients. Knowing the prevalence of HIV infection will help planning for appropriate health care
Non-maleficence	Explain how harm is going to be avoided or minimize	In
Justice	Explain how benefits and risks are going to be distributed among study participants	All subjects are going to be screened for HIV/AIDS. Diagnosis will be returned to all subjects except if they refuse to have it (by writing). All infected subjects will be provided with counseling and referenced for confirmation of diagnosis and appropriate treatment/ follow-up.
Autonomy	Explain how respect for autonomy of individuals will be guaranteed and how the rules of privacy and confidentiality are going to be addressed	We will explain to all individuals that they are free to participate in the study and no harm will come from refusing participation. The risks and discomfort of having the test done as well as the consequences of the results will be orally explained and written in the informed consent. Subjects will be informed that in case they test positive they will have to undergo further tests and that best available treatment will be provided. Subjects will be asked to sign informed consent. The tests and inquiry will be done in an room where only the researcher and the subject will be present. The test will be identified by code. Only the researcher will be able to link the code to the subject using a key.

All requests to external bodies (e.g., ethical committees) should be mentioned and the letters can be presented in annexes to the protocol.

Usually we include a work plan, often in a Gantt chart format, where we present activities and the time line for their completion, as well as the relationship between the several activities (e.g., independent, dependent, happening at the same time, etc.). In some cases it might be useful to include milestones and deliverables (required by some funding agencies like FCT and European Union). Milestones identify critical points in the implementation of the study that have to be achieved in order for the study to proceed. Deliverables are products resulting from the study (e.g., questionnaire or article for publication).

Even if we are not applying for funding, we should present an estimate of resources (e.g. financial, material, human) needed to conduct the study. When we present a budget this should be realistic and account for the time lapse between planning the study and procurement/acquisitions. In the budget we should mention items (often organized by type – e.g., human resources, travel, and procurement), unit cost and total cost. Sometimes, when we came into this stage, we realize that we not have the sufficient resources (time or money) to conduct the study and we might have to review the entire protocol.

This section can also include a plan and methods for dissemination of results (e.g., publications and conferences).

Ethical considerations

Epidemiological studies cannot be detached from social context.¹² We are often faced with the need to obtain accurate information but with the moral imperatives of avoiding harm, do good, promoting justice, and respect autonomy.

In the ethical considerations chapter of the research protocol we should explain how we are going to deal with ethical constrains to the design and implementation of the study. Usually we should do this by analyzing our study in light of the ethical principles of beneficence, non-maleficence, justice and respect for autonomy.¹³ It is not sufficient to state that we are going to respect these principles but how we are going to achieve that (Table 2). Nevertheless, the abstract principles before mentioned provide a starting point and specification depending on the context is always needed (e.g., selection of subjects, choice of comparison group, interviewing procedures, avoidance of bias).¹³

Bear in mind that this is the section that ethical committee will be analyzing more carefully.

CONCLUSION

Although occasionally being time consuming and tedious, writing a good research protocol is a prerequisite for a good epidemiological study, independently of its objective (academic or funding).

All parts of the research protocol should be carefully written using simple and clear language, defining and explaining essential concepts in order to allow others besides

the research team to replicate the study. All decisions must be presented with their rationale and it is good practice to reference methodological options and terms.

By carefully preparing a research protocol we avoid further bias to the study, guarantee rigor and (in)directly enhance study's internal validity. Besides that, a good research protocol is also a good contingency plan for all the things that can impair the implementation of the study. Sometimes it can work as a plan to deal with risk (of, for instance, not ending the study). More, a good research pro-

ocol is half-way work done for a good paper or scientific report.

CONFLICT OF INTEREST

The author has no competing interests to declare.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not for-profit sectors.

REFERENCES

1. Bowling A. Research methods in health: investigating health and health services. Philadelphia: Open University Press; 2000.
2. dos Santos Silva. Cancer epidemiology: principles and methods. Lyon: International Agency for Research on Cancer; 1999.
3. Creswell J. Research design: qualitative, quantitative and mixed methods approaches. 3rd ed. London: Sage; 2009.
4. Abramson J. Survey methods in community medicine: epidemiological studies, programme evaluation, clinical trials. 4th ed. London: Churchill Livingstone; 1990.
5. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. *J Mixed Methods Res.* 2007;1:112-33.
6. Sale J, Lohfeld L, Brazil K. Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality Quantity.* 2002;36:43-53.
7. Porta M, Greenland S, Last J. A dictionary of epidemiology. 5th ed. Oxford: Oxford University Press; 2008.
8. Fronteira I. Estudos observacionais na era da medicina baseada na evidência: breve revisão sobre a sua relevância, taxonomia e desenhos. *Acta Med Port.* 2013;26:161-70.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandembroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4:e296.
10. Daniel J. Sampling essentials: practical guidelines for making sampling choices. London: Sage; 2012.
11. Machin D, Campbell M. Design of studies for medical research. John Wiley & Sons, Ltd; 2005.
12. Coughlin S. Ethically optimized study designs in epidemiology. In: Coughlin S, Beauchamp TL, editors. *Ethics and epidemiology.* Oxford: Oxford University Press. 1996;145-55.
13. Coughlin S, Beauchamp TL. Ethics, scientific validity and the design of epidemiologic studies. *Epidemiology.* 1992;3:343-7.
14. Beauchamp TL, Cook RR, Fayerweather WE, Raabe GK, Thar WE, Cowles SR, et al. Ethical guidelines for epidemiologists. *J Clin Epidemiol.* 1991;44:S151-69.
15. Dickens BM. Issues in preparing ethical guidelines for epidemiological studies. *Law Med Health Care.* 1991;19:175-83.

Inês FRONTEIRA

How to Design a (Good) Epidemiological Observational Study: Epidemiological Research Protocol at a Glance

Acta Med Port 2013;26:731-736

Publicado pela **Acta Médica Portuguesa**, a Revista Científica da **Ordem dos Médicos**

Av. Almirante Gago Coutinho, 151

1749-084 Lisboa, Portugal.

Tel: +351 218 428 215

E-mail: submissao@actamedicaportuguesa.com

www.actamedicaportuguesa.com

ISSN:0870-399X | e-ISSN: 1646-0758



ACTA MÉDICA
PORTUGUESA

