Planning Undergraduate Research:

A Step-by-Step Guide for Medical Students

R. Kumar, P.A.D. Coonghe, R. Surenthirakumaran

Department of Community and Family Medicine
Faculty of Medicine, University of Jaffna
2020

Table of Contents

1. Introduction to the undergraduate research project 4
2. Identifying a research problem 7
3. Surveying the literature 11
4. Analyzing the research problem 13
5. Developing general and specific objectives 16
6. Designing a study protocol 19
    6.1 Selecting an appropriate study design 19
    6.2 Defining the study population 22
    6.3 Selecting the study sample 24
        6.3.1 Estimating sample size 24
        6.3.2 Selecting a sampling technique 26
    6.4 Defining variables, indicators, and measures 31
        6.4.1 Tips on questionnaire development 35
        6.4.2 Validity and reliability 37
        6.4.3 Measurement error 38
    6.5 Planning for data collection 39
        6.5.1 Pilot study 40
    6.6 Planning data analysis 40
        6.6.1 Descriptive statistics 40
        6.6.2 Inferential statistics 41
    6.7 Ethical considerations 44
7. Writing up the research proposal 45
    7.1 A note on plagiarism 50
    7.2 Referencing 51
    7.3 Reference Management Software 51
References 52

Preface

This research guide outlines the requirements and expectations of the Department of Community and Family Medicine (DCFM) in relation to the undergraduate research project – a core component of the undergraduate community medicine curriculum at the Faculty of Medicine, University of Jaffna. The ultimate aim of the guide is to make research an enjoyable and rewarding process for students, and, in turn, motivate them to become life-long learners by integrating research into day-to-day medical practice.

This guide is inspired by the late Prof. C. Sivagnanasundaram, the pioneering head of the then Department of Community Medicine, Faculty of Medicine, University of Jaffna, who worked tirelessly to guide and motivate hundreds of medical students in undergraduate research, including two of the authors of this guide—P.A.D. Coonghe and R. Surenthirakumaran.

We acknowledge the feedback received on the undergraduate research process from medical students, external supervisors, and members of the Ethics Review Committee, Faculty of Medicine, all which has been integrated into this revised and updated version of this research guide. We thank Prof. A. Pathmeswaran, Department of Public Health, Faculty of Medicine, Ragama, for his contribution at the Workshop for Research Supervisors, held at the Faculty of Medicine, University of Jaffna, in 2018. Thanks also to Aathithya Kugathasan and Sanduni Ayeshika, the students of the 36th Batch whose work is cited as an example throughout the guide.

R. Kumar, P.A.D. Coonghe, R. Surenthirakumaran
December 31, 2020

1. Introduction to the undergraduate research project

The community medicine research project is carried out in the third and fourth years of the undergraduate medical programme at the Faculty of Medicine, University of Jaffna. You are expected to complete your research activities and submit a final research report before sitting for the Second Examination for Medical Degrees. The research project accounts for 15 per cent of the final mark in Community and Family Medicine at the Second Examination for Medical Degrees.

1.1 Learning outcomes

By the end of the research activity, you should be able to:

- Demonstrate knowledge of basic research concepts at the undergraduate level;
- Conceptualize, design and implement a research project at the undergraduate level;
- Assess scientific evidence and arrive at conclusions by way of logical deduction; and
- Contribute towards advancing the practice of evidence-based medicine.

1.2 Supervision

You will undertake the research project in groups of five. An internal supervisor will be assigned from the Department of Community and Family Medicine (DCFM) to each group. After you identify a research problem, you will meet your internal supervisor who will advise you on selecting an external supervisor with relevant expertise. External supervisors may be members of academic staff at the Faculty of Medicine or other faculties of the University of Jaffna, a consultant or an experienced medical officer at the Teaching Hospital Jaffna or other hospitals in the Northern Province, or a public health practitioner based in the Northern Province.

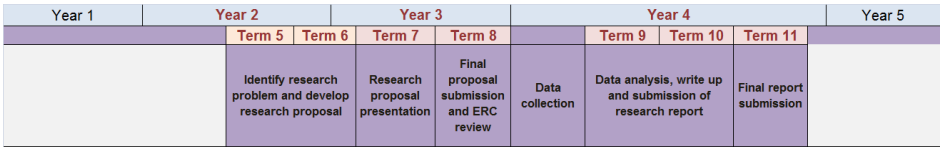
You are expected to demonstrate professionalism in contacting prospective external supervisors, discussing the possibility of supervision, obtaining

their consent, and in interacting with supervisors thereafter. In turn, supervisors will offer guidance, direct you to relevant experts and other resources, and provide feedback on the *Monitoring of Student Research Projects by Supervisors Form*. DCFM places the responsibility for initiating meetings, communicating, and completing the monitoring form, with the students. Please remember to obtain your supervisors' signatures at the end of each meeting.

1.3 Timeline

The research project spans terms 5 through 10 of the undergraduate medical curriculum (Figure 1). You are expected to identify your research problem, develop a research proposal, and submit the research proposal in terms 5 through 7. In term 7, you will present your proposal to DCFM, attend to revisions, and then present your proposal to the Ethics Review Committee (ERC), Faculty of Medicine, University of Jaffna. Once ERC approval is obtained and the final proposal is submitted, you will be granted time for data collection activities between terms 8 and 9. Data analysis commences in the early part of term 9. You are expected to write up the research report (on which you will be graded) and submit it to DCFM by the end of term 9/beginning of term 10. The research report will be returned to you for correction, and you will submit the final research report to DCFM before sitting for the Second Examination.

Figure 1. Research project timeline



1.4 Assessment

Your research activities will be assessed formatively by your supervisors throughout the research activity.

Three formal summative assessments will take place:

1) Final research proposal and proposal presentation – 20 %
2) Final research report – 50 %
3) Research project viva – 30 %

Together, these marks will contribute to 15% of your final grade at the Second Examination for Medical Degrees.

2. Identifying a research problem

The research problem you select should be relevant to one or more of the following disciplines: community medicine/public health, family medicine, anatomy, physiology, biochemistry, microbiology, pharmacology, pathology, parasitology, medicine, surgery, paediatrics, obstetrics, gynaecology, psychiatry, other medical specialties, medical education or medical sociology.

You may identify research problems based on your own experiences in school, at the university, in the hospital, or in the community. It could be a problem observed during a clinical attachment or a contemporary social issue. Alternatively, you could identify a problem by speaking with health professionals or based on your reading of the literature. A research problem is essentially "a perceived gap between what is and what should be" or "a difficulty or obstacle … between a present situation and a desired future objective" (Sivagnanasundaram 1999:51). Often there may not be data available on the "present situation" in Jaffna, and this, itself, could be a research problem (Sivagnanasundaram 1999).

Most importantly you must select a *researchable* problem. The FINER criteria may be used to evaluate whether your chosen research problem is researchable or not (Hulley et al. 2007). These criteria ask whether the research problem is: **F**easible, **I**nteresting, **N**ovel, **E**thical and **R**elevant (Table 2).

Feasibility: First, it is important that you narrow down the scope of the research problem so that your study can focus on achieving two to four tightly defined specific objectives. Second, the availability of research participants is important because you have fixed deadlines. Other feasibility considerations include whether you have the skills, experience, and equipment required to undertake the study, and the expenses associated with the study.

Table 1. FINER Criteria for a good research question

Criterion	Characteristics
Feasible	Manageable in scope
	Adequate number of subjects
	Adequate technical expertise
	Affordable in relation to time and funds
Interesting	Getting the answer intrigues the investigators
Novel	Confirms, refutes or extends previous findings
	Provides new findings
Ethical	Will receive ERC approval without difficulty
Relevant	To scientific knowledge
	To clinical and/or health policy
	To future research

Adapted from Hulley (2007)

Interest: It is extremely important that you select a research problem that will keep you interested through the research activity (Hulley et al. 2007). Therefore, *you* should select your research problem under the guidance of your supervisor. it is not desirable for your supervisors to select the research problem for you. Their role is to guide you towards a research problem that is of relevance to public health/medicine and that will advance the status of knowledge on the topic.

Novelty: A study that duplicates an existing study is a waste of time and resources. However, as Hulley and colleagues (2007:21) point out, "a research question need not be totally original – it can be worthwhile to ask whether a previous observation can be replicated, whether the findings in one population also apply to others, or whether improved measurement techniques can clarify the relationship between known risk factors and disease."

However, it is important that you avoid duplicating research. A list of research projects undertaken by medical students as the Community and Family Medicine research project (from the 7th Batch onwards) is available on the DCFM web page under the 'Research' tab (http://www.med.jfn.ac.lk/index.php/community-medicine/research/).
Please ensure that you select a research problem that will make a unique

contribution to the status of knowledge on your topic. Please avoid adapting a study design developed by seniors to a different population. Think outside the box!

Ethics: It is desirable that the study involves minimal risk to study participants. Some aspects to consider are: physical or mental risks and invasion of privacy. For instance, a study on unsafe abortion may not be feasible given the timeframe, difficulty in accessing participants, and the sensitive nature of the topic that demands the skills of an experienced researcher. A practical consideration is that if you select a topic that involves greater risk to participants, you may take longer to obtain ethical clearance.

Relevance: The relevance of a research question may be assessed by thinking about the impact of the findings, particularly whether and how the findings will inform future research, policy and/or practice. In other words, a good research question should pass the 'So what?' test by contributing meaningfully towards advancing knowledge in the research area (Hulley et al. 2007).

Here are some examples of research problems:

- Unhealthy food habits are increasing among school-going adolescents in Jaffna
- Patients with chronic renal spend a lot of money on healthcare in Sri Lanka
- Falls among institutionalized elders is a growing problem in Jaffna

Your research problem could be stated in a sentence, but specify: 'what, who, and where.' What is the problem? Who experiences it? And, in what setting? Once you submit your research problem, you will be assigned and an internal (DCFM) supervisor with whom you will discuss your research problem and make changes, as necessary. Suitable external supervisors will be identified thereafter.

Suggested reading:

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. 2007. *Designing Clinical Research*, 3rd Ed. Philadelphia, PA: Lippincott Williams & Wilkins. (Available at FOM Library)

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

3. Surveying the literature

Having selected a research problem, the next step is to conduct a literature survey. A good literature survey will help you better understand the research problem (its magnitude or severity, who is affected and where, factors known to be associated with the problem, the impact of the problem, etc.). This initial survey will also help you improve your understanding of fundamental concepts relevant to the research topic, which may be subsequently applied to the research problem.

A common mistake is to confine the literature survey to Google. While you can obtain an understanding of the current state of knowledge on the research topic by accessing online journal publications, such as those available via Google or Google Scholar, you may miss the literature/sources relevant to Sri Lanka, and, particularly the Northern Province or Jaffna. This is why **consulting a librarian, speaking with local experts, contacting local authorities, including the Office of the Regional Director of Health Services**, may be crucial components of your literature survey.

You may also find the following sources helpful:

- Ministry of Health – Annual Health Bulletin and other publications
- Epidemiology Unit – Publications
- Family Health Bureau – Annual Reports
- National Programme for Tuberculosis Control & Chest Diseases – Annual Reports
- National STD/AIDS Control Programme Sri Lanka – Annual Reports
- National Dengue Control Unit – Publications
- Antimalaria Campaign – Annual Reports
- Epidemiology Unit – Polio Eradication Programme
- Dept. of Census and Statistics – Census 2012 and Demographic and Health Survey 2016
- National Science Foundation – Sri Lanka Journals Online

- Statistical Handbooks published by the Provincial Department of Health Services Northern Province and Office of the Regional Director of Health Services Jaffna are available at the Faculty of Medicine Library.
- University of Jaffna – Subscribed e-resources
- World Health Organization – Global Health Observatory Data and World Health Statistics
- World Bank- Data Bank
- Wiley Online Library – The Cochrane Library
- US National Library of Medicine - PubMed
- Google – Google Scholar

If you are unable to obtain the information you need, please communicate with your supervisors.

Suggested reading:

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research (http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf).

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

4. Analyzing the research problem

Having conducted a thorough literature survey, you are now prepared to analyze your research problem at a deeper level. Problem analysis involves brainstorming and pooling existing knowledge on the research problem, and helps to identify the scope of the study, its objectives, the study variables (areas that will be covered by data collection instruments), and measures. Problem analysis is an extremely important part of research because, as depicted in Figure 2, it shapes the entire research process.

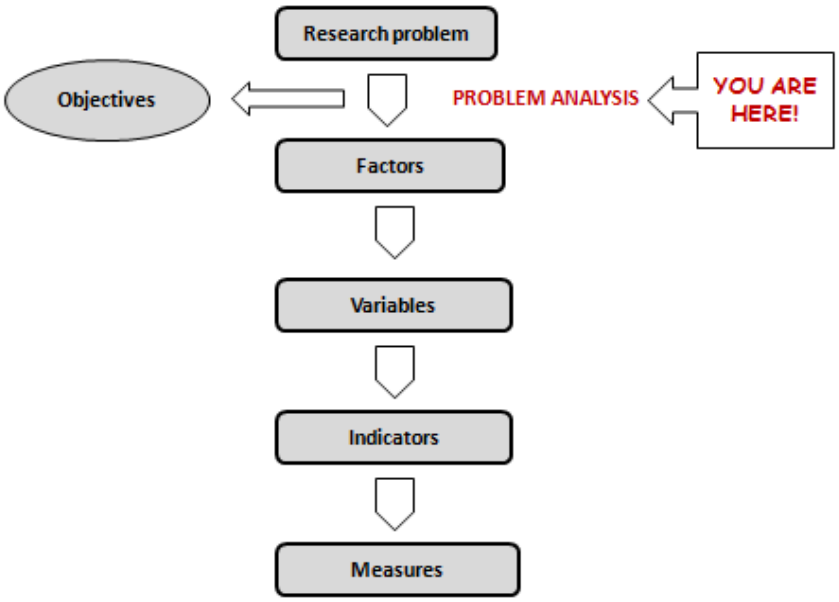


Figure 2. From research problem to indicators and measures

The literature survey is an ongoing process during problem analysis; access as many resources, not only literature but also expert opinion, to achieve a comprehensive level of analysis. Reflecting the importance of problem analysis, you are required to include an image of the problem analysis chart in your initial proposal presentation.

4.1 Steps of problem analysis

Step 1 – Clarify the problem → What is the problem? State precisely.

Step 2 – Describe the problem → What is its nature, distribution, size/intensity?

Step 3 – Identify contributing factors → Develop a problem analysis chart (Figure 3)

Step 4 – Decide on the focus and scope of the study → Which aspects of the problem will you study? Select the factors based on importance and feasibility (refer FINER criteria on p.6).

A common mistake is to select a research title before carrying out problem analysis. **Do not do this because you will have difficulties designing a coherent research project and you may miss important aspects that need to be addressed by your study objectives.**

How to develop a problem analysis chart:

- Draw a central bubble that contains the problem (primary bubble)
- Identify key factors that contribute to the problem (secondary bubbles);
- Identify contributory factors that shape the key contributory factors (tertiary bubbles);
- Connect the bubbles based on the relationships between factors; all bubbles connect to the primary bubble; a tertiary bubble may connect to a secondary bubble as well as the primary bubble; and each bubble may connect to others at the same level.
- Continue expanding levels of bubbles until the problem is comprehensively analyzed.
- Now organize related factors into larger categories (e.g., sociodemographic factors, service-related factors, etc.) and develop the final chart.

Figure 3. Developing a problem analysis chart (Adapted from Singh et al., 2008).

A model problem analysis chart is given below. Note the way in which all factors are connected to the core problem and the factor categories are placed peripherally.

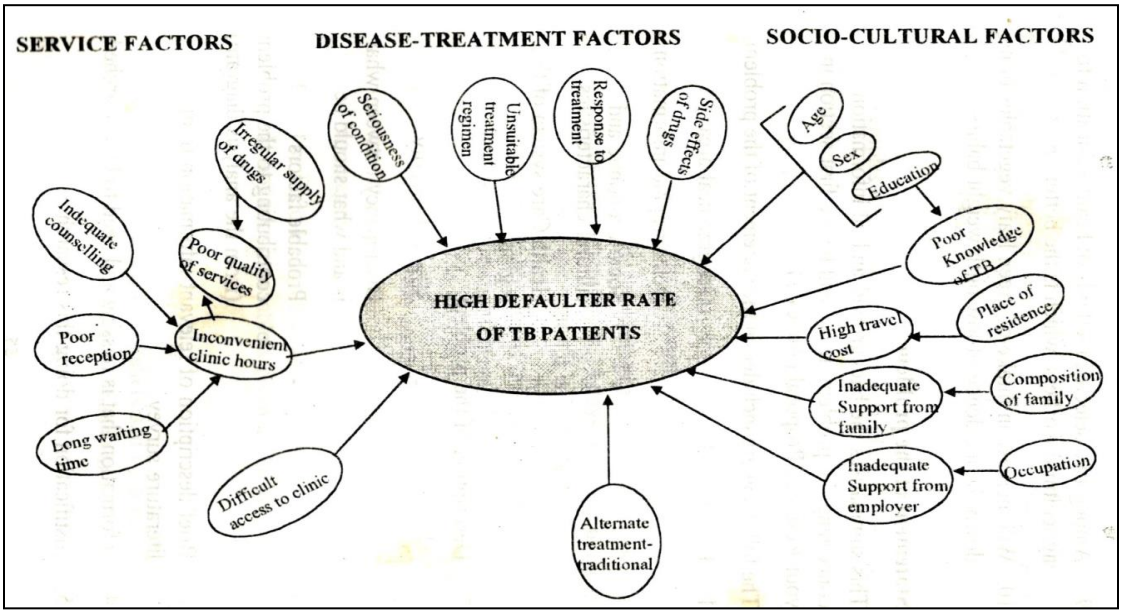


Figure 4. Problem analysis of high default rate among patients with TB (Extracted from: Sivagnanasundaram, 1999)

Suggested reading:

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research (http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf).

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

5. Developing general and specific objectives

In descriptive studies, where the relationship between variables cannot be assumed, you state general and specific objectives. Some studies predict a relationship between two or more variables. In such instances, a hypothesis is stated (Sivagnanasundaram 1999). Studies that begin with a hypothesis are analytical in nature. Most undergraduate projects are descriptive, with many including analytical components. It is best to develop your specific objectives before your general objective.

5.1 Specific objectives: In a descriptive study, the first specific objective usually deals with the magnitude of the problem, while the rest focus on its contributory factors. Developing SMART (S-specific; M-measurable; A-achievable; R-realistic; and T-time bound) specific objectives will help you define the scope of your study. They will guide you from start to finish, including your literature review, methodology, results, discussion, and conclusion. The specific objectives keep the study focused on essentials, prevent unnecessary data collection, and, most importantly, break the study into 'doable' parts.

S	**Specific**
M	**Measurable**
A	**Achievable**
R	**Realistic**
T	**Time bound**

Formulating your specific objectives enable you to: define the sample in more specific terms; write inclusion and exclusion criteria; define variables, categories and measures; select and modify study instruments; keep data collection focused; construct dummy tables for analysis; select appropriate statistical tests; discard unnecessary information gathered from the literature survey; interpret the findings; and write up a coherent research report (Sivagnanasundaram 1999).

The specific objectives need to be listed in logical sequence; 3 to 4 specific objectives are adequate.

In a study of dysmenorrhoea among A/L students undertaken by a group of students in the 36th Batch, the following specific objectives were stated:

- *To determine* the prevalence of dysmenorrhoea among A/L students in the Jaffna Educational Division;
- *To assess* the severity of dysmenorrhoea among A/L students in the Jaffna Educational Division;
- *To describe* menstrual characteristics associated with dysmenorrhoea among A/L students in the Jaffna Educational Division; and
- *To assess* the influence of menstrual characteristics on the presence and severity of dysmenorrhoea among A/L students in the Jaffna Educational Division.

In surveying the literature, the students did not find evidence that sociodemographic factors play a significant role in the experience of dysmenorrhoea. Therefore, the study focused on studying menstrual characteristics known to be associated with dysmenorrhoea.

When finalizing the specific objectives, you should ask yourself: *Are the specific objectives clear, precise and simple? Do they answer the research question? Do they address important factors identified in the problem analysis? Are the action verbs* (italicized in the list above) *appropriate?* (Sivagnanasundaram 1999). Common mistakes include, listing specific objectives that are not relevant to the general objective; mismatch between problem analysis and specific objectives; and use of inappropriate action verbs. Please take care in selecting action verbs as they indicate specific research actions and should not be used interchangeably (e.g., determine, assess, describe, etc.). Please ask your supervisors for guidance.

5.2 The general objective: The general objective encompasses all specific objectives, and should be stated as simply as possible. For example, the general objective of the study on dysmenorrhoea mentioned earlier was stated as follows: To determine the prevalence and assess the severity of

dysmenorrhoea and associated menstrual characteristics among A/L students in the Jaffna Education Division.

Suggested reading:

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research (http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf).

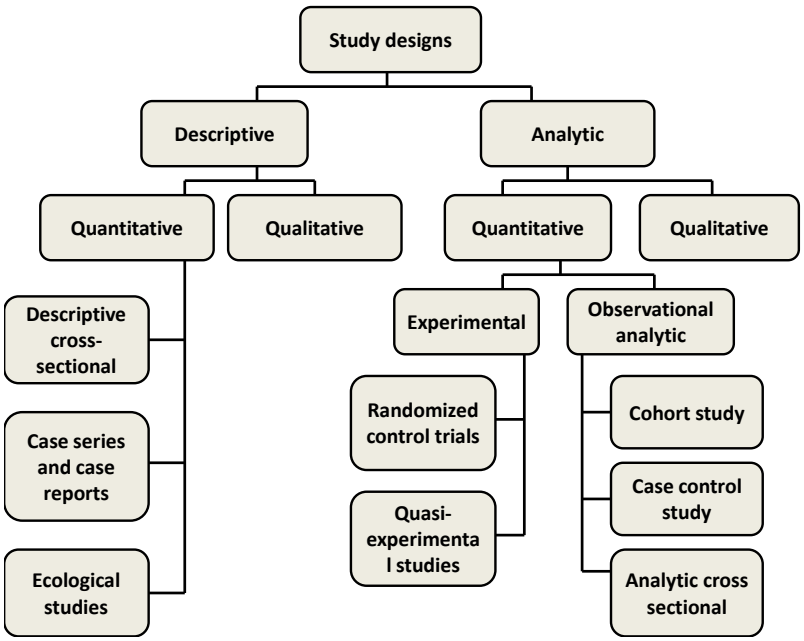
Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

6. Designing a study protocol

6.1 Selecting an appropriate study design

Study designs are classified in various ways (Figures 5 and 6). As depicted in Figure 5, they may be broadly divided into descriptive and analytic study designs, which, in turn, may be quantitative or qualitative. Because you are expected to use a quantitative study design for the purposes of the DCFM undergraduate research project, this guide focuses on designing a quantitative study.[1]

Descriptive studies involve assessing a sample at any given time to learn about the characteristics of the population, phenomenon, or observation of interest. There is no comparison group; making causal inferences is not an objective of research. By contrast, analytic studies test hypotheses about exposure-outcome relationships, and involve making comparisons between two or more groups.



[1] You may combine both quantitative and qualitative methods in a mixed methods design, if you wish to develop your qualitative research skills.

Figure 5. Types of study designs

Epidemiological study designs are mostly quantitative and are divided into **observational** and **experimental** designs (Figure 6). In an observational study, the researcher studies/observes phenomenon but does not alter what occurs. By contrast, the researcher intervenes to assign exposures and observe outcomes in an experimental study (The Open University 2018; University of Ottawa n.d.). Observational studies include cohort studies, case control studies, cross-sectional studies, case reports/case series and ecological studies. Experimental studies include randomized control studies and quasi-randomized control studies (Grand Canyon University n.d.).

Carrying out a **cross-sectional study** involves drawing a sample from the selected study population and recording information from participants in a systematic way during a given time. Cross-sectional studies are well-suited for describing variables and their pattern of distribution. They may be used to determine **prevalence** (the number of cases of a disease that are present in a particular population at a given time expressed as a proportion) at either one point in time (point prevalence) or over a defined period of time (period prevalence).

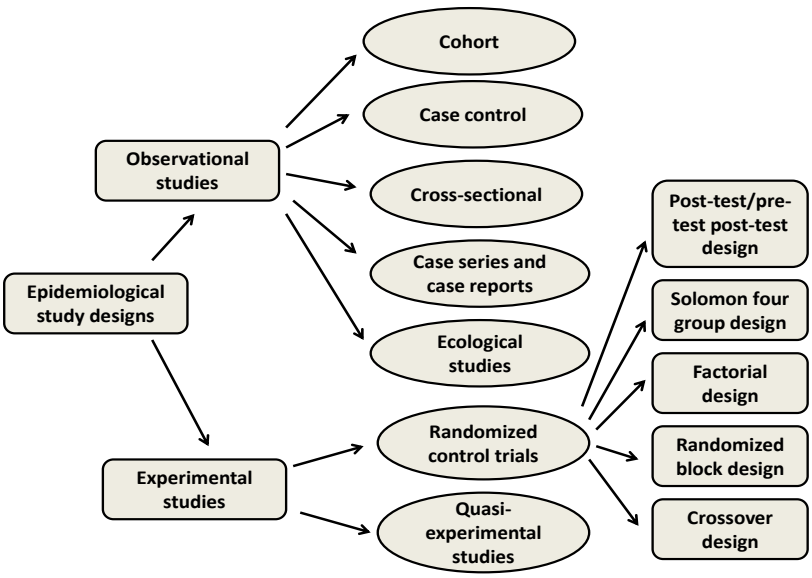


Figure 6. Types of quantitative research designs

Period prevalence is determined when it takes time to obtain sufficient information on a disease in a population, e.g., what proportion of 35-year-old women residing in Nallur MOH Area undergo Pap smear over a year.

Cross-sectional studies may be used to examine associations; the proportion with the outcome is compared between those with and without the exposure. These are called analytic cross-sectional studies. In other words, analytic studies may compare the proportion of exposed persons who are diseased with the proportion of non-exposed persons who are diseased to assess the relationship between exposure and outcome. PLEASE NOTE: it is not possible to make causal inferences using a cross-sectional design because data on exposure and outcome are collected simultaneously. The labeling of variables as exposures and outcome in a cross-sectional study depend on the cause-effect hypotheses of the researcher. (Hulley et al. 2007; Sivagnanasundaram 1999; University of Ottawa n.d.).

The research design you select will depend on your research objectives and the availability of time and funds. Most students select a descriptive cross-sectional study design for the purposes of the DCFM research project. If you decide to undertake a descriptive cross-sectional study, it is desirable to include an analytical component to ensure that you gain some experience with the application of inferential statistics to research. You may select a different study design, if you wish, but be mindful of the time constraints associated with the research project.

Suggested reading:

Alexander, L.K, Lopes, B, Riccheti-Masterson, K, Yeats, K.B. (2015). Cross-sectional studies. https://sph.unc.edu/files/2015/07/nciph_ERIC8.pdf

Grand Canyon University (n.d.). Types of experimental research. https://cirt.gcu.edu/research/developmentresources/research_ready/experimental/design_types

Grimes, D. A., & Schulz, K. F. (2002). An overview of clinical research: the lay of the land. *The Lancet*, 359(9300), 57-61 http://www.geocities.ws/mim_ebm/LancetEpi-01.pdf.

Petrie, A. & Sabin, C. (2009). Medical statistics at a glance. https://leseprobe.buch.de/images-adb/28/42/28428804-b42f-417d-b28e-e407f7fd73ff.pdf

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

The Open University (2018). Types of epidemiological studies. http://www.open.edu/openlearn/science-maths-technology/science/health-sciences/epidemiology-introduction/content-section-3.2

6.2 Defining the study population

The **theoretical population** is the entire group of people (or objects) you wish to generalize the study findings to. For example, your theoretical population may be all adolescents aged 18 years in the Northern Province. The study population is the entire collection of *possible* observations that you will have access to (Figure 4). Returning to the same example, the study population could be defined as all Grade 11 students registered at schools in the Northern Province. The study population is made up of study units. Note that the latter do not always refer to people, and may include other units such as families, households, operating theatres, clinics, and so forth.

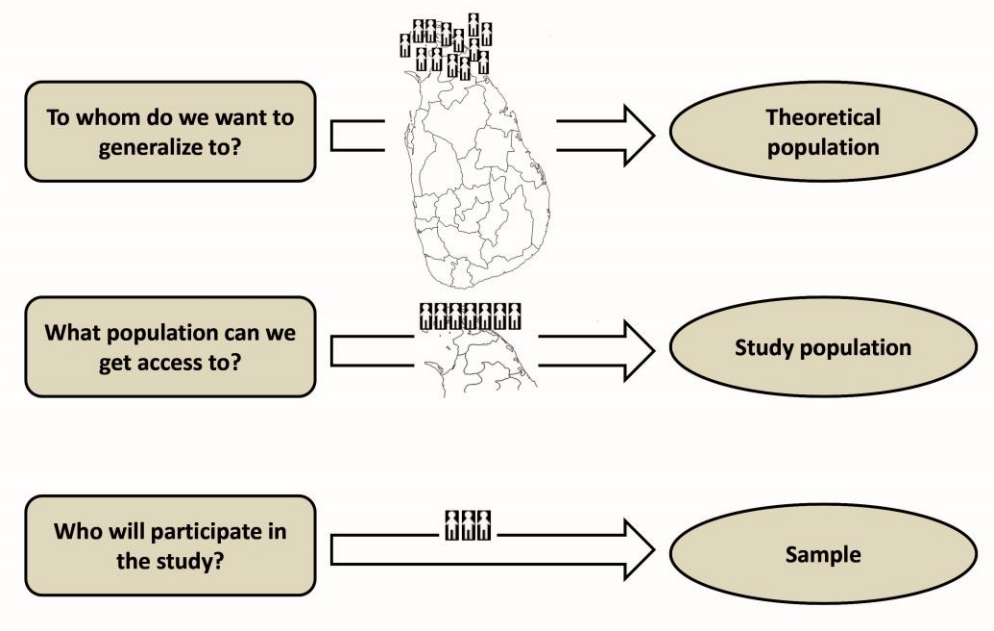


Figure 7. Theoretical population versus study population versus sample
(adapted from Web Centre for Social Research Methods 2006)

6.2.1 Inclusion and exclusion criteria

The study population is defined by inclusion and exclusion criteria. Inclusion criteria are attributes of participants that are essential for them to be eligible to participate in the study. For example, the inclusion criteria for a study of knowledge on sexually transmitted infections among A/L students in the Northern Province might be: being a student in an A/L class of a school in the Northern Province for at least 2 years. On the other hand, the exclusion criteria are attributes of eligible participants that disallow their participation. Returning to the same example, if the study instrument is a self-administered questionnaire, those who are unable to read and complete a questionnaire on their own may serve as exclusion criteria. While this would exclude students with severe visual impairment, the ethical implications of exclusion must be considered. Students being absent on the day of data collection and/or unwillingness to participate are <u>not</u> exclusion criteria.

6.3 Selecting the study sample

6.3.1 Estimating sample size

As the entire study population cannot be enumerated in a reasonable period of time, a subset of the population, or a sample, needs to be selected. Before sampling, the researcher must estimate the minimum sample size required to obtain a valid result. At the undergraduate level, you are not expected to be familiar with the complexities of sample size calculation. But you should be familiar with concepts that are used in estimating sample size, including the expected proportion, the level of statistical significance, margin of error, and power.

Expected proportion refers to the proportion the researcher expects to find in a prevalence study. Researchers often draw on previous studies to make an informed estimation of the effect size, or, when data are not available, undertake a pilot study to estimate the proportion.

The **level of statistical significance** or α is the probability of making a Type I error (rejecting the null hypothesis when it is true), usually set at 0.05, which means the researcher has set the maximum probability of incorrectly rejecting the null hypothesis at 5%.

The **margin of error** refers to the maximum expected difference between the true population parameter and a sample estimate.

β is the probability of making a Type II error (failing to reject null hypothesis when it is not true). The quantity (1- β) is referred to as the **power** of the study or the probability of correctly rejecting the null hypothesis (Hulley et al. 2007). In other words, the study's power is the probability that a statistically significant difference is found between the groups when it exists; a greater power requires a larger sample size. β is generally set at 0.2, which means that the researcher is willing to accept a 20 per cent chance of missing an association of a given effect size if it exists. This translates to a power of 80% or an 80% chance of finding an association of

that size or greater. Note that power is not included in the sample size equation of a prevalence study.

A key point is that sample size should be estimated in the design phase of the study. Sample size planning entails selecting an adequate number of subjects to keep α and β acceptably low within the limits of feasibility. This formula is used to calculate sample size in a prevalence study:

$$n = \frac{Z^2 \times p(1-p)}{d^2}$$

Where,

n = Sample size

Z = Critical value of the normal distribution at $\alpha/2$ (for a confidence level of 95%, α is 0.05, and the critical value is 1.96)

p = Expected proportion (based on results of a previous survey)

d = Acceptable margin of error at 5% (0.05)

Add 10% to the estimated sample size to account for non-responders.

Students usually draw the expected proportion from a previous survey. The study should preferably be carried out in Sri Lanka. If a similar study has not been undertaken in Sri Lanka, you could use a proportion from a study that has been undertaken in a similar socio-cultural context. If such a study is not available, you may use 50 per cent, which gives the maximum sample size for a given margin of error.

If a prevalence study seeks to estimate the prevalence of more than one variable (e.g., the prevalence of >1 factors associated with a given disease), then the sample size for the study should be decided based on the expected proportion of the variable that requires the largest sample size (Hulley et al. 2007).

Sample size impacts the *margin of error* or the level of precision of a study. A narrower margin of error requires a larger sample size. Conversely, reducing the margin of error in the sample size equation, will increase the

required sample size. The expected proportion should be considered when setting the margin of error. For instance, a small expected proportion (such as 1%) would require a margin of error equal to or less than 1%, which translates to a very large sample size.

6.3.2 Selecting a sampling technique

Sampling is the process of selecting study units (e.g., people, households, schools, etc.) from a population of interest. In quantitative research, this is usually undertaken in a manner that allows the researcher to generalize the results to the theoretical population represented by the study population.

6.3.2.1 Probability sampling

Sampling methods fall into two broad categories: probability sampling and non-probability sampling. Probability sampling allows researchers to generalize the findings from the study sample to the theoretical population. Probability sampling methods use a random process where each study unit has a specified chance of being included in the sample. It allows for estimating the likelihood of the study findings representing the characteristics of the theoretical population by significance testing and estimation of confidence intervals (Hulley et al. 2007). Widely used methods of probability sampling include: simple random sampling, systematic sampling, stratified sampling, and cluster sampling.

A **simple random sample** is drawn by listing all study units in the study population and selecting a subset with the help of a random number table or a random number generator. Each member in the study population has an equal chance of being selected to participate in the study (Figure 8). The feasibility of this method is low, particularly in large scale population-based studies where complete lists of study populations (or sampling frames) are not available. Where available, the geographic spread of the population poses practical difficulties for the researcher. However, this method could be used where the sampling frame is readily available and geographic spread is not a concern. For instance, to select a random sample of pregnant women registered with Public Health Midwives in the Nallur MOH

Area, you could obtain a list of all pregnant women residing in the Nallur MOH Area during a given time from the Nallur MOH Office and then use a random numbers table to select a random sample.

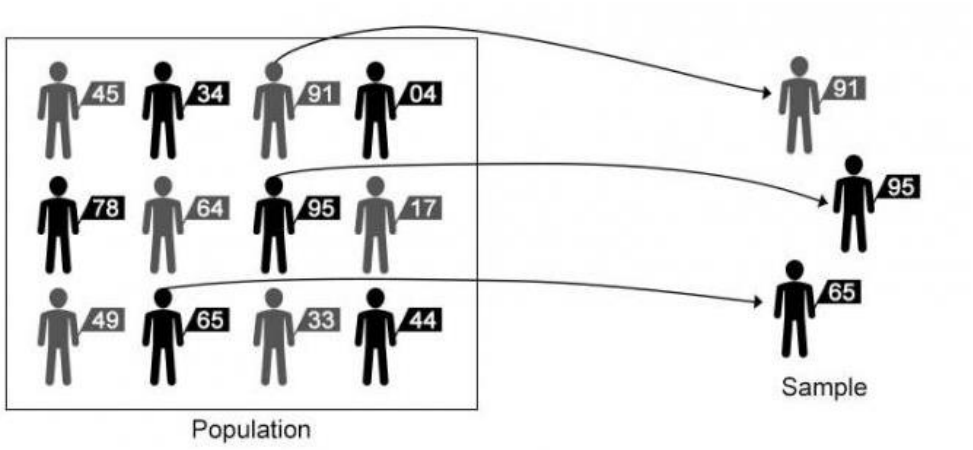


Figure 8. Simple random sampling

A **systematic sample** is advantageous when the study population cannot be readily enumerated. Instead of using random numbers, this method uses a pre-defined periodic process (e.g., every third individual – Figure 9) to select the study sample. This process is open to manipulation by the investigator and may also be influenced by natural periodicities in the population (Hulley et al. 2007). An example of systematic sampling would be to select every third patient who visits the OPD at DH Kondavil. This number is defined based on the estimated sample size and the availability of participants.

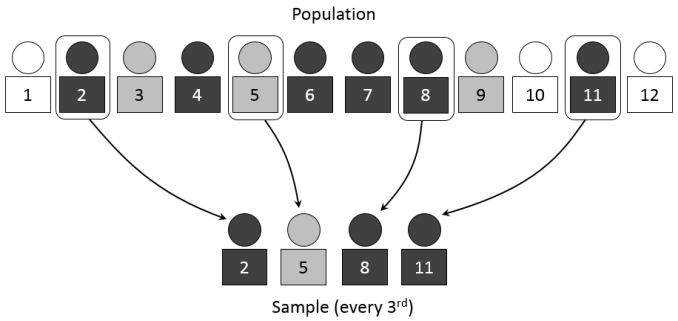


Figure 9. Systematic sampling

A **stratified sample** allows for stratification of the study population by various characteristics in order to ensure inclusion of subsets of the population that are of interest to the researcher (Hulley et al. 2007). Returning to the example of pregnant women in the Nallur MOH Area, you may wish to study how a variable varies with age. Stratified random sampling would allow you to obtain adequate numbers of pregnant teenagers and pregnant mothers over forty, who may be few in the study population, by dividing the population into subgroups based on age (<20 years, 20-39, ≥ 40 years) and taking a random sample from within each stratum of the population (Figure 10). *Proportionate stratified sampling* would involve selecting participants to match their proportions in the study population, while a *disproportionate sample* would draw the number of participants required for statistical analysis.

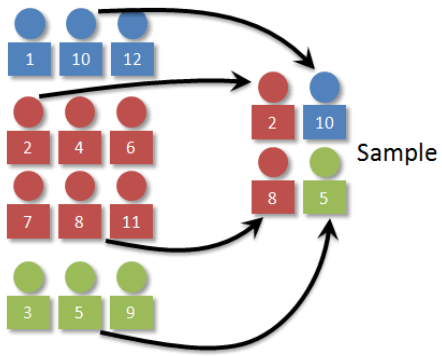


Figure 10. Stratified sampling

Cluster sampling is widely used in population-based research to manage the geographic spread of the population under study. For instance, if you wished to study the nutritional status of O/L students in the Jaffna District, cluster sampling would involve selecting naturally occurring clusters (e.g., schools, classes) (Figure 11) rather than randomly selecting participants from a list of all O/L students in the Jaffna district. Cluster sampling may be carried out at multiple stages. For instance, returning to the example of O/L students, the investigators would select schools and then select classes within the selected schools.

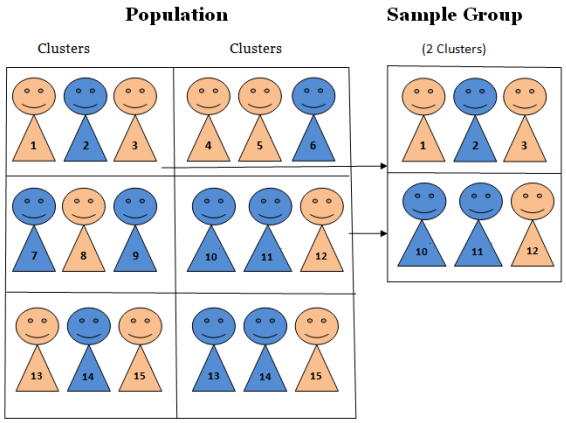


Figure 11. Cluster sampling

Cluster sampling does increase feasibility and substantially lowers costs, but naturally occurring clusters tend to be homogenous. As the number of stages increases, the cluster effect increases, as does the required sample size (Hulley et al. 2007). In other words, the cluster effect needs to be accounted for in calculating sample size. While there are statistical methods used to calculate the cluster effect, you are not expected to be able to do this at the undergraduate level. Thus, if you use cluster sampling, double your estimated sample size to account for the cluster effect.

6.3.2.2 Non-probability sampling methods

Commonly used non-probability sampling methods include: convenience sampling, purposive sampling, and snow-ball sampling. A **convenient sample** is made up of people who meet the inclusion criteria and are easily accessible to the investigator. An example of convenient sampling is consecutive sampling where participants are selected on a consecutive basis. For instance, the selection of each patient who arrives at the out-patient department of DH Kondavil during a specified period.

Purposive sampling is widely used in qualitative research, and permits the researcher to select participants to serve a specific purpose. For example, in a study of patient satisfaction at the Teaching Hospital Jaffna, the researcher may want to approach the research question from a spectrum of social positions. This form of purposive sampling is called maximum

variation sampling. Purposive sampling may be undertaken to serve other purposes.

Snowball sampling is generally employed to access participants who are difficult to reach. For example, we may want to explore the healthcare experiences of a stigmatized population (e.g., LGBTQI persons). It is likely that members of this social group may have contact with others of a similar orientation. Snowball sampling involves drawing on the networks of participants we recruit to our study to access more participants. It is called snowball sampling because a snowball gathers more snow and becomes larger when it rolls downhill. Whether non-probability sampling is appropriate or not depends on the methodological approach adopted and the purpose of the research.

Suggested reading:

Gregg, M.B. (2002). Field Epidemiology, 2nd Ed. Chapter 8 http://www.ciphi.ca/hamilton/Content/documents/fieldepi.pdf.

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. 2007. *Designing Clinical Research*, 3rd Ed. Philadelphia, PA: Lippincott Williams & Wilkins. (Available at FOM Library)

Kumar, R. (2011). *Research Methodology: A Step-by-Step Guide for Beginners*. London, UK, Thousand Oaks, CA, New Delhi, India and Singapore: SAGE http://www.sociology.kpi.ua/wp-content/uploads/2014/06/Ranjit_Kumar-Research_Methodology_A_Step-by-Step_G.pdf.

Petrie, A. & Sabin, C. (2009). Medical statistics at a glance. https://leseprobe.buch.de/images-adb/28/42/28428804-b42f-417d-b28e-e407f7fd73ff.pdf

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

Web Centre for Social Research Methods (2006). Sampling terminology. https://www.socialresearchmethods.net/kb/sampterm.php

6.4 Defining variables, indicators, and measures

The problem analysis would have led you to identifying several groups of factors that influence/are influenced by the research problem (pp. 11-12). At this stage, you would have selected the factors that you wish to focus on in your study. This selection should have been guided by a survey of the literature, discussions with experts, and may be your own experiences. Your literature survey should have covered both international and local literature to ensure you did not miss important factors that are known to influence/be influenced by your core problem.

In quantitative research, these factors need to be converted into variables, and the variables are subsequently operationalized into indicators and measures (Figure 12). Variables take on varying values; they may be a characteristic, object or phenomenon that is measurable. For example, sociodemographic factors, such as age, sex, education level, and income level, are variables.

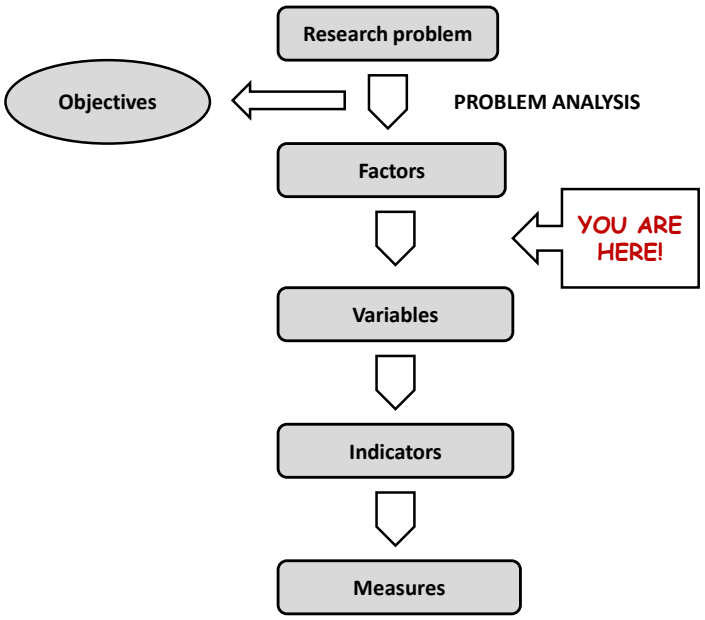


Figure 12. From factors to variables, indicators, and measures

Variables are classified as **categorical** (qualitative) and **numerical** (quantitative) variables (Figure 13). In turn, categorical variables may be divided into nominal and ordinal variables. **Nominal variables** have two or more categories that cannot be ordered (e.g., sex/gender, ethnicity, stream of study, etc.). **Ordinal variables** have two or more ordered categories that may be arranged from highest to lowest or vice versa (e.g., education level, stress levels, etc.).

Numerical variables are divided into discrete and continuous variables. **Discrete variables** are those that can take a finite number of values; they do not have values between values (e.g., number of children in a family → may be 1 or 2 but not 1.5). By contrast, **continuous variables** may take on an infinite number of values between values (e.g., weight, height → maybe 1, 1.1, 1.11, and so on).

Variables may also be classified as: independent, dependent, and controlling variables. Understanding the difference between independent and dependent variables is important. The researcher designs the study to observe changes in the **dependent variable** by using varying values of the **independent variable** (Singh et al. 2006). For example, the researcher may select participants of various age groups (independent) to find out how diet (dependent) may change with age.

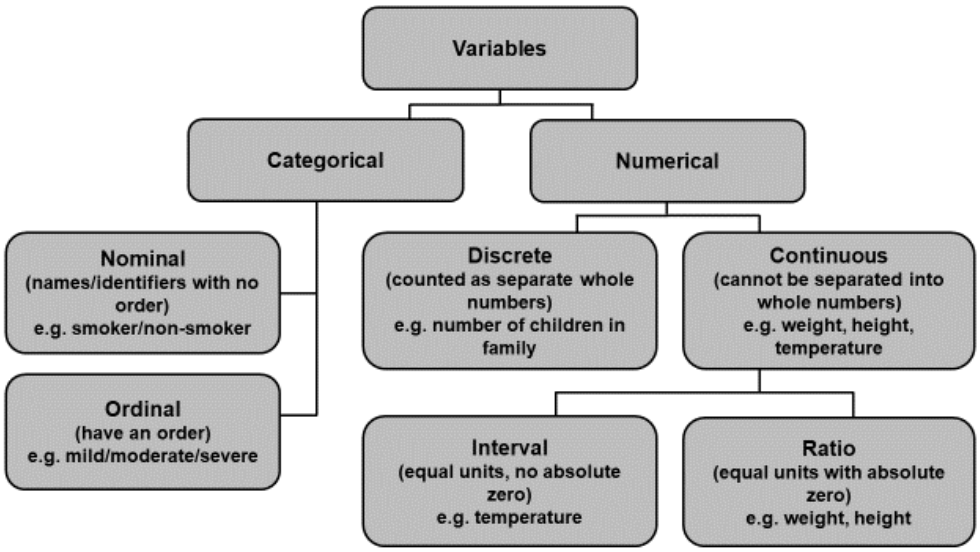


Figure 13. Types of variables

(adapted from Sivagnanasundaram 1999)

Controlling variables are extraneous variables that the researcher does not wish to investigate. However, these variables need to be kept constant to investigate the true nature of the relationship between independent and dependent variables. For instance, in the study on diet and age, family income may influence the availability of food and, in turn, diet, and may need to be controlled for. You are not expected to control for variables for the purposes of the undergraduate research project.

Indicators are formed by defining and stating variables in operational terms (Sivagnanasundaram 1999). In developing indicators, you need to clearly define the variables and their attributes. An **attribute** is a specific value of a variable. The attributes of a variable must include all possible responses ('exhaustive') and respondents should not be able to respond to two attributes simultaneously ('mutually exclusive') (Web Centre for Social Research Methods 2006a).

For example, tobacco use (variable) may be operationalized as 'smoking status' (indicator) and could have two attributes: 'smoker' and 'non-smoker.' You would need to clearly define a smoker and a non-smoker in a

way that ensures that research participants cannot belong to both categories. For example, the attributes of 'smoking status' may need to be updated to 'current smoker' and 'current non-smoker' to make sure that non-smokers with a history of smoking can only belong to one category. The purpose is to operationalize the variables in such a way that makes it absolutely clear about what you mean by each variable you intend to measure (McLeod 2008; Sivagnanasundaram 1999).

Table 2. Variables and their measurement

Variable	Definition of variable	Scale of measurement
Age	Child's age at admission to hospital, calculated from date of birth and date and admission	Months and years
Patient has dengue fever	Dengue fever confirmed by serology (Dengue antigen positive, IgM positive or four—fold rise in IgG titre) or virology.	Yes/no/not available
Social class	Head of household's main occupation as stated by respondent in answer to a question in a structured questionnaire.	Detailed occupation, classified into social class I-V
Haemoglobin level	Haemoglobin concentration in capillary blood, measured by haemoglobinometer.	g/dL

Source: Singh et al. (2008)

Measurement scales have their basis in the type of variable that is to be measured (Singh et al. 2008). Four scales of measurement are widely used in quantitative research: nominal, ordinal, interval and ratio. **Nominal** scales measure categorical data by simply assigning numbers to names or other identifiers, which have no rank/order (e.g., male-1/female-2 or yes-1/no-2). **Ordinal** scales order a series of relationships, such as ranked categories (e.g., socio-economic status (I-1 to V-5). The interval between two ranks is not equal in ordinal scales.

Interval scales measure quantities and consist of equal units, but zero represents simply an additional point of measurement (e.g., Fahrenheit scale). The **ratio** scale is similar to the interval scale in that it also represents

quantity and comprises equal units, but it always has an absolute zero. Physical measures generally represent ratio data (e.g., height and weight). A few examples of variables, their operationalization and scales of measurement are presented in Table 2.

Researchers may group ratio data into categories for the purposes of measurement. For example, age could be measured in months and years and converted into age groups. By grouping, ratio data may be transformed into ordinal data where the difference in age between age groups cannot be quantified in a meaningful way. However, it is always desirable to collect higher level data (age in months and years) and convert them to categories, instead of measuring age as a category in the study instrument. This way you would be able to use both grouped and ungrouped data as needed (Cornell University n.d.).

6.4.1 Tips on questionnaire development

Once variables are operationalized and scales of measurement selected, the next step is to develop your study instrument(s). Students tend to use questionnaires, data extraction forms, checklists, and/or portable measuring tools (e.g., weighing scales, Snellen chart, etc.) in the research project.

Before embarking on developing a questionnaire, you should consider whether a questionnaire is the most appropriate tool to collect the required information. For instance, if you intend to assess academic performance by term test marks, a self-administered questionnaire may not be the best method to obtain this information. Instead, you could extract the marks from school registers (after obtaining permission and consent, of course). Similarly, measuring height and weight of participants instead of asking participants for their height and weight will increase the validity of your results.

Once you confirm you will use a questionnaire, the next step is to decide whether it will be self-administered or interviewer-administered. The advantages of each are listed below on Table 3.

Table 3. Comparison of self-administered and interviewer-administered questionnaires

Self-administered questionnaires	Interviewer-administered questionnaires
No interviewer bias	Respondent literacy is not necessary
Less time spent on administration	Questions and responses can be clarified by interviewer
Easier for larger numbers of people	Fewer blanks; allows probing for additional information
May permit more careful responding	Complex and open-ended questions are possible
More anonymity; may yield higher quality data on sensitive issues	Answering of questionnaire by intended person is assured
Printed visual aids may be incorporated	Participation may be increased by personal contact.

Adapted from WHO (n.d.).

It is best to use a tool that has been validated in the language of administration. You may use tools that have been validated in other settings but these may need to be adapted to our setting. You may also develop your own questionnaire. At undergraduate level, you are not expected to validate your questionnaire, but you do need to be familiar with related concepts (see p. 30).

Seeking guidance when developing your questionnaire will facilitate smoother ERC approval. You can also follow the guidelines provided in Chapter 10 of *The Medical Research Handbook* (Singh et al. 2008). You need to ensure that questions are clear, easy to understand, and appropriate to the respondent's level of education. The layout should be easy to read and the questions must be placed in logical sequence so that the respondent does not have to move back and forth unnecessarily between topics. The questionnaire should not be unnecessarily intrusive or lengthy, and should not include sections that do not help you to achieve your specific objectives. If translation is required, please translate from English to Tamil and back-translate to improve accuracy of the translation.

6.4.2 Validity and reliability

Validity refers to the tool's ability to measure what needs to be measured accurately. Important dimensions of validity include: content validity (including face validity), criterion-related validity, and construct validity. You are expected to address content and face validity.

Content validity refers to the extent to which the measurement covers all aspects of the phenomenon of interest. **Face validity**, a component of content validity, refers to the extent to which a tool appears valid at face value. Addressing the aspects outlined in Section 6.5.1 and asking experts to review your questionnaire/study tool will enable you to improve content and face validity. Please note that a comprehensive problem analysis at the initial stages of research also improves content validity.

Criterion-related validity refers to the extent to which the measurement obtained from the tool correlates with proven measures. This is usually measured against an external criterion. For instance, self-reported data on sanitary facilities obtained from a questionnaire can be validated by reviewing census data. You are not expected to assess criterion-related validity.

Construct validity addresses the extent to which the assumptions of the tool conform to theoretical constructs. For instance, if a questionnaire is developed to measure knowledge on dental caries and it is hypothesized that high knowledge scores are associated with fewer dental caries, the questionnaire would demonstrate construct validity if there is a marked difference in the number of dental caries in those who obtain high and low knowledge scores. You are not expected to assess construct validity.

Reliability refers to the tool's precision or its ability to obtain nearly the same value each time the variable is measured. Precision of an instrument is affected by measurement variability (due to the tool itself), observer variability (due to researcher), subject variability (variability in study subjects), or variability of the environment (Hulley et al. 2007). You are not expected to assess the reliability of your tool.

6.4.3 Measurement error

Five common types of errors influence the research process. These include: population specification (when the researcher selects an inappropriate population to address the research objective), sampling error (when the researcher uses a probability sampling method and the selected sample does not represent the population concerned), selection error (when there is selection bias in non-probability sampling), non-responsive error (when a large proportion of the sample does not respond), and measurement error (when the error emerges from the research process itself).

Measurement error is related to the concepts of validity and reliability, and may be classified as being either random or systematic. *Systematic error* consistently occurs in one direction, and is generally attributed to inaccuracy of the study instrument. It is therefore related to the validity of the instrument.

Random error is due to chance and occurs because of unknown and unpredictable changes in the experiment. Random error can be minimized by repeating measurements (or increasing sample size). Random error may result in Type I error (rejection of null hypothesis when it is true – 'false positive') or Type II error (failure to reject null hypothesis when it is not true – 'false negative'), and is related to the reliability of the study instrument (Hulley et al. 2007).

Suggested reading:

Cornell University (n.d.). Introduction to Measurement and Statistics http://lsc.cornell.edu/wp-content/uploads/2016/01/Intro-to-measurement-and-statistics.pdf.

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. 2007. *Designing Clinical Research*, 3rd Ed. Philadelphia, PA: Lippincott Williams & Wilkins. (Available at FOM Library)

Kothari, C.R. (2004). *Research Methodology: Methods & Techniques*. New Delhi, India: New Age International Publisher http://www.modares.ac.ir/uploads/Agr.Oth.Lib.17.pdf.

Kumar, R. (2011). *Research Methodology: A Step-by-Step Guide for Beginners*. London, UK, Thousand Oaks, CA, New Delhi, India and Singapore: SAGE http://www.sociology.kpi.ua/wp-content/uploads/2014/06/Ranjit_Kumar-Research_Methodology_A_Step-by-Step_G.pdf.

McLeod, S. (2008). Research Variables. https://www.simplypsychology.org/variables.html

Petrie, A. & Sabin, C. (2009). Medical statistics at a glance. https://leseprobe.buch.de/images-adb/28/42/28428804-b42f-417d-b28e-e407f7fd73ff.pdf

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

Web Centre for Social Research Methods (2006a). Variables. https://www.socialresearchmethods.net/kb/variable.php

WHO (n.d.) Questionnaire design http://www.who.int/foodsafety/publications/foodborne_disease/Annex_4.pdf

6.5 Planning for data collection

The selection of data collection methods depends on the study objectives, how others have answered similar questions in previous work, opinions of

experts, and feasibility concerns. Students generally use self- or interviewer-administered questionnaires and/or basic clinical measures to collect data. However, you are encouraged to explore other methods if you wish. By the time you submit your draft proposal, you should have planned the data collection process thoroughly.

6.5.1 Pilot study

You are expected to carry out a **pilot study** - a small scale preliminary trial on a sample recruited from a different study population (not a **pre-test**, which is a trial of a data collection technique)—before data collection proper. The purpose is to implement your research protocol to identify problems with the study design, and make amendments accordingly. Please note that you must collect data and analyze them as part of your pilot study. The pilot study is extremely important because it enables you to make adjustments to both your study design and study instruments.

6.6 Planning data analysis

Data analysis should be planned at the proposal stage. Dummy tables are useful for planning your analysis. Dummy tables do not contain data. They help the researcher plan the analysis based on the specific objectives of the study. In creating dummy tables, you should address the following: Which descriptive statistical methods will be used for analysis of which variables? And how will the descriptive data be presented? Which inferential statistical methods will be used for analysis of which variables? And how will they be presented? It will be helpful to go back to the literature for ideas.

Please note that the dummy tables should address both descriptive and analytical components of the study, and may include tables/charts/graphs, cross tabulations for associations, and tests to assess statistical significance. Note that you are expected to formulate your data analysis plan at the proposal stage, and include it in your research proposal.

6.6.1 Descriptive statistics

Descriptive statistics allow the researcher to summarize large amounts of data and present them in ways that are easily accessible. They describe the

sample under study and are specific to the sample – they do not aim to generalize the results to the theoretical population. Widely used descriptive statistics include measures of frequency (e.g., percentages, proportions), measures of central tendency (e.g., mean, median, and mode), measures of dispersion (e.g., range, variance, and standard deviation), and measures of position (interquartile range, percentiles) (Campus Labs 2018). You are expected to present your descriptive data in clear and accessible ways. Tables and diagrams (e.g., bar charts, pie charts, histograms, box plots, etc.) may be helpful for this purpose (Duke University 2017; Sivagnanasundaram 1999).

6.6.2 Inferential statistics

Inferential statistics are used when you want to move beyond simple description and draw conclusions based on your data. They help you to make comparisons (by different groups and even across time) or predictions based on data that has been collected. These predictions (inferences) are about the theoretical population.

Inferential statistical tests may be broadly divided into parametric and non-parametric tests. Parametric tests assume a normally distributed or symmetric outcome variable, while non-parametric tests do not make this assumption (Marshall n.d.). Various algorithms have been developed to guide selection of statistical tests (Gonzalez-Chica et al. 2015; Vassarstats n.d).

Students tend to use the Chi-Square test and t-test for inferential statistical analysis. However, you are encouraged to use other tests when necessary. The **Chi-Square test of independence** is used to assess whether the association between two categorical variables is significant. It may be used regardless of the number of categories of the outcome or the exposure variables. Its assumptions include: 1) the two variables are measured by an *ordinal* or *nominal* scale (i.e., categorical data); 2) the data in the cells are frequencies, or counts of cases (not percentages); 3) the levels (or categories) of the variables are mutually exclusive – that is, a particular subject fits into only one level of each of the variables; 4) no expected value

in the contingency table is equal to 0; and 5) the expected frequencies in the contingency table may not be lower than five in more than 20% of cases (none of the expected values for dichotomous exposure and outcome measurements may be <5). If frequencies are lower than this, the Chi-square test with Yates' continuity correction (when the total sample size is greater than 20) or Fisher's exact test may be used (Gonzalez-Chica et al. 2015; Laerd Statistics 2013; Marshall n.d.).

The **independent t-test** compares the means of two unrelated groups on the same continuous, dependent variable. It assumes, 1) the dependent/outcome variable is measured on a continuous scale (i.e., interval or ratio level); 2) the independent variable consists of two categorical, independent groups; 3) there is independence of observations, which means there is no relationship between the observations in each group or between the groups – this means there must be different participants in each group with no participant being in more than one group (if two different values on the same subjects need to be compared, the paired-samples t-test may be required); 4) there are no significant outliers; 5) the dependent/outcome variable is approximately normally distributed for each exposure group (this may be tested using the Shapiro-Wilk test of normality on SPSS); and 6) there is homogeneity of variances (this may be tested using Levene's test for homogeneity of variances on SPSS) (Laerd Statistics 2013).

The **Pearson correlation** is a measure of the strength and direction of association that exists between two continuous variables. It assumes, 1) the two variables are continuous; 2) there is a linear relationship between the two variables (may be checked with a scatter plot on SPSS); 3) there are no outliers; and 4) that the variables are normally distributed (Laerd Statistics 2013). The test is used when both variables are measured by the researcher.

Stating you will use SPSS for analysis is inadequate as a data analysis plan. You need to specify which descriptive and inferential statistical methods you will use in your analysis.

Suggested reading:

Campus Labs (2018). Types of descriptive statistics. https://baselinesupport.campuslabs.com/hc/en-us/articles/204305665-Types-of-Descriptive-Statistics

Duke University (2017). Descriptive statistics. http://sites.nicholas.duke.edu/statsreview/descriptive/

Gonzalez-Chica, D. A., Bastos, J. L., Duquia, R. P., Bonamigo, R. R., & Martínez-Mesa, J. (2015). Test of association: which one is the most appropriate for my study? *Anais brasileiros de dermatologia*, 90(4), 523-528 http://www.scielo.br/pdf/abd/v90n4/0365-0596-abd-90-04-0523.pdf.

Gregg, M.B. (2002). Field Epidemiology, 2nd Ed. Chapter 8 http://www.ciphi.ca/hamilton/Content/documents/fieldepi.pdf.

Kothari, C.R. (2004). *Research Methodology: Methods & Techniques*. New Delhi, India: New Age International Publisher http://www.modares.ac.ir/uploads/Agr.Oth.Lib.17.pdf.

Marshall, E. (n.d.). The statistic's tutor's quick guide to commonly used statistical tests. http://www.statstutor.ac.uk/resources/uploaded/tutorsquickguidetostatistics.pdf.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

Vassarstats (n.d.). Parametric and non-parametric http://vassarstats.net/textbook/parametric.html.

6.7 Ethical considerations

You are expected to think through the ethical implications of your research. The purpose is to demonstrate that you have considered the rights of research participants. Please pay attention to the following areas: informed consent, voluntary participation (including withdrawal without penalty), privacy and confidentiality, assessment of risk and benefits of participation, medical/psychological support for participants (if necessary), complaint procedures, and plans for data dissemination.

Suggested reading:

Mandal, J., Acharya, S., & Parija, S. C. (2011). Ethics in human research. *Tropical parasitology*, 1(1), 2–3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3593469/

7. Writing up the research proposal

The research proposal should consist of 10-12 pages (excluding annexures), be printed on both sides with 1.5 spacing and 3 cm margins, and Times Roman font size 12. It should contain the following sections:

- *Cover page*

This should include the title of project, your names and registration numbers, names and signatures of your supervisors, and the place and date of submission. Please insert the following statement at the bottom of the page:

"This research proposal is submitted as a requirement in Community and Family Medicine at the Second Examination for Medical Degrees [YEAR]*"*

- *Summary*

This section should provide a brief overview of your project. It should introduce the research problem, state the general and specific objectives, and briefly describe how you intend to achieve your objectives (methods). It should end with a description of the potential use of carrying out the study, i.e. how would your study findings be useful to practitioners, policymakers and/or researchers.

- *Introduction*

The introduction should consist of a background and justification for research. The background is essentially where you introduce the research problem and other related concepts the reviewer needs to be familiar with to understand the contents of your proposal. The justification should explain the importance of carrying out your research, particularly in relation to your research setting. The introduction should end with a clear statement of the general and specific objectives, presented in bullet points or numbers. A good background and justification flow naturally to the objectives.

- *Literature review*

At the proposal writing stage, you would have performed a comprehensive survey of the literature. The literature review presents the current state of knowledge on your topic obtained through the literature survey. A common mistake is to confine the literature review to journal publications accessible via Google or Google Scholar. You are expected to widen the literature search to other online databases (e.g., PubMed) and local literature that may not show up on these data bases. Sri Lanka Journals Online provides open access to journals published in Sri Lanka. Visiting the library or contacting local authorities or reviewing past undergraduate research reports will give you a better understanding of the local context in relation to your research problem, and will make your literature review much stronger.

The purpose of a literature review is to critically engage with the literature and present it in a way that sets up the study that is to be undertaken. In other words, your literature review should convey the study designs/methods used in previous work in sufficient detail to justify the methods you will use in your own study. However, literature reviews compiled by students are often a series of abstracts of studies that use similar research methods. Please note **it is not acceptable to 'cut and paste' from online sources** (see Section 7.1).

By the time you write up your research proposal, you would have analyzed the research problem and developed your research objectives. Instead of simply listing the findings of similar studies, you should organize the review into themes according to the research objectives. For example, the literature review for a study on the prevalence and severity of dysmenorrhoea and associated menstrual characteristics among A/L students in the Jaffna Educational Division could be organized under the following subsections:

- Prevalence and severity of dysmenorrhoea (presented in a manner that justifies selected age group and highlights gaps in the literature on dysmenorrhoea in Sri Lanka/Northern Province/Jaffna);

- Factors associated with the prevalence and severity of dysmenorrhoea (presented in a manner that justifies focus on menstrual characteristics over other factors in the study);
- Approaches to measuring dysmenorrhea and its severity (presented in a manner that suggests that you are familiar with standard approaches/tools used to assess the prevalence and severity of dysmenorrhea, and justifies your selection of study instruments)

Organizing the literature review in this way allows the reader to gain an understanding of the status of knowledge on the topic as relevant to the research objectives. Writing up a strong literature review will not only make a stronger research proposal but will also save time on the long term as you would not need to work on expanding and improving this section in the final research report.

➢ *Methods*

Your methods should be finalized by the time you submit your proposal. This section should clearly lay out how you intend to achieve your objectives, and should address the sections listed below.

i) Study design

Specify the study design you will adopt (see Figures 5 and 6). For the purposes of the undergraduate research project, most students use a descriptive cross-sectional study design.

ii) Study setting

Specify the location(s) in which you will carry out research. Provide some details of the research setting. For example, if your research setting is the medical clinics of Teaching Hospital Jaffna, mention the number of clinics and when they function.

iii) Study population

State your study population clearly with inclusion and exclusion criteria as needed.

Estimate the study population to which you will have access to demonstrate that you will be able to achieve your estimated sample size. Consider the time you have for data collection and the feasibility of accessing participants. For example, if you carry out a clinic-based study, you can estimate the study sample by multiplying the number of clinics that will be conducted during the data collection period by the average number of patients attending each clinic (this information can be obtained from the clinic).

iv) Sampling

i) Estimated sample size

Estimate the sample size. You are not expected to be familiar with the various sample size equations that are used in quantitative research. DCFM expects you to estimate sample size using the sample size equation for a descriptive cross-sectional study (pp. 20-21). If you are using a different study design, seek further guidance on sample size estimation from your supervisors.

ii) Sampling technique

Specify your sampling method. If the study population is less than your estimated sample size, you should state that you will include all prospective participants (e.g., patients who attend the relevant clinics) during the data collection period. If the numbers exceed your estimated sample size, you may adopt a suitable probability sampling method to select your sample considering the availability of a sampling frame, and the geographic dispersion of your study population.

v) Study instrument(s)

Provide details about your study instrument(s). If you intend to use a questionnaire, list the sections the questionnaire will contain. PLEASE NOTE: the questionnaire/data extraction form/check list should be annexed to

your proposal. Remember to reference the relevant annex in brackets in this section.

vi) Data collection plan (including pilot study)

Provide details about your intended pilot study and your plans for data collection. Include when, where and by whom data will be collected. If you plan to recruit data collectors, please describe their training.

vii) Data analysis

Provide details about your data analysis plan, including the descriptive and inferential statistical methods you will use. Please note that stating you will use SPSS to analyze your data is insufficient.

viii) Ethical considerations and permissions

Please follow the guidelines set by the Ethics Review Committee, Faculty of Medicine, University of Jaffna, and use the templates provided by the ERC when developing the informed consent forms, available here. Include consent forms for illiterate participants, if you expect such persons may participate in your study. If persons of <18 years of age are expected to participate, include informed consent forms for parents and assent forms for participants. Please annex all forms to your proposal.

Outline plans for obtaining permission from the relevant authorities, as applicable.

- *Timeline*

You are expected to include a Gantt chart to indicate the intended timeline for data collection, analysis, and write up of the final research report. Please consider the timeline provided by the department.

- *Budget*

Anticipated expenses and sources of funding (if applicable) should be included.

- *Annexures*

Questionnaires, data extraction forms, checklists, other tools (if applicable), information sheets, informed consent forms, and assent forms (if applicable) must be appended along with translations.

Suggested reading:

Al-Riyami, A. (2008). How to prepare a Research Proposal. *Oman Medical Journal*, *23*(2), 66 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3282423/.

Biribo, S. (2016). Guidelines for the development of a health research proposal http://www.fnu.ac.fj/research/images/forms/updated/Guidelines_for_the_Development_of_a_Health_Research_Proposal.pdf

University of Nevada, Reno (n.d.). Undergraduate research: Prepare your research proposal https://www.unr.edu/undergradresearch/proposal.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

7.1 A note on plagiarism

The work you submit must be your own. **Please do not 'cut and paste' from online resources. Any proposal containing 'cut and pasted' material will be returned, and the proposal will need to be resubmitted.** Cutting and pasting is a form of plagiarism. As a medical student you are expected to understand concerns around plagiarism.

Suggested reading:

University of Oxford (2018). Plagiarism. https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism?wssl=1.

7.2 Referencing

You are expected to use Harvard referencing style consistently throughout your research proposal. Pay attention to both in-text citations and your reference list.

Suggested reading:

Imperial College (2012). Citing and referencing: Harvard Style https://www.otago.ac.nz/library/pdf/Harvard_referencing.pdf

7.3 Reference Management Software

Reference management software helps researchers store, track, and cite reference information about their research materials, including books, journal articles, archives, interviews, maps, works of art, and many other types of materials. There are many types of software available, some of them, such as Mendeley or Zotero, are available at no cost to the researcher (Imperial College 2017). DCFM requires that you gain experience using Mendeley in your research activities. The software may be downloaded free-of-charge here. Guidance for working with Mendeley is available here.

Suggested reading:

Bodleian Library (2015). Managing your references: Mendeley. https://libguides.bodleian.ox.ac.uk/reference-management/mendeley

Imperial College London (2017). Reference management software comparison. https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/library/public/Reference-management-software-comparison.pdf

References

Gregg, M.B. (2002). *Field Epidemiology*, 2nd Ed. Chapter 8 http://www.ciphi.ca/hamilton/Content/documents/fieldepi.pdf.

Grimes, D. A., & Schulz, K. F. (2002). An overview of clinical research: the lay of the land. *The Lancet*, 359(9300), 57-61 http://www.geocities.ws/mim_ebm/LancetEpi-01.pdf.

Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. 2007. *Designing Clinical Research*, 3rd Ed. Philadelphia, PA: Lippincott Williams & Wilkins. (Available at FOM Library)

Kothari, C.R. (2004). *Research Methodology: Methods & Techniques*. New Delhi, India: New Age International Publisher http://www.modares.ac.ir/uploads/Agr.Oth.Lib.17.pdf.

Kumar, R. (2011). *Research Methodology: A Step-by-Step Guide for Beginners*. London, UK, Thousand Oaks, CA, New Delhi, India and Singapore: SAGE http://www.sociology.kpi.ua/wp-content/uploads/2014/06/Ranjit_Kumar-Research_Methodology_A_Step-by-Step_G.pdf.

Marshall, E. (n.d.). The statistic's tutor's quick guide to commonly used statistical tests. http://www.statstutor.ac.uk/resources/uploaded/tutorsquickguidetostatistics.pdf.

Petrie, A. & Sabin, C. (2009). *Medical Statistics at a Glance*. https://leseprobe.buch.de/images-adb/28/42/28428804-b42f-417d-b28e-e407f7fd73ff.pdf

Singh, A., Bakar, A.A., & Sararaks, S. 2008. *The Medical Research Handbook: Planning a Research Project*. Kuala Lampur, Malaysia: Clinical Research Centre Perak and Institute for Health Systems Research http://www.crc.gov.my/wp-content/uploads/2016/07/01_the_medical_research_handbook.pdf.

Sivagnanasundaram, C. 2003. *Learning Research: A Guide to Medical Students, Junior Doctors, and related Professionals,* 2nd Ed. Jaffna, Sri Lanka: C. Sivagnanasundaram. (Available at FOM Library)

WHO (n.d.) Questionnaire design http://www.who.int/foodsafety/publications/foodborne_disease/Annex_4.pdf