



SCHOOL OF MEDICINE

PREPARING A RESEARCH PROPOSAL OR PROTOCOL

September 2025

CONTENTS

Introduction	1
The difference between a research proposal and a research protocol	1
Learning through preparing a proposal and protocol	3
Purpose of this document	4
Learning approach	4
Preparing a proposal	9
Writing the proposal	9
Research problem, research question, aims and objectives	10
Research problem and research question	12
Hypothesis	15
Aims	17
The rationale is not the aim	17
Learning about research aims	17
Reflection and application	18
Objectives	19
Learning about objectives	20
How many aims and objectives should there be?	20
Primary and secondary aims and objectives	21
Further learning	22
Reflection and application	22
Taking stock of research question, aims and objectives	23
The next step	23
Back to the beginning	24
Title Page	25
Title	25
Supervisors	25
Investigators	26
Date	26
Background and rationale	27
Learning about the background and literature review	27
The research problem	27
Literature review	28
Learning about scientific writing	28
Summary of existing research related to the topic	28
Rationale	28
Referencing	28

Research methodology	31
Proposal versus protocol	31
Study design	31
Study site	33
Study population	33
Inclusion and exclusion criteria	35
Sampling strategy	37
Metrics	39
Statistical methods	41
Sample size	43
Timeline and study duration	45
Methodology	46
Ethical considerations	46
Data analysis and Statistical methods	47
Scientific rigour	48
Reflection, consideration and editing	50
Conclusion	51

INTRODUCTION

THE DIFFERENCE BETWEEN A RESEARCH PROPOSAL AND A RESEARCH PROTOCOL

These are two related documents. Although often developed together, each serves a different purpose. The

Research proposal

The **research proposal** is an initial document that outlines the proposed research. It explains the importance of the study, its objectives, and the general methods that will be employed. Written prior to the commencement of the study, the proposal is designed to demonstrate that the study is scientifically valid, relevant, feasible, and ethically sound. It is the first formal step in a PhD. A well-structured, well-written proposal provides about 40% of the content of the final protocol. Once the proposal has been approved, you will develop your proposal into a protocol by strengthening the literature review and providing finer detail on participants, methods, logistics, and ethics compliance sections. The two documents are compared in the following table.

Section	Proposal	Protocol
Title	Short, working title	Final, precise title
Background, literature review	Concise, focused, only the most relevant studies	Expanded, comprehensive, with full referencing
Problem statement	Present	Present, updated
Rationale and significance	Present, brief	Present, possibly elaborated
Aims and objectives	Present, clearly listed	Present, updated
Study design	Broad description (e.g. cross-sectional, cohort)	Full detail: exact design, setting.
Study population	General description	Detailed eligibility criteria (inclusion/exclusion)
Sampling and recruitment	Outlined broadly	Detailed recruitment strategy, sampling frame, procedures
Variables and measurements	Outlined broadly	Each variable defined precisely; measurement instruments described
Samples size	Tentative estimate	Full calculation with justification
Data collection	Broad outline	Step-by-step process, instruments appended
Statistical analysis	Broad outline of proposed statistical methods	Detailed plan: descriptive stats, tests, handling of confounders/missing data, software
Data management	Brief mention	Plan for storage, coding, quality control, confidentiality
Ethics	Brief mention	Full detail: informed consent, risk/benefit, safety monitoring, compliance with regulations
Timeline	Preliminary outline	Full timeline

Budget	Brief mention, including potential or committed funding sources	Itemised and costed budget, funding sources
References	Key references only	Full, formatted list
Appendices	Absent	Questionnaires, CRFs, consent forms, letters of permission/support

Purpose of the proposal

1. *Clarifying and consolidating the researcher's understanding of their project.*
2. *Presenting the project for independent review before it is finalised as a protocol.*

All good research plans are independently reviewed. Having the benefit of an independent review almost always leads to identifying weaknesses in the protocol, or opportunities to improve the research's efficiency, quality and value. By studying your proposal, reviewers can evaluate the study's scientific soundness, validity and feasibility and suggest improvements before you commit yourself to the hard work of writing every section in the detail required for a protocol.

To develop an acceptable project, the researcher must clearly understand two things: the research question they intend to answer, and how the research uses the *scientific method* to answer it. This requires a firm understanding of the scientific method. Your completed proposal should demonstrate that you are familiar with the scientific method and that you have a deep understanding of the purpose of your project and what you will do to bring it to completion. The alternative is GIGO (*garbage in, garbage out*). Remember the adage: *if you don't know where you are going, you may end up somewhere else*. A poorly designed research project will produce poor results: poor results result in poor (or unpublished) papers, and poor (or possibly failed) dissertations. More importantly, master's and PhD research require a lot of time and commitment, and it is, in fact, something of a tragedy if all this time and commitment result in work which is at best mediocre and perhaps worse.

Proposal defence

The proposal is presented to a panel of senior academics, whose role is to assist the student in designing the best possible project by commenting on it and suggesting areas for improvement. This is known as a *proposal defence*. (*Defence* is an old and traditional term for an opportunity for an individual to explain and justify their work publicly. It does not imply that the student is under attack!) The panel may well suggest modifications which will result in a better project. It is even possible that they will feel that the project is unlikely to succeed and that it would be better to change the project substantially. This may be disappointing, but it is ultimately in your best interests. It is better to change course at the proposal stage than only when you have completed an entire protocol and had a project rejected (or, worse, actually commenced research only to find that your project is going nowhere).

This document is intended to assist the student in writing a research proposal to a high standard. Perhaps more importantly, it is intended to guide the student's learning as they do so, such that they develop a sound understanding of the scientific method and how research is planned and conducted.

Research protocol

The *proposal* introduces the study concept to secure approval, while the *protocol* is the definitive guide for the research.

Following approval of the proposal, the **research protocol** is developed as a comprehensive and detailed plan for conducting the study. The protocol follows on from the proposal. It includes the material from the proposal, possibly updated.

Purpose

1. Guiding the research

The protocol serves as a roadmap, ensuring the study is conducted systematically and consistently, reducing the risk of bias and errors.

2. Managing resources

The protocol includes precise descriptions of each part of the methodology, covering data collection, participant recruitment, and data analysis. It also addresses relevant regulatory standards for clinical or regulated studies to ensure that the study is reproducible and conducted safely. It helps in planning and managing the resources required for the study, including time, budget, study participants and helpers such as data capturers and laboratory staff.

3. Ensuring compliance with norms and standards

Ethics review of the protocol determines that the research complies with ethical standards, regulatory requirements, and institutional guidelines, particularly in studies involving human or animal subjects. Remember that poor-quality research *which does not produce a beneficial outcome* is **unethical** if human or animal subjects are inconvenienced or placed at risk during the research.

4. Reproducibility

By documenting all aspects of the study design, including the objectives, hypotheses, methodologies, and analysis, the protocol allows other investigators to reproduce the research, which is an important check on its credibility.

LEARNING THROUGH PREPARING A PROPOSAL AND PROTOCOL

Preparing a research proposal is a critical part of the research process, and it typically accounts for a substantial portion of the overall learning experience of master's and doctoral degrees. Together, your preparation of proposal and protocol is estimated to contribute **about 30% to 50%** of all the learning gained from a master's degree or PhD! Any student who hurries through the preparation of their proposal is depriving themselves of the most important opportunity for learning. It is a common mistake to think that preparing the proposal and protocol are tasks that must be done as quickly as possible so that the student can start with "the real research", i.e., data collection. Thinking this way has serious consequences:

- The student will not have assimilated all the knowledge, insights and understanding of the scientific method and of research method that come from engaging deeply with their proposal. They have missed out on the 30-50% of learning described above.
- Lacking this knowledge, they will produce a low-quality research protocol. Poor protocols result in poor research: *garbage in, garbage out*.

Key learning derived from the conscientious preparation of the research proposal and protocol

If you use the process of preparing your proposal and, subsequently, your protocol as an opportunity for deep learning, you will achieve the following:

- You will understand the scientific method and how scientific reasoning is employed to gain new knowledge
- You will learn to formulate a scientifically valid and defensible research question, hypothesis, aims and objectives
- You will learn how to access the scientific literature, recognise the current state of knowledge, identify gaps, and justify a research question.
- You will understand the wide spectrum of study designs, and how to select the appropriate study design (e.g., qualitative, quantitative, mixed methods) for a given research question.
- You will have confidence in selecting the right data collection, sampling, and analysis methods, which involves understanding the strengths and limitations of various approaches.
- You will be familiar with ethical issues, including obtaining informed consent, maintaining confidentiality, and securing ethical approval.
- You will understand the importance of careful planning, budgeting, resource management and time management in research.
- You will develop ability in critical thinking.
- You will practise your scientific writing skills by drafting a clear, concise, and well-organised document.

PURPOSE OF THIS DOCUMENT

This document has two purposes. Firstly, as it guides you through preparing your proposal and protocol, it will provide you with multiple opportunities for self-directed learning about the topics necessary to understand a research project. If you engage with this conscientiously, you will find that your understanding of research, and of your own project specifically, is hugely improved. This leads to the second purpose. It guides you in the preparation of your own proposal and protocol to the point where they are ready for scientific and ethics review. The first and second together mean that your eventual protocol, and therefore your project, will be of a far higher standard than if you had just attempted to work through the protocol template unsupported. The amount of information required for a proposal or protocol respectively is suggested throughout the document in a box under the section heading.

LEARNING APPROACH

The approach we follow is as follows:

1. Each section of the proposal and protocol is presented in the order that makes the most sense in terms of its preparation.
2. There is a brief introduction to that section of the document.
3. You are invited to engage deeply with ChatGPT or a similar AI (artificial intelligence) engine on the concepts and facts which underlie that section.

4. Having done so, you will find yourself greatly assisted in completing that section, as you will now have the background necessary to do so.
5. You are invited to reflect on what you have written, and to decide whether it is genuinely suitable for your purposes.

Learning with ChatGPT

ChatGPT is an advanced AI language model developed by OpenAI, designed to generate human-like text based on the input it receives. It can engage in conversations, answer questions, provide information, and assist with tasks across various topics. Trained on a vast amount of text data, ChatGPT is used in various applications, from customer support to educational assistance. Its responses are generated based on patterns in language and are continually improved as it learns from user interactions. ChatGPT is free and easily accessible. You ask it any question you wish using **everyday, natural language**, and it puts together a concise summary of what it finds. It is extraordinarily effective at this. A commonly used alternative AI engine is *Claude* (<https://claude.ai/>). Try both and see which you prefer.

Use ChatGPT freely to learn all the underlying concepts necessary for a good research project and protocol. This will also help you gain that 30-50% of learning which should accompany the development of the protocol, meaning that you are on the way to becoming a competent scientist.

Setting up ChatGPT

Starting with ChatGPT is easy. You do not have to download or install any software. You interact directly with the ChatGPT website (so you must be online). Visit the site <https://chatgpt.com/>. If you are not already registered, then register now. You only need to register once. Signing in using your Google credentials is very convenient. Bookmark the site so you can return to it easily. Follow the same process if you wish to set up Claude.

ChatGPT's natural language processing

Interacting with ChatGPT is like having a conversation with a *person*, not a *computer*. Here are some of the most practical and impressive aspects.

1. ChatGPT responds to everyday conversational instructions, such as:

Tell me about critical thinking.

I don't understand critical thinking.

(You do not even have to phrase this as a question. It knows that you are asking it to help you.)

3. ChatGPT can carry on a conversation. It remembers what came before.

You do not have to remind it what the conversation is about or repeat yourself. It will pick up where you left off. This means you can ask it to rephrase, clarify, give examples, correct mistakes (yes, it makes mistakes), etc., without repeating your question. This is entirely unlike doing a Google search. This makes it easy to repeat something until it is perfect. All you have to do is enter phrases such as: *Repeat that in simpler language* or *Do it again using medical examples* or *Can we do it again without the formulae* or *Please repeat with more emphasis on practical applications* or even *I don't understand this*. This last example will prompt ChatGPT to repeat the output but rephrase it more

simply. Note that it does not matter how you phrase the instruction. AI intelligence understands the intentions behind the request and responds appropriately. If you make a mistake, you can correct yourself and ChatGPT will generate output in response to what you wanted.

3. It overlooks spelling mistakes and bad grammar as long as your meaning is clear.

You do not have to go back to correct typing errors when you chat with ChatGPT. It will know what you meant. This greatly speeds up the interaction if you are not a great typist. If you use dictation software, you can use your voice to enter the request in the submission box. If you make a mistake in your question, for example, by asking something about quantitative rather than qualitative research, you can redirect ChatGPT like this: *Sorry, I meant qualitative research.* You do not need to retype the whole question. ChatGPT will generate a new reply in line with what you asked.

4. It changes the context of its replies and examples to match your desired context.

This means it can keep its discussion and examples focused on whichever area is most relevant to you. You might, for example, say: *Discuss this in a clinical context,* or *Give me examples from immunological research,* or *Stick to biomedical research.* You can focus it on something as specific as your own project: *I am studying the differences in survival between patients from rural and urban areas who present to Greys Hospital with oesophageal cancer. What do I need to know to calculate my sample size?*

Why ChatGPT is such a powerful learning tool

This is best shown by contrasting ChatGPT as a learning tool with more traditional sources of learning. Say, for example, you are presented with the following information in a lecture or textbook:

The t-test is a statistical procedure used to evaluate whether the difference in means between two groups is statistically significant, based on the assumption that the variances are roughly equal. It calculates a t-statistic by dividing the difference in group means by the standard error, which is derived from the pooled variance of the two groups. This t-statistic is then compared to a critical value in the t-distribution, helping determine whether the observed mean difference is likely due to random variation or a true effect. A significant t-test result, indicated by a low p-value, implies that the variances between groups do not fully account for the observed difference in means.

If you have not previously studied statistics, it is likely that you will not actually understand all of this. Among other difficulties, you will not know the meaning of the double-underlined words in the example or appreciate the finer points of the concepts implicit in a phrase such as *random variation*. If you are an extraordinarily conscientious student, you may go away and spend hours looking up these phrases in textbooks or online resources. Students being students (and humans, humans), it is quite likely you will never get around to doing so. Consequently, you will learn only a fraction of what was offered in that lecture or textbook.

Unlike the lecture or textbook context, you can immediately query ChatGPT on *anything* you don't understand. Look at the following "conversation" (and it really is a conversation) between student and ChatGPT on the question of *variance*, one of the underlined words in the abstract:

Student: What are variances?

ChatGPT: Variance is a measure of the spread or dispersion within a set of data points. It quantifies how much individual values in a dataset differ from the mean, showing whether data points are generally close to the mean or widely scattered. Mathematically, variance is calculated by finding the average of the squared differences between each data point and the mean. A high variance means the data points are spread out, while a low variance indicates they are close to the mean.

Student: Why are the differences squared?

ChatGPT: The differences are squared in the calculation of variance to achieve two main purposes. First, squaring each difference ensures that all deviations from the mean are treated as positive values; otherwise, negative and positive deviations would cancel each other out, giving a misleading result close to zero. Second, squaring emphasizes larger differences more than smaller ones, making variance sensitive to data points that are far from the mean. This helps capture the overall spread more effectively, especially in datasets with outliers or widely scattered points.

Student: Isn't the standard deviation also used to measure spread?

ChatGPT: Variance is simply the standard deviation squared. When we calculate variance, we square the differences between each data point and the mean, then average these squared differences. Standard deviation is just the square root of this result, bringing the measure back to the original units of the data. So, while variance and standard deviation both measure the spread of data, standard deviation is the square root of variance and gives a more intuitive, unit-consistent measure of how spread out the data points are.

Student: What has the square root got to do with it?

ChatGPT: Taking the square root in the standard deviation calculation reverses the squaring step used in variance, bringing the measure back to the original units of the data. Since variance involves squaring the differences between each data point and the mean, it results in squared units (e.g., meters squared if the data is in meters), which can be harder to interpret directly. By taking the square root of the variance, standard deviation returns this measure of spread to the same units as the data, making it more interpretable. This adjustment makes it easier to understand how data points deviate on average from the mean.

Student: Am I right if I say that it doesn't really matter if I describe the spread of my results in terms of variances or standard deviations?

ChatGPT: You are partially right, but there is a subtle difference in how variance and standard deviation are used to describe the spread of results. While both measure the spread of data, standard deviation is generally preferred because it is in the same units as the original data, making it easier to interpret. For example, if your data is in meters, the standard deviation will also be in meters, so you can easily understand how much the values deviate from the mean in real-world terms. On the other hand, variance is in squared units (e.g., meters squared), which makes it less intuitive to interpret directly. Variance is more often used in statistical formulas, like in hypothesis testing and regression analysis, but standard deviation is typically used when you want to communicate the spread in a more understandable way. So, while they are mathematically related, standard deviation is usually the more practical choice when describing the spread of your results to others.

When the student met the word *variance* in the lecture or textbook, it meant almost nothing to them. Can you imagine how much more deeply the student understands the term *variance* after this short conversation with ChatGPT? By the time they have worked through all the underlined words in the initial description of the t-test, they will truly understand what it is saying. It now contains *meaning*, and not just *words*. In other words, they are starting to *understand* a t-test in a broader, deeper and applicable context. When they come to discuss statistical methods and sample size in the protocol, they will know what they are doing. ChatGPT performs the role of a personal tutor, but one who will never get bored, impatient or sarcastic, will not think you are stupid, and will not say, "Sorry, I have to leave now"!

One of ChatGPT's most useful abilities is the ability to provide examples. Examples immediately give context to what you are learning, make the principles easier to understand, and get you thinking about your own work and how you might apply what you are learning.

Student: Give me an example of how variance is used in a study

ChatGPT: In a study investigating the effectiveness of two different treatments for high blood pressure, variance could be used to assess the variability in blood pressure reductions across participants within each treatment group. Researchers enroll two groups of patients, one receiving Treatment A and the other receiving Treatment B. After a set period, blood pressure readings are taken for all participants in both groups. The researchers want to understand not only the average blood pressure reduction in each

group but also how much variability exists in the responses to the treatments. If the variance in Treatment A is small, it suggests that most patients in that group had similar blood pressure reductions. If the variance in Treatment B is large, it indicates that there were significant differences in how patients responded to the treatment, with some patients showing large reductions, while others saw little or no improvement. By comparing the variance in each group, the researchers can understand the consistency of the treatment effects. A higher variance might suggest that Treatment B is less predictable or that it works better for some patients and worse for others, while a lower variance for Treatment A would indicate more consistent results across the group.

Student: Give me examples in the context of psychiatry research...

Suggestions to promote deep learning

1. Do not restrict yourself to straightforward questions.

Hold a **conversation** on the topic you are interested in, just as you might with a real person. Discuss different aspects, and approach the topic from different perspectives. You saw examples of this above. The questions ranged much more widely than just the obvious *What are variances?* They included *Why are the differences squared?* and *Isn't the standard deviation also used to measure spread?* And *What has the square root got to do with it?* and *Am I right if I say that it doesn't really matter if I describe the spread of my results in terms of variances or standard deviations?* Then ask for *good* and *bad* examples (for example, *Give me examples of a good research question*, followed by *And a bad research question?*)

2. Do not pass over anything you do not understand in ChatGPT's responses.

In other words, do not treat words and concepts in ChatGPT's answers that you do not understand by overlooking or glossing over them, as we tend to do with lectures and textbooks! Ask for an explanation or examples of every phrase, word or term you do not understand in ChatGPT's responses, just as you would in a human conversation. Do this with anything you do not fully understand in the second round of responses too... and the third round... and the fourth round... Follow the trail further and further until there is **nothing** in the entire conversation you do not understand. After a while, the same terms and concepts will start coming up repeatedly. Not only does this repetition reinforce learning, but it also indicates that you have mastered the concept and can move on. ChatGPT has brought about what is called *deep learning*, whereas the lecture and textbook represented *superficial* or *surface learning*: the words were heard or read; even if they are remembered (which is unlikely), they will be essentially devoid of meaning and context and can therefore not be put to good use.

3. Save your conversations. Make your own personalised textbook.

Copy all ChatGPT's responses and paste them on successive pages of a new Word document. (There is a copy icon below every ChatGPT response.) Save this document each time you add a new response. As the session continues, you will find that you are producing your very own, personalised textbook, which directly addresses all the things *you* want to know and all the things *you* did not understand, discussed in the context most relevant to *you*, and which contains *nothing* that you do not understand!

Understanding and enjoying interactions with ChatGPT will give you a good foundation for writing your protocol. In the following section, we move on to the proposal and protocol themselves.

PREPARING A PROPOSAL

WRITING THE PROPOSAL

A research proposal is a document that describes the objectives, design and organisation of a research project. It outlines the methodology you intend to use, but does not require detailed methods. Why? Because there is no point in investing time in detailed methods until you are absolutely certain of the purpose and shape of the project.

Addressing sections of the proposal intelligently

The natural approach is to write the proposal in the conventional order, which is optimised for the document's reader, not the writer. Thus, the title appears first, followed by the background to the problem and literature review, followed by details of the study planned by the researchers.

This is not the best order in which to plan your project. Your research question, aims and objectives are the central focus of your project: all the sections which come before or after this section play a supporting role. We strongly recommend that you finalise your research question, hypothesis, aims and objectives first, and then work forwards and backwards from there. (Of course, you must have completed your reading and survey of the literature before you start on the protocol, but writing these up is best left until later.)

Therefore, we go straight to the research problem, question, aims and objectives in the next section of this guide.

RESEARCH PROBLEM, RESEARCH QUESTION, AIMS AND OBJECTIVES

Proposal – described in detail
 Protocol – updated, described in detail

Start your proposal development in the middle, by defining the research problem, research question, aims and objectives. (Let us abbreviate these to RPQAO, as they are tightly interconnected.) They are central to your study, and all the rest are built around them:

- Your study *title* captures the research question, research problem and aims of the study.
- Your *literature review* provides the background necessary to understand the research problem and research question.
- Your *study rationale* provides the justification for addressing this *research problem*.
- Your *methods* describe how you plan to answer the research question, achieve the aims and meet the objectives.

It is therefore best to finalise the research problem, question, aims and objectives before you write anything else, even the title. Only after these have been finalised, and you are fairly certain they will not change, should you shift your attention to the earlier items in the proposal.

Comment

The RPQAO are **critical** to the success of your study. It is not a matter of writing down something vaguely relevant to the topic to fill this space in the protocol. The RPQAO tell you, very precisely, where you want to go, such that your study will be scientifically valid. (Remember: *If you don't know where you are going, then you may end up somewhere else.*)

Defining RPQAO is difficult

Settling on a precise research question, hypothesis (if necessary), aims and objectives is difficult. **All our students struggle with it.** Scientific reasoning is not the same as clinical reasoning, so clinicians are not comfortable with defining RPQAO, and are often not very good at it. Excellent clinical reasoning makes excellent clinicians; excellent scientific reasoning makes excellent researchers. They are not the same thing. We do not allow non-clinician researchers to treat patients in our hospitals. So why should we expect clinicians to be capable researchers without further training? Once you fully understand the concepts behind the RPQAO, you are well on your way to understanding the scientific method.

Defining your RPQAO takes time

You need to be certain that your RPQAO are clear, scientifically valid, practically achievable and appropriately phrased before proceeding to any other aspect of project planning, proposal, or protocol writing. This cannot be rushed. As you develop your RPQAO, you must discuss them with your supervisor and as many others as possible. External commenters, formal or informal, are extremely helpful: if you share your RPQAO with them, it is very likely that they will pick up weaknesses for you to correct, or that they will see opportunities in your project which could make it very much better than what you initially had in mind. There are many ways to achieve this, from asking mentors and advisers to read your proposal to presenting it at a meeting that includes both peers and seniors, where it can be discussed. Nor should you wait until you have completed it: discussion at earlier and intermediate stages allows for changes without too much disruption.

Terms used in research protocols have specific meanings and contexts.

Most of the headings you encounter in writing your protocol sound like simple English words: “problem”, “question”, “aims”, “study design”, etc. but, in the context of the scientific method, they have **precise meanings**. Beware! You **cannot** prepare your protocol relying on your sense of their everyday meaning in English. You must understand how they are used in the scientific sense before you attempt to write anything in your protocol. Each term has a meaning and a *context* in which it is understood and used. You must understand both to write an adequate proposal and perform decent research. As you learn and understand them, you will start understanding the scientific method. Effectively, you will be doing that 30-50% of the learning associated with a master’s or PhD degree, which attaches to mastering the proposal and protocol. You will also prepare yourself to produce an adequate protocol. As you saw earlier, ChatGPT is an excellent learning tool for this.

RESEARCH PROBLEM AND RESEARCH QUESTION

The research problem and research question are two sides of the same coin. The problem gives rise to the question, and the question demands an answer, which will be the answer to that problem. Logically, one defines the problem first, then composes a question that addresses it. Although you must understand the problem to write the question, it often makes sense to work in reverse when writing it down. Having defined your research question, go back and commit your research problem to words in a way which matches the two. Your research question must match the research problem perfectly. If not, one or the other needs adjusting.

Warning

Many serious problems with research projects start with a faulty research question. Reviewers and examiners home in on the research question. They know that a poorly-defined research question is a red flag: *Warning! This candidate does not know what they are doing! Bad science ahead!* They expect the research problem, the question, the aims and the objectives to be perfectly matched.

Analogy

A precise research question is analogous to a precise address, e.g. *I want to go to 18 Marlowe St, Scottsville, Pietermaritzburg*. A poorly-defined research question is analogous to *I want to go to Scottsville, or I want to go to Pietermaritzburg* (too vague to define a precise destination), or, worse still, *I want to drive in a car* (which describes a *method*, not a destination).

Students frequently make statements such as *My research question is about glomerulonephritis*. This is useless. It is analogous to saying *My destination is about Pietermaritzburg*, which is logically nonsensical and completely unhelpful in navigating to a specific address. Apart from anything else, a research question must be phrased as a genuine question, followed by a question mark, typically beginning with interrogative terms such as *what, how, why, who, and where*. *About* is not such a word.

Answering the research question

The research question must be answerable by the project. The entire purpose of your study is to answer the research question. Make sure the question is really the question that your results will answer.

Example

Research Problem: Many patients with type 2 diabetes in KwaZulu-Natal have poor glycaemic control, but the reasons for this are not well understood.

- a. **Well-matched Research Question:** What are the main patient-related factors associated with poor glycaemic control among adults with type 2 diabetes attending public-sector clinics in KwaZulu-Natal?
- b. **Poorly Matched Research Question:** What is the average fasting blood glucose level among type 2 diabetes patients attending public-sector clinics in Durban?

There are at least two problems here. Measuring blood glucose levels tells us nothing about the reasons for poor control. Secondly, Durban is not synonymous with KwaZulu-Natal. It is possible that Durban is not fully representative of the KZN population.

There is actually a third problem, too. Even the first, “well-matched” question is, in fact, not perfectly matched. The question confines itself to public-sector patients. The research problem deals with a broader problem: diabetes across the province, including affluent patients and those with medical insurance (since they are not explicitly excluded). Answering the research question will, therefore, not address the problem completely. It will only have dealt with *part* of the problem—the reasons for poor control in public-sector patients. It often takes a lot of thought, insight, and reformulation to achieve a perfect alignment between the problem and the question. It is not just a matter of keeping reviewers and examiners happy. It is ensuring that your project is scientifically valid, that you fully understand what you are doing, and that you fully understand the limits to the interpretation you will eventually place on the results (this case, case *public-sector*, not the entire KZN population, with all that that means for income, social advantage or disadvantage, ethnicity, etc, all of which are potential confounders).

Warning!

This example shows just how careful you have to be in defining your research problem and research question.

Learning about research questions

1. Have a conversation with ChatGPT about research problems and questions, how they relate to each other, their importance to a protocol and a study, and what factors contribute to a research question being good or bad, suitable or unsuitable.
2. Ask ChatGPT about the three logical approaches to formulating and answering a research question: *induction*, *deduction* and *abduction*. Understand the difference, and get ChatGPT to provide examples of research questions built on each.
3. Now discuss your proposed research question with ChatGPT. For instance, type *My research problem is* (followed by the research problem you have decided upon) into ChatGPT. Then type *Is the following a good research question?* (followed by your proposed research question). Continue the conversation with ChatGPT until you understand all the possible strengths and weaknesses of different formulations of your research question.

Remember the “intelligence” in *artificial intelligence*. ChatGPT is perfectly capable of an intelligent conversation. You’re not restricted to asking at straightforward questions. Try asking it less obvious but interesting questions such as *Why do doctors struggle to develop good research questions?* or *Why does my supervisor keep rejecting my research question?* Continue the conversation as long as you need to reach the point where you have confidence in your own ability to put down a good research problem and question and feel competent to defend them in the face of critical questioning.

Look at the following example.

Student	I want to know how common asthma is in schoolchildren in part of Durban. Does this sound like a good research question: What is the prevalence of asthma in schoolchildren in the suburb of Inanda?
ChatGPT	Yes, that's a clear and focused research question! It specifies the condition (asthma), the population (schoolchildren), and the location (Inanda). This question will guide you to determine how widespread asthma is in that specific area, which can provide valuable insights for local health initiatives or school

health programs. If needed, you could refine it further by specifying the age range or the school level (e.g., primary or secondary) for even more precision.

Student How can I improve it?

ChatGPT To improve your research question, you could make it more specific by including details such as age groups, asthma control, or whether you are focusing on diagnosed cases or symptoms. Here are a few refined options:

What is the prevalence of diagnosed asthma among primary and secondary schoolchildren in the suburb of Inanda, Durban?

Adding “diagnosed” narrows the focus to medically confirmed cases.

What is the prevalence and control level of asthma in schoolchildren aged X to Y years in the suburb of Inanda, Durban?

Including age range and control level adds depth by addressing not just prevalence but also how well asthma is managed.

What factors are associated with asthma prevalence among schoolchildren in the suburb of Inanda, Durban?

If interested, this question allows exploration of environmental or lifestyle factors linked to asthma rates in the area. These adjustments could yield more targeted insights, especially if you're looking to inform specific health policies or interventions in the community.

Can any student say that a deep interaction with ChatGPT would not improve their protocol?

Warning

Remember that this exercise *is not about getting ChatGPT to write your research question for you*. It is all about ChatGPT helping you to develop the understanding you need to write a good, scientifically valid research question. When you discuss your research question subsequently with your supervisor or anyone else, you will have a better suggestion to present to them, and you will be able to discuss it with them more confidently and knowledgeably, and even to defend it against criticism.

The University has a policy governing the use of AI in academic work. Ensure you understand. This whole document is based on the expectation that you will use AI for learning and to check and improve your work, not that you will allow it to do it for you.

Consideration and reflection

Once you thoroughly understand the principles behind good research problems and the research questions, you can decide on those for your study and write them into your proposal. If you and ChatGPT have used the time constructively, you will be amazed at how much tighter your research question is now compared to your first draft. Review your research problem and research question. Do they meet all the positive attributes and avoid all the negative attributes you discussed with ChatGPT? Do they match each other?

Remember, a *research question* requires an *answer*, which is provided by the study. If it cannot be answered, it is just a philosophical question, not a research question. Are you confident that the question is one you can realistically answer in the course of your degree?

HYPOTHESIS

Not all research questions require a hypothesis. As a general rule, *experimental* studies need a hypothesis, since you are testing out a particular prediction, e.g. *Patients receiving Drug A will live longer than those receiving Drug B.* (This is deductive reasoning.) *Observational* studies may or may not require a hypothesis. A purely observational study (inductive reasoning) does not. However, if, in studying a particular observational dataset, e.g. a retrospective study of patient records, you have a clear intention, not just to see “what’s going on”, but to prove or confirm a particular suspicion or expectation, then a hypothesis is very appropriate. Here is an example of two observational studies. Both require a retrospective analysis of patient records.

No hypothesis:

Which factors best predict mortality from ruptured aortic aneurysm? (There is no prediction to be tested. This is pure observation, to see what answers pop out. No hypothesis is required. (This is inductive reasoning.)

Hypothesis required:

Is mortality from ruptured aortic aneurysm associated with a delay in presentation to hospital? This sounds similar, but on reflection, it is very different. **You are hoping to prove something specific.** The research question can (and should) be rephrased: *Does delayed surgery correlate with mortality?* This is a prediction. It suggests a hypothesis. The hypothesis is that *Delay in surgery contributes to mortality in patients with ruptured aortic aneurysm.* Your study will be designed to prove this (the hypothesis), or, alternatively, fail to prove this (the null hypothesis). This is deductive reasoning.

A common error

A common error in proposals is that students present their study as purely observational (*What interesting things are there to see if I look at these 50 patients?*)—which has no hypothesis, yet, when they discuss their projects, it becomes clear that they actually do have a hypothesis in mind (*I’m interested to study these 50 patients because I think we will find that patients from the rural areas have worse outcomes than those from the city.*) This is a hypothesis. Understanding the difference and recognising how it applies in your own research is an example of the *clarity of thinking* that characterises good scientific reasoning, and a skill you need to develop. If you are unsure whether your project requires a hypothesis, then you do not understand what it is that you are hoping to achieve with your study.

Learning about hypotheses

The hypothesis sits at the heart of the **scientific method**. You cannot practise good science without a deep understanding of the scientific method. Take the time now to deepen your understanding of the scientific method, drawing on ChatGPT.

Have a conversation with ChatGPT. Here are some key terms to start you off:

Hypothesis, hypothesis-driven research, observational study, descriptive study, interventional study

Now dig deeper into the logic on which the scientific method is based. Here are some more key terms and approaches:

Inductive, deductive and abductive logic, the scientific method, falsifiability; the difference between *proving*, *failing to prove* and *disproving* a hypothesis.

You will start to see how experimental and observational research concepts, inductive and deductive logic, hypothesis testing and the scientific method all fit together. You will start to understand the philosophical framework in which your project is situated. **Do not restrict yourself to asking *What is...* with each of the above terms.** The key is that they are all interrelated and support the structure we call the scientific method. Prompt ChatGPT with queries that engage with the interactions. *What has deductive logic got to do with the scientific method? Is failure to prove the same as disproving a hypothesis? Can I do good research without a hypothesis?* Even oddball questions, such as *Is the scientific method compatible with a belief in magic? Do politicians make good scientists? and Do politicians need a hypothesis?* will throw up answers which deepen your understanding. Your tutor might laugh at you or tell you to stop wasting their time with such a question. ChatGPT will remain courteous and provide an often surprisingly wise answer. If you are dissatisfied with a reply, ask again.

As always, **do not skip over terms you do not understand in the replies.** Drill down on them by taking them up with ChatGPT. By filling in **all** the gaps, you develop a complete understanding. This is deep learning, which, ultimately, corresponds to the construction of complex conceptual structures (so-called *mental models*) in the mind. These structures contain not just items of knowledge, but items bound together in a complex web of associations. Interacting with ChatGPT in this way is a powerful way to build up these mental models. Textbooks, by contrast, are poor at this, and lectures usually fail utterly.

Bring your own research project into the conversation, e.g. *If I am studying 50 patients with glomerulonephritis to see if there is a threshold above which patients need dialysis, should I have a hypothesis?* Use your conversations with chat GP to learn around issues, and then apply what you have learnt to your research question.

Reflection and application

Now that you have all this knowledge at your disposal, decide whether you need a hypothesis or not. If the answer is *yes*, write out your hypothesis. It is the thing you wish to prove. Run it through ChatGPT and ask for comments. If you decide that you do not need a hypothesis, write a sentence explaining why a hypothesis is not required. Finally, given your new insights, do you perhaps need to go back to your research question and update it?

Time for reflection: changing your research question

Both observational and hypothesis-driven studies are useful and scientifically valid. One is not better than the other. However, hypothesis-driven research has a role beyond observational studies, and may be of greater interest and impact in the concept of your research problem. It is appropriate to ask yourself at this stage: *Would I like to change my research question, such that it is hypothesis-driven, and not just observational?* It is often just a change in emphasis, with little extra work involved, but results in a better and more meaningful study.

This is a good time for some extra learning. *Discuss differences in value, impact and rewards between these two types of research with ChatGPT.*

Are you still happy with your research question?

AIMS

Proposal – described in detail
 Protocol – perfected and finalised

THE RATIONALE IS NOT THE AIM

As a concept in research, *aims* are poorly understood by many students. The research aim is *to answer the research question*. It is not to do anything else. It is *about the research*, not external considerations such as possible uses for the findings.

Example

Research Question: *Is Drug A more effective than Drug B in prolonging life expectancy?*

Correct aim: *To determine whether Drug A is more effective than Drug B in prolonging life expectancy.*

Incorrect aim: *To improve the care of cancer patients.*

The last statement is not an aim in the research sense. It is termed the **rationale** for the study—the reason *justifying* the research.

Because the *aim* of the study is to answer the *research question*, they are frequently worded very similarly:

- **Research question:** *Is Drug A more effective than Drug B in reducing mortality from cancer?*
- **Aim:** *to determine whether Drug A is more effective than Drug B in reducing mortality from cancer.*

Because a *hypothesis* links question and aim, it, too, is often worded very similarly:

- **Hypothesis:** *Drug A is more effective than Drug B in reducing mortality from cancer.*

LEARNING ABOUT RESEARCH AIMS

Have a lengthy conversation with ChatGPT. See what it can tell you about the *aims* of a study, and *how they relate to the rationale for the study, the research question, the research problem, the hypothesis and the objectives*. Ask it for good and bad examples. Keep asking until you thoroughly understand what goes where in the proposal regarding study rationale, research aims and objectives.

Think holistically in choosing your questions. Do not just stick sequentially to *research question, hypothesis, and aims*. Ask ChatGPT questions which integrate all three of these. You might also want to run ahead by including objectives in your conversation. (We move on to objectives in the next section.) Try asking in different ways. Do not just stick to *Tell me about* or *Describe*. You can also ask direct questions, such as *Is the study rationale the same thing as the study aim?* As we have stressed repeatedly, such questions elicit more meaningful answers. The more thoughtful and intelligent the question, the more instructive and helpful the answer. You can also make your prompts personal: *I get confused between aims and objectives, or Why am I struggling to understand the difference between aims and objectives?* ChatGPT will respond appropriately.

Warning

Do not rush this section! This is possibly the single most important part of your proposal. Clear and correct aims and objectives are essential. Theses fail because poorly conceptualised aims and objectives do not speak to the research question or match the work done for the thesis.

REFLECTION AND APPLICATION

Now write down your research aim. Look back at your research problem, research question, hypothesis and aims. Are they well-matched? If not, make the necessary adjustments. Discuss your aims with ChatGPT. Could they be improved?

OBJECTIVES

Proposal – described in detail
Protocol – updated and perfected

Students frequently confuse aims, objectives and methods. Each has a distinct identity, purpose and place. The *aim* of the study is what you intend to learn, understand or prove from the study as a whole. Typically *a study aims to answer the research question*. The *objectives* of the study are the steps you need to complete such that, when all objectives have been met, the aim will have been achieved and the research question answered. (The *methods* are the actions you will undertake to achieve those objectives.)

Analogy

Here is an analogy, using an imaginary journey from Durban to Mthatha:

- Our **aim** is to *arrive* at Mthatha. Note that the **aim** is **not to drive** to Mthatha. The aim is an *outcome* - arriving. Driving is the *method* we will employ to achieve our aim. And there are alternative methods: we could ride a bicycle, or take a bus: these are different methods to achieve the same aim.
- Objectives:
 - Our **first objective** might be to reach Port Shepstone, where the route leaves the coast and moves inland. Having met this objective, we have made progress towards meeting our aim.
 - Our **second objective** might be to reach Kokstad, where the route changes direction from north to west, leaves KZN and enters the Eastern Cape.
 - Our **third objective** would be to reach Mthatha.

Once we have met **all three objectives**, we have achieved our aim. Our **methods** are the processes we will apply to achieve our objectives: in this example, a combination of navigation and driving.

Example

Aim: to determine the utility of Drug A in the treatment of cancer.

Objectives:

1. To compare the mortality and side-effect profile of patients treated with Drug A with those receiving conventional therapy.
2. To compare the quality of life of patients treated with Drug A with those receiving conventional therapy.
3. To compare the costs associated with Drug A-based therapy with conventional therapy, controlling for life expectancy, side effects and quality of life.

Here you can see that the *aim and objectives are neither the same nor interchangeable*. As each objective is ticked off, we get closer to achieving the aim. In many PhD studies, each objective might match a particular chapter or research paper emanating from that section of the study. From the above example, it is easy to see that it is very likely that the researcher will publish three separate papers, corresponding to objectives 1, 2 and 3 respectively.

A word on master's projects

Master's projects often have a single objective (and, if published, result in a single paper). With such a project, it happens quite frequently that the aim and objective are very similar and possibly identical. Essentially, the aim is being achieved in one step. Using our driving analogy, *arrival in mThatha* (aim) required three distinct navigational steps: *getting to Port Shepstone* (objective 1), *getting to Kokstad* (objective 2), and *getting to mThatha* (objective 3). If, however, our eventual destination was less ambitious, *arrival in Umhlanga* (aim), then this could be achieved in only one step, *getting to Umhlanga* (objective 1, the only objective).

LEARNING ABOUT OBJECTIVES

1. Continue your conversation with ChatGPT, exploring the differences and synergies between aims and objectives and studying multiple examples.
2. Ask ChatGPT to provide examples of both good and bad objectives, and examples where aims and objectives are confused.

HOW MANY AIMS AND OBJECTIVES SHOULD THERE BE?

This is the wrong question to be asking. It depends entirely on the complexity of your research question and how many steps are required to achieve your aim. The best advice is to reduce your aim to as few outcomes as necessary for your research question. That is sufficient if this adequately describes what you need to achieve your research question. This is frequently the case in a simple master's research project. For many master's projects, there may be a *single aim* and a *single objective*, and the aim and objective *are frequently more or less identical*. So, in summary, simple projects are likely to have a single aim, and often just one or two objectives. More complex projects, particularly for the PhD, may have more than one aim and 4-6 objectives.

Do not pad your aims with unnecessary statements. This is evidence of poor scientific thinking because conciseness and clarity are prized in science. Appropriate scientific thinking reduces all elements to the minimum necessary to describe a situation. If you can answer your research question with a single aim, then that is all the aims you need.

Sometimes a student thinks that having more aims and more objectives improves the project. This is not so. At least, you can expect the examiner to ask you to remove or combine some objectives if doing so would indicate a clearer concept of what is being done. More dangerous are orphan aims and objectives, which appear in the thesis but are never actually addressed by the research described in the thesis. Examiners are very hard on this.

How many objectives a researcher puts down may also be affected by their particular writing style. For example, one researcher may list *describe the patient's quality of life*, *calculate life expectancy* and *assess the cost-effectiveness of treatment* as three separate objectives. Another may string all three together under one objective (as in the previous sentence). *It does not really matter*. All that matters is that the researcher clearly understands what they intend to do and has expressed this clearly and unambiguously in the proposal.

PRIMARY AND SECONDARY AIMS AND OBJECTIVES

Primary and secondary aims do not have the same meaning as Aim 1 and Aim 2. A *secondary* aim is one which, though not the main aim for which the research is performed (i.e. the primary aim), is something that can also be addressed during the course of the study. Think of it as an additional opportunity the researcher can take advantage of.

Example

Researchers plan to study the effectiveness of disease control in KwaMashu school children with asthma.

Aim: to determine the extent to which disease is controlled in asthmatic schoolchildren in KwaMashu, Durban.

The method they employ is to identify *all* children with asthma in five schools in the area and determine how symptomatic they are, using a variety of instruments. But wait—identifying every child with asthma in all five schools will **also** allow them to estimate the prevalence of childhood asthma in KwaMashu. All they have to do is divide the number of children with asthma by the total number of children enrolled in the five schools. This is not the study's primary aim, but it is something valuable that could emerge as part of the study. This might be a *secondary* aim.

Primary aim: to determine the extent to which disease is controlled in asthmatic schoolchildren in KwaMashu, Durban

Secondary aim: to determine the prevalence of asthma amongst KwaMashu schoolchildren.

Primary objectives will address the primary aim: secondary objectives will address the secondary aim. In this example:

Primary objectives

1. to characterise asthma control via a researcher-administered questionnaire
2. to characterise asthma control prospectively using an asthma diary to be kept by the subjects
3. to determine asthma control using spirometry administered at two time points during the study period

Secondary objective

1. to determine the prevalence of asthma in this population using school enrolment records as the denominator.

Discretion

There is always room for discretion here. One researcher may decide to put down a primary and secondary aim for the project above. Another may decide to put them down as two primary aims.

- Aims:**
1. to determine the appropriateness of disease control in asthmatic schoolchildren in KwaMashu, Durban.
 2. to determine the prevalence of asthma amongst KwaMashu schoolchildren.

A third researcher may decide to state these two aims in the reverse order. It all depends on how much weight you place on each of them: is the first more important to your research than the second,

the second more important than the first, or the first clearly the most important, with the second being demoted to a “secondary” aim. *There is no right or wrong as such: what is necessary is that you have a very clear idea of what you intend to do in your research* and that your aims and objectives reflect this accurately.

FURTHER LEARNING

1. Have a conversation with ChatGPT about primary and secondary aims and objectives and about how many aims and objectives are the most appropriate.
2. Get ChatGPT’s opinion on the manner in which you have formulated your own aims and objectives.

REFLECTION AND APPLICATION

Are you confident that you fully grasp the concepts of a research question, aims and objectives? Could you explain everything you have learnt to someone else? If not, read through all the notes you saved in your notebook and continue the conversation with ChatGPT until you really do understand everything.

You are now in a position to write down your objectives. Do so. Discuss them with ChatGPT (and, obviously, your supervisor!)

TAKING STOCK OF RESEARCH QUESTION, AIMS AND OBJECTIVES

Your next task is to review the research question, aims and objectives you have written for your proposal. Do they fit together perfectly? Common errors include objectives which do not appear to match a particular aim, or aims that cannot be met using only the objectives listed. Worse still is when the aims and objectives appear incapable of answering the research question.

Comment

The research question, aims and objectives **are the most critical elements of your protocol**. (After all, if you do not know where you are going, you may end up somewhere else!) Returning to our driving analogy: if you fail to state that your aim is reaching Mthatha, you could end up anywhere in the country if you get in your car and start driving. Similarly, if you fail to put down reaching *Kokstad* as an objective, then you might not turn at Port Shepstone, continuing to drive south until eventually getting stuck in the mud on a bad road somewhere in Pondoland.

Reviewers and examiners pay special attention to your aims and objectives. If these are incorrect or inappropriate for the project, this is taken very seriously as **it is assumed to reflect a lack of understanding of the scientific method**. Where examiners pick up these problems, they almost always recommend that the thesis be rewritten and returned for re-examination. Sadly, in some cases, it turns out that the situation is beyond repair. There is just too much mismatch between the work done and any possible combination of acceptable aims and objectives (even if rewritten). **In this case the thesis fails**. (Remember that, according to university rules, a failed thesis is the end of the road. It cannot be fixed and resubmitted.) This is analogous to being so hopelessly stuck in the mud in Pondoland, that it is no longer possible to get to Mthatha without returning to Durban and starting again.

Last word

The research question, aims and objectives you select and put down in your proposal must be appropriate for your project. Be prepared to discuss them and even defend them, not just with your supervisor, but with external advisers.

THE NEXT STEP

Now that you have some confidence in your research problem, question, aims and objectives, and therefore in the outline of your study as a whole, you are in a good position to think up a title and to write up your study rationale, background and literature review.

BACK TO THE BEGINNING

By this stage, you should have an excellent understanding of your project in terms of research question, hypothesis, aims and objectives. This allows you to choose an appropriate title, write a literature review that speaks directly to the project, and write down, accurately, a research problem that is directly addressed by the research question and the project as a whole.

TITLE PAGE

TITLE

Proposal – tentative Protocol – finalised
--

Provide a concise and descriptive title of the study.

Titles are often very poorly chosen. Make sure that your title captures the essence of the project. A reviewer reading your title should immediately know the topic of your study. Think carefully before including methods in your title.

Here are some examples of well-chosen titles:

Effectiveness of a novel mRNA vaccine in reducing severe COVID-19 cases among high-risk populations. The title clearly states what is being studied (mRNA vaccine), the outcome of interest (reducing severe COVID-19 cases), and the target population (high-risk groups). It also hints at the innovation aspect (novel vaccine).

The role of inflammatory markers in predicting cardiovascular events in patients with chronic kidney disease. This title identifies the key variables (inflammatory markers, cardiovascular events) and the population (patients with chronic kidney disease), and suggests the research's predictive focus.

A retrospective analysis of risk factors for stroke in patients with atrial fibrillation: a 10-year observational study. "Retrospective analysis": indicates that the study looks back at previously collected data, "Risk factors for stroke": specifies the main focus of the study, "Patients with atrial fibrillation": defines the population being studied, "10-year observational study": provides a timeframe, emphasising the long-term nature of the data and the observational study design.

The above examples suggest that the title is mainly about *what my project is about*. Sometimes it makes sense to incorporate *How I am going to do it*. If so, a two-part title is sometimes appropriate, such as: *The impact of gut microbiota on metabolic syndrome in obese adults: a randomised controlled trial*. This title is specific, indicating the population studied (obese adults), the focus (gut microbiota and metabolic syndrome), and the study design (randomised controlled trial). It gives a clear picture of the scope of the project.

Learning about titles

Have a discussion with ChatGPT about good and bad titles for research projects.

Reflection and consideration

Decide on your own title and run it past ChatGPT

SUPERVISORS

Proposal – list supervisors and co-investigators if known Protocol – final list of supervisors and coinvestigators in detail

Provide the names, titles, qualifications and affiliations of your supervisors.

INVESTIGATORS

Provide the names, titles, qualifications and affiliations of the principal investigator, supervisors and co-investigators. You, the student, are typically the principal investigator. Your supervisors are automatically assumed to be co-investigators and should not be listed as co-investigators. As co-investigators, you will list anyone else fully involved in the project who is not a supervisor.

Who is entitled to be listed as a co-investigator?

This is sometimes problematic. People may wish to be listed as co-investigators and co-authors in a subsequent publication, even though they are not contributing to the project. People providing patients or data, running tests or performing statistical analyses are not considered co-investigators. Nor is anyone entitled to claim to be a co-investigator or a co-author purely by virtue of their position, such as the academic head of department, unless they are supervising. Journals, funders, and ethics boards are now very strict about knowing who does what in the project and ensuring there is no free-riding.

Ask ChatGPT who qualifies to be an investigator on a project, and who qualifies to be an author on a paper.

DATE

Record the version (draft) number and date of the latest version. No proposal or protocol is ever perfect on the first draft. **As with all good scientific writing, multiple drafts are almost always required before it is acceptable.** If you want to submit your protocol on the first or second draft, you are either a scientific genius (unlikely) or have not fully understood the scientific method and the principles behind the protocol (perhaps more likely).

BACKGROUND AND RATIONALE

Proposal – brief, just sufficient to contextualise the research problem
 Protocol – in some detail, fully referenced

This section is here to contextualise the research project by providing the background to the research problem, what is known about it and previous attempts to solve it. **It justifies the research question, aims and objectives.** This means you cannot adequately complete this section until you have definitively finalised your question, aims and objectives. This is why we have suggested in this guide that you leave this section until after attending to these sections.

It goes without saying that you must have done an adequate literature search before starting the writing of the proposal, as you require this background to do it properly. We are not recommending that you leave the *reading* until you have written most of the proposal, only that you leave the *writing up* of the background until this time.

Keep the purpose of this section in perspective. You are not writing a textbook on a particular topic of interest to your project. This section is primarily about your research-centred focus within a topic, not the topic itself. For example, if your research question is about the prevalence of asthma in Durban children, then your background and literature review focus on asthma prevalence in children, prevalence in contexts similar to Durban, and prevalence in Durban itself. *Do not spend time on asthma in general, its pathophysiology, diagnosis or treatment.*

It is, however, permissible to introduce the topic at a high level extremely briefly, often with one or two sentences:

Asthma is a chronic inflammatory disorder of the airways characterised by variable airflow obstruction and bronchial hyperresponsiveness. Childhood asthma represents a major public health challenge, with implications for growth, development, and long-term respiratory health. In emerging economies, the burden of childhood asthma is increasing, driven by rapid urbanisation, environmental exposures, and disparities in healthcare access.

This is followed by a referenced discussion of the literature in a context appropriate to your research: paediatric asthma in lower- to middle-income countries, Africa, South Africa, Durban and so on, which forms the central part of the discussion.

LEARNING ABOUT THE BACKGROUND AND LITERATURE REVIEW

Have a conversation with ChatGPT on the purpose of this section, and what you should include in it.

THE RESEARCH PROBLEM

Provide a brief description of the problem or knowledge gap the study addresses. Discuss your research problem and research question with ChatGPT and ask whether they are well-matched.

LITERATURE REVIEW

Proposal – brief, just sufficient to contextualise the research problem
 Protocol – in some detail, fully referenced

LEARNING ABOUT SCIENTIFIC WRITING

If you are a master's student, this section may be your first exercise in scientific writing. Scientific writing is very prescriptive and requires knowledge and practice to do it well.

Ask ChatGPT to summarise the principles of scientific writing for you. When you have finished writing our review, copy what you have written and paste it back into ChatGPT. Ask ChatGPT to comment on whether you have achieved a reasonable style, and ask it to help you improve it.

SUMMARY OF EXISTING RESEARCH RELATED TO THE TOPIC

Briefly summarise what work has been done in the area of your research question, describing the current state of knowledge. Provide references for key papers in the field.

RATIONALE

Provide a justification for the study, explaining why the research is necessary, and what the benefits of the results may be for knowledge or society. Discuss the rationale you devise with ChatGPT.

REFERENCING

Provide key references relevant to the foregoing. Being able to reference properly is a fundamental skill in academic writing. This includes three aspects: knowing what should be referenced, how to reference using a standard referencing style (e.g. Vancouver), and how to use referencing software, such as EndNote. The following guidelines are reproduced from the SCM document: *SMPG-1-A guide to the preparation of the MMed thesis*.

The quality of your references

The references you provide should be *the most appropriate* for your proposal. Examiners are asked whether the student has demonstrated an adequate understanding of the literature in the field. Students who appear unable to identify the most appropriate references do not meet this criterion and should fail.

The appropriate references are (1) most relevant to your work and (2) most credible. If your research project concerns diagnosis, do not quote references that concentrate on therapy. Do not reference a small case series in a low-quality journal when there are far better alternatives, such as large case series in high-quality journals. Doing so usually implies that you are lazy and just making use of any reference that comes to hand. Under no circumstances should you reference the popular press and web pages other than where the web page contains an important report from an authoritative body, for example, a piece of legislation or official guidelines.

Never use a reference without having read it. In particular, do not rely on an article's abstract or a reference to that article from another source. You will be surprised how often abstracts and other people's references fail to reflect the article's contents accurately.

Always ensure that the references you supply agree with whatever you are referencing in your text. It is common to find that a student uses a reference to support a particular point of view when that reference states the opposite.

Formatting your references

This must be done accurately and consistently. The Vancouver system should be used for all proposals and theses except for some in the social science domain. You must note that the Vancouver system lays down precise rules for how citations (the references in the text) and bibliographies (the reference list at the end of the manuscript) are structured and formatted. You must have studied the instructions for their use, either independently or by following the instructions given to authors by a particular journal. It is not just a matter of putting a number in the text and writing some words in the bibliography.

Note that printed journal references are generally simple to format. References to chapters in books, books, e-journals, official reports and webpages are more complex and often require you to put some extra work into obtaining all the details necessary to reference them correctly (for example, the city in which a book was published, or the DOI number of an electronic article). You must provide these references correctly.

You must know how to use a reference manager like EndNote or Mendeley. Your references must be marked as the correct type so that your reference manager can format them properly in the bibliography. If chapters in books, books, electronic articles and webpages are marked as the default "journal articles", they will be formatted incorrectly. Here is a summary of the essential reference types and how they should be structured in Vancouver style.

CONTEXT	FORMAT	EXAMPLE
Journal Articles	Author(s). Title of the article. <i>Abbreviated title of the journal</i> . Year;Volume(Issue):Page numbers.	Smith AJ, Clark A, Johnson B. The impact of hypertension on kidney disease. <i>J Nephrol</i> . 2020; 15 (3):145-52.
Books	Author(s). Title of the book. Edition (if not the first). Place of publication: Publisher; Year.	Brown HJ, Miller T. Renal Pathophysiology. 3rd ed. New York: Springer; 2018.
Chapters in books	Author(s) of the chapter. Title of the chapter. In: Editor(s), editor(s). Title of the book. Edition (if not the first). Place of publication: Publisher; Year. p. Pages of the chapter.	Peters MD. Diagnosis of chronic kidney disease. In: Harris JL, editor. <i>Advances in Nephrology</i> . 2nd ed. London: Academic Press; 2017. p. 89-102.
Webpages:	Author(s). Title of the webpage/document. Website name. URL. Published date (or updated date); Accessed date.	World Health Organization. Chronic kidney disease. WHO [updated 2023 May 12; cited 2024 Jul 19]. Available from: https://www.who.int/health-topics/chr .

Electronic journal articles	Where you know the DOI Author(s). Title of the article. <i>Journal Name</i> . Year; Volume :Pages. doi:DOI.	Smith AB, Johnson CD. Impact of lifestyle interventions on metabolic syndrome. <i>J Clin Nutr</i> . 2023; 50 :200-10. doi:10.1016/j.jcnut.2023.03.010.
Electronic journal articles	Where you do not know the DOI Author(s). Title of the article. <i>Journal Name</i> [Internet]. Year; Volume (Issue):Pages [cited Year Month Day]. Available from: URL.	Smith AB, Johnson CD. Impact of lifestyle interventions on metabolic syndrome. <i>J Clin Nutr</i> [Internet]. 2023; 50 :200-10 [cited 2024 Oct 22]. Available from: https://www.jcnut.org/article/12345 .

Explanatory notes

Formatting with bold and italic. Journals have different expectations for using bold and italic in Vancouver references. A standard format is to use italics for the journal name and bold font for the volume number. This helps break up the parts of the reference. We suggest you do so.

Journal names. All journal names should be abbreviated using the official abbreviation (e.g., S Afr Med J). Do not use your own abbreviations, e.g., SAMJ. No full stops are used in the abbreviations. The abbreviations are available from Index Medicus or the journal website.

Issue number, e.g. J Nephrol. 2020;15(3):145-52. The issue number here is the (3). The rule is that this need only be included where each issue has its own numbers, starting from 1. (There must be very few journals which still do this.) For journals where the whole volume is numbered consecutively, it is unnecessary to put the issue number.

Page numbers in books and book chapters. In most cases, it is sufficient to reference the book or chapter without supplying exact page numbers.

Webpages. The reason for providing the cited date is that web pages, unlike printed pages, can change. Your cited date says that you are saying that that is what the website said on the day you accessed it.

Electronic journal articles with DOI. Many e-articles now carry a DOI (digital object identifier). This is linked to that article, as it was published at a particular time before any update. The article linked to it cannot change. You can also access the article directly from the DOI. Therefore, it is unnecessary to state the date you cited it, nor provide its web address.

Electronic journal articles without DOI. The page can still be updated if there is no DOI but only a website address, meaning your reference may become outdated or inaccurate. Therefore, you have to give a citing date that says that the article said that on that date. You must also provide the URL and web address so your readers can access it. Wherever possible, use the DOI.

RESEARCH METHODOLOGY

Now that you have clarity on your research question, aims, and objectives, you can decide how you will answer the question, or, in other words, decide on the methods you will employ in your study. When you undergo surgery, you expect your surgeon to know exactly what they are doing, to be able to choose the correct technique, justify why they chose it, and apply it correctly. They can only do all this if they have a broad knowledge of all the possible techniques available. **The same applies to research.** Many different research methods and methodologies can be applied to answer a research question. You need to have a sufficiently broad knowledge of them to choose the most appropriate. Do not just do what someone tells you to do, or what seems appropriate at first sight. A different choice of study design could significantly improve many studies! You cannot take advantage of these alternatives if you are unaware of them.

PROPOSAL VERSUS PROTOCOL

From this point on, proposal and protocol start to diverge. Your proposal will cover all the following in broad terms, enough for protocol reviewers to judge whether you are likely to achieve your aims using the subjects and methods you describe, or whether there might be better ways of achieving this. In the protocol, full detail is required; essentially, you must record, in detail, precisely what you will do at each step of the project, which includes which subjects are included in the studies, where the study is conducted and so on.

Indeed, your ethics permission will hold you to this detail, more or less word-for-word. If your protocol states that you will recruit patients aged 70 and younger, and you subsequently find that 60 would be a more appropriate cut-off, you are obliged to submit a formal amendment request to the ethics committee. Similarly, if you state that patients will be recruited from a number of clinics, which you name, and you subsequently wish to include patients from an additional clinic, you are obliged to submit a formal amendment request. The same holds for a change in method, sample size, supervisor, co-investigator and many other details.

This degree of detail is inappropriate for the proposal. Your proposal has to be sufficiently flexible to allow for changes in response to reviewers' suggestions. It would be very inefficient to have committed yourself, at this stage, to the degree of detail listed above, at the stage where it is likely that change will be necessary.

STUDY DESIGN

Proposal – broad outline only Protocol – detail
--

Type of study

This describes the study design (e.g., randomised controlled trial, cohort study, case-control study, etc.) You cannot answer choose the right study if you know nothing about these. One of the goals of the master's and PhD programmes is that you learn about the full spectrum of study design, even though you may only employ one such design in your own project.

Learning about study design

The following is a list of some key concepts which must be understood by *anyone* drawing up a research proposal:

- Positivist, interpretivist and mixed-methods research
- Quantitative and qualitative research
- Research paradigms
- Retrospective and prospective studies
- Study design (descriptive, observational, interventional, cross-sectional, cohort, etc)

The following terms will have come up in your conversation:

- Bias and confounding

Have a deep conversation with ChatGPT about all these concepts. Ask for plenty of examples so that you understand the concepts well. Remember that one of the most effective methods of deep learning is by comparison and contrast, and ChatGPT is ideal for this. For example, phrase your request as *Is a retrospective study the same as a prospective study?* (It doesn't matter that the question is nonsensical; it will elicit a useful answer!), or *Is a retrospective study better than a prospective study?*, or *Is a prospective study more scientific than a retrospective study?*, or *Can a controlled trial be retrospective?*

Remember to use the key phrases that help unlock ChatGPT's potential: *Explain further*, *Give me more detail*, *I don't understand* and *Give me examples*.

Decision on study design

Now that you understand the study design spectrum, decide which method is most relevant to your project. This is usually a two-way process: You and your supervisor have probably mapped out a methodology for your study. Consider the possibility that a different design would answer your research question better. You may well find that this is so. You may even find that one of the research designs you have read about seems so appropriate that it is worth going back and adjusting your aims and objectives to enable you to utilise this design and conduct a much more worthwhile project. (Ask ChatGPT for examples of this.)

This is another critical aspect of the scientific method. Rather than plodding through your proposal, one step after another, you should always think backwards and forwards, ending up with the most scientifically valid and productive combination of research question, aims, objectives and methods possible.

Consideration and reflection

Enter your study design into your protocol. Are you confident in your choice of an interpretivist or a positivist design? Are you confident in your choice of a retrospective or prospective study? Are you sure that your choice of study design can answer the research question? Is it feasible? That means that you have sufficient time and resources to complete the study. If the answer is *no* to any question, this may invalidate your study, and you need to rethink your project.

What does ChatGPT think about your choice of study design? Will it answer the research question? *Ask it!*

Do not continue until you understand what you have written, explain it to somebody else, justify why it is the appropriate design, and describe the strengths and weaknesses of some of the other designs you might have chosen instead.

Remember that the purpose of training master's and PhD students is not just to get a research project done and see the results. It is primarily about training those students to understand science and scientific methods, and to think like scientists. This means you must be able to position your own chosen study design in relation to other possible study designs.

Warning

It is common for students to plan a project that is acceptable from the point of view of the degree, but a slight change in study design could result in a **much stronger project** with a higher chance of being published in a high-quality journal.

STUDY SITE

Proposal – broad outline only Protocol – detail
--

Comment

You may wonder whether these items are specific to studies involving human subjects. Do they apply to animal experimentation or laboratory work?

Have a conversation with ChatGPT on this issue. By the end of that your question should have been answered.

Study site(s)

Record the locations where the research will be conducted. Be specific, e.g.: *This study will be conducted in the outpatient section of Wentworth Hospital, Durban. This study will be conducted in three schools in KwaMashu Durban: School A, School B and school C.*

Think carefully about your study site. The choice of site can have a major effect on the impact of your research. Often, feasibility and impact correlate negatively. (Type this phrase into ChatGPT and ask it to explain and give examples.)

STUDY POPULATION

Proposal – It is worth planning this in some detail. Small changes in population can have major effects on your study's feasibility, validity and generalisability. Protocol – detail
--

Describe the study population. Narrow it down **as far as possible to the group you wish to study**. You will then fine-tune with inclusion and exclusion criteria, as discussed in the next section.

Common error: What is a population?

The term *population* in research does not mean the same as the term *population* in general use. As you saw in your previous conversation with ChatGPT, your protocol would employ the concept of a population with animal research and, in more restricted circumstances, with basic science research (quoting, for example, the source of cell lines), even though you are not dealing with people. The concepts of population and sample are closely related in research and are foundational to statistics.. Understanding this relationship is critical in planning your protocol, and even more so considering the application of statistics to your study.

Have a conversation with ChatGPT about the concept of a population in research, and how populations and samples are related. Ask it what the concepts of population and sample have to do with statistical testing. Ask it for examples of populations used in the context of the research protocol for human, animal and laboratory research.

Example

We return to the example of the asthmatic schoolchildren in KwaMashu we discussed earlier.

- Aim:**
1. to determine the appropriateness of disease control in asthmatic schoolchildren in KwaMashu, Durban.
 2. to determine the prevalence of asthma amongst KwaMashu schoolchildren.

What is the study population? The following would be *wrong* as study population:

- Population of KwaMashu
- KwaMashu children
- Asthmatic children in KwaMashu

The following is *correct*:

- Children with diagnosed asthma attending five selected schools in KwaMashu.

All the wrong answers are too broad. It is *not true* to say that a child attending some other school in KwaMashu is part of the study population. Indeed, they have nothing to do with the study. You might not even know that their school exists. The correct answer accurately defines the population which will be studied.

Note, however, that *Children with diagnosed asthma attending five selected schools in KwaMashu* is the correct study population for **the first** aim: to determine the extent to which disease is controlled in asthmatic schoolchildren in KwaMashu, Durban. It is **incorrect** for the **second** aim we described: to determine the prevalence of asthma amongst KwaMashu schoolchildren. The difference is that, for this aim, we effectively have to study all the children to determine who has asthma and who does not. Hence, for **this** aim, a more accurate reflection of the study population would be: *All children attending five selected schools in KwaMashu.*

Do not include inclusion and exclusion criteria when you describe the study population. The study *population* remains the same: the *sample* you choose is narrowed down by the additional criteria.

Learning about study populations

Have a detailed conversation with ChatGPT about study populations.

You will find that it brings up all sorts of related, but important, detail, much of which we will look at in more detail in later sections. This includes inclusion and exclusion criteria, sampling methods, etc. Clearly, all these concepts are related and work together to describe the subjects will be studying, which is why ChatGPT brings them up. Query ChatGPT repeatedly, to focus its attention on specific areas you would like to know more about, and to broaden your understanding of these concepts and how they relate to each other. You will also better understand what is going on as you move down our list of tasks for the proposal.

Consideration and reflection

Enter your study population into the proposal. Think about what you have written. Is it a concise and accurate reflection of the group you intend to study?

What does ChatGPT think of your study population? Ask it to comment.

Recruitment process

Proposal – broad outline only Protocol – detail
--

Explain how participants will be identified, approached, and enrolled, including details about informed consent procedures. (Actual details of informed consent will be entered into the protocol under a later section, falling under *Ethics*.)

INCLUSION AND EXCLUSION CRITERIA

Proposal – It is worth planning this in some detail. Small changes in inclusion criteria can have major effects on your study's feasibility, validity and generalisability. Protocol – detail
--

Inclusion criteria *restrict* the sample to a smaller subset of the study population. Typically these criteria are used to enrol only those subjects whom it is ethically appropriate to study (e.g. *those who provide informed consent*), those meeting criteria which confirm their appropriateness for the study (e.g. *patients with a CK level >1000 ng/l* (if this criterion was not applied, participants might be admitted to the study who do **not** actually have the disease we are studying), and some characteristic which makes it feasible to include them in the study (e.g. *subjects who can attend all five study sessions*.)

Do not repeat criteria which are already characteristic of the *study population*. For example, if your study population is *Male patients with...* then you will not put down *Male* as an inclusion criterion, similarly with exclusion criteria. Do not put down *female* as an exclusion criterion. It is redundant: being male is a criterion for being in the study population.

Exclusion criteria are criteria which make a subject's inclusion in the study inappropriate for any reason. Whether a particular criterion is placed as an inclusion or exclusion criterion is often a matter of choice; one is just the negative counterpart of the other. For example, we may **include** patients aged 70 or younger, or **exclude** patients older than 70. The effect is the same. Do not write both!

Learning about inclusion criteria

Ask GPT to explain and give examples of inclusion and exclusion criteria for your type of study. Follow up until you are confident you have understood all the issues involved. In particular, concentrate on the advantages and disadvantages of restricting your study sample.

Warning

Think carefully before restricting your study population by adding inclusion and exclusion criteria. Why?

Positives

From a scientific standpoint, the main reason for restricting your study population is to limit or remove possible confounders. A classic example is comorbidity. If we wish to know the mortality of a particular disease, such as diabetes, it may be best to *exclude* patients who are also HIV-positive because, if a patient in the study dies, we may not know whether it is the diabetes or the HIV that was responsible. (This is a superficial example, but it makes the case.)

Negatives

1. By reducing sample size, you reduce *statistical power*: in other words, it becomes more difficult to show significant differences.
2. By excluding specific subjects, you may introduce *bias*.
3. By excluding specific subjects, you reduce the *generalisability* of your results.
4. By excluding some subjects, you *throw away all the useful things you might learn from studying them*.

The first three are important concepts in their own right. *Have a comprehensive discussion with ChatGPT about statistical power, bias and generalisability.*

The fourth negative consequence is significant but often not considered. It is often appropriate to use a larger sample to answer those questions that can be answered in them, and then restrict the sample in a later part of the study requiring data unavailable for every patient. If you exclude every subject with partial information, you unnecessarily reduce your sample size for the first part of the study.

Example

Our study population includes 90 patients with cataracts who underwent surgery. We are interested in: (1) describing patients with cataracts as they present to Durban hospitals, and (2) their surgical outcome. Forty were treated in a tertiary hospital with excellent records. Fifty were treated in peripheral hospitals where the quality of follow-up after surgery is poor.

The first reaction of many students is to restrict the patients to those treated at the tertiary hospital. This reduces the sample size to 40. But there is **no reason** why you could not include the whole population for the first objective, giving you a sample of 90 subjects. (There is no reason to exclude patients whose data on the initial presentation are adequate, just because data on subsequent follow-up are poor.) You can then restrict the sample for the *second* objective, relating to surgical outcomes, to the 40 where the outcome was documented correctly.

SAMPLING STRATEGY

Proposal – broad outline only Protocol – detail
--

Comment

For some studies, sampling is not an issue. This is when you choose to study every available subject that meets your inclusion criteria. For example, *We will study all patients admitted to XYZ hospital with disease ABC between January 2018 and December 2022.* This is your study population. You do not have to **choose a sample** from this study population since you will be studying all of them.

The situation is different when you do not have the time, resources or even the necessity to study every one of a large study population. It is then appropriate to choose just a sample of them. For example, imagine you wish to assess the satisfaction of outpatients attending a particular hospital with the care they received over a month. The proposed method is a personal interview. It would be impossible for you to interview all 1000 patients who attended clinics in a month. Let us assume that appropriate calculations indicate that a sample size of 50 will be sufficient to estimate patient satisfaction accurately. *Sampling strategy refers to the method you will use to select those 50 out of a possible 1000.* The problem is that you wish to maximise the chances that the 50 you select are properly representative and do not introduce a *bias*. There are many possible sources of bias: imagine, for example, that you choose the first 50 in that month. It may be that those subjects indicate satisfaction with their care, but had you chosen the last 50 of the month, when the pharmacy has run out of stock, or several staff have gone on leave, the satisfaction level might be very different. Similarly, if you only select patients attending on Thursdays, it might be that they will be drawn from specific clinics whose staff are better or worse than the average in the way they approach their patients.

Learning about bias

Unrecognised bias destroys the credibility of research. You must understand the many forms it can take, and consider your sampling strategy very carefully. Picking up bias is another important reason for discussing your proposal with outside people, beyond you and your supervisor.

Have a comprehensive discussion with ChatGPT about bias. Learn about the different forms it takes. (Though they have names, they are not the names that are important. It is the principle.) Ask for many examples and work through them. Discuss the issue of **confounding** in some detail: it is a common problem in science. Ask how to recognise and avoid it. Describe your own study and have a conversation about possible bias that you need to look out for.

Types of sampling strategy

Many sampling strategies are formally named and described. Each is suited to a particular type of study. You need to know and understand them to choose the most appropriate one for your study design.

Learning about sampling strategy

Have a conversation with ChatGPT about sampling strategy. Pick out those parts of its replies which appear relevant, interesting or which you do not understand, and explore them further.

Remember that you can have a **very specific** conversation with ChatGPT about your project. For example, I typed the following into ChatGPT: *I want to evaluate patient assessments of the quality of care using a study population of 1000 patients seen over a single month in the clinic. Is systematic sampling a good idea?* The answer was highly informative. I then followed this up with: *Is there a better technique?* The answer was more informative still.

Exploit ChatGPT's capacity for natural language. You should be using prompts such as *I don't understand*, or *Can you simplify that*, or *Tell me more about that*, or *Why did you not mention bias?*, or *But that is not a confounder—it is actually an association I am looking for* etc. (ChatGPT sometimes errs: when you spot an error and take it up with ChatGPT, you are engaged in deep learning.) **It really is a conversation!**

Consideration and reflection

Now that you understand the extensive range of possibilities for your sampling strategy, choose one that fits your study design and record this in your protocol. Add a sentence justifying your choice.

Discuss your choice with ChatGPT. Does it seem appropriate? Are there better alternatives? Is it feasible? These are all matters you can discuss with ChatGPT.

METRICS, STATISTICS AND SAMPLE SIZE

METRICS

Proposal – broad outline only
Protocol – detail

What are you going to measure?

This is an aspect which is often very poorly handled by new researchers. Positivist research almost always requires you to measure something. **Choosing what to measure may make all the difference between a good and a bad study.** For example, imagine you want to look for an association between the length of a surgical incision and the pain experienced postoperatively. If you are lazy, you may rely on just asking patients how they feel. With more effort, you could utilise a visual analogue scale, or record the total dose of opioid required by the patient, applying the appropriate conversion factors for morphine, pethidine and fentanyl. With still more effort, you might enlist the collaboration of pain researchers in the physiology department, who might have a sophisticated neuro-electrical method for measuring the intensity of impulses conducted by pain fibres. With each step up, the accuracy of measurement increases. So do the likelihood of finding statistical significance (assuming a positive result), the impact of your research and the chances of it being accepted by a good journal.

New researchers often know *what*, in broad terms, they want to measure, but struggle to decide *which specific measurement* they will make to measure it. Here is an example from a PhD protocol in nuclear medicine:

Research question: *Is Isotope A a better radionuclide tracer than Isotope B in the clinical management of liver cancer?* This is promising, but too vague in terms of what is being tested. This becomes apparent when the researcher is asked what, precisely, they are going to measure:

- How are we going to measure *clinical management*? We cannot. It is not a measurable property. What we can measure are *clinical outcomes*.
- This leads to the next decision. How are we going to measure clinical outcomes? Via life expectancy? Mortality? Disease-free survival? Relapse rate? Tumour size? Severity of symptoms? Pain?
- All these give rise to a further set of questions. *What measurement will be made to assess each of these?* For example, tumour size— will this be assessed as size on CT? Or PET? Or angiography? Or macroscopic size at post-mortem? Or on a resected specimen?. All these techniques will give different values. They are not directly comparable. You cannot compare tumour size on CT in one group with tumour size on MRI in another. You have to choose one or more of these metrics as your measurement, and they have to be the metrics **most likely to give a clear-cut result**, while being feasible regarding cost and practicability.
- Each measurement comes with its own issues. Is it affordable? Is it accurate enough? Do we have access to it?
- Each measurement requires a different method. Each has biases and confounders, so the methods cannot be written until the researcher has settled on which measurements will be made to answer the research question. The choice of metric can make the difference between low-quality, low-impact and high-quality, high-impact research.

Accuracy and precision

These are two related but not identical qualities. They are critical for any research involving measurement and are not just high-tech measurements. They are just as applicable to the determination of the prevalence of disease in a sample.

Learning about accuracy and precision

Engage ChatPT on accuracy and precision. The term precision is sometimes used in two different contexts. It will not take you long to understand their importance. Have a discussion on the key issues. Some measurements are inherently more appropriate than others. One has to consider possible metrics in terms of accuracy, precision, affordability, availability, safety and many other considerations. Ask ChatGPT for plenty of examples of both good and bad choices for measurement. Start generally, and then home in on your research project in particular.

How your choice of metric and statistical method may affect precision and accuracy, and therefore your p value

Here is a common example. Two researchers performed the same measurements on the same sample of patients. They are interested in whether life expectancy in patients with a particular form of cancer (measured in weeks) correlates with the patient's age. Researcher A records the exact age of each patient in years. She can perform a Pearson's correlation of age versus survival and finds the result is statistically significant. Researcher B groups the patients into age bands (30-39, 40-49 etc.). Statistically, she now has to look for a difference in mean survival between each age band, using a test such as ANOVA. *The result is now statistically non-significant.* Why? It's the same patients! The answer is that (1) by grouping patients, you are losing information (you have lost precision: now estimating age to the nearest five years, rather than to the nearest year), and (2) you must now utilise a statistical test which only tests groups for a difference in means, but, unlike the Pearson's correlation, is not designed to recognise a consistent trend.

See the section on statistical methods below.

Surveys and questionnaires

Students sometimes choose to base their research on surveys and questionnaires, as they find these less intimidating than other research methods. **Beware: such research is particularly difficult to do well. The outcome is often research of poor quality with an unsatisfying result.** Students expend a huge amount of energy conducting surveys in large sample groups, only to find that the results are disappointing, with no real outcome, and nothing statistically significant.

Survey and questionnaire items are particularly prone to lack of precision and accuracy, bias, confounding and misinterpretation, all of which invalidate the research. Ask ChatGPT what could go wrong with the following simple questionnaire items:

What is your gender? Your age? How long have you been sick? Do you take your medication regularly? As a doctor, are you comfortable treating anxiety?

If you include surveys and questionnaires in your research, you must interrogate every item extremely carefully, ask ChatGPT to evaluate each one, and discuss the instrument with your supervisor and outside commentators in detail. Every survey or questionnaire should be piloted to identify and correct its weaknesses. It should also be validated, especially at the PhD level.

Validating your instrument

This is an essential concept. Discuss it thoroughly with ChatGPT. Ensure you know what is meant by an *instrument*. Master's-level projects may get away with not formally validating instruments such as surveys and questionnaires. At PhD level it is unwise to assume that you do not have to do so.

Warning!

If your research includes a survey or questionnaire, do not ignore the advice given above. This particular method of data gathering is highly prone to error. Do not just write down questions that seem to make sense to you. Examine every question, looking for all the ways subjects and research assistants may misinterpret it.. Ask others to scrutinise your questions. Run them through ChatGPT, asking how that question could be misinterpreted or provide wrong answers.

Consideration and reflection

Finalise the measurements you intend to make. By discussing them with your supervisor and other helpful people, make sure that they are the most appropriate measurements for your study.

Word in closing

Choosing the wrong measurements can **destroy your project**. If it turns out that the techniques are unavailable or unaffordable, your project will never get finished. If the measurements are not precise enough, then you will never reach statistical significance. If the measurements are fundamentally inappropriate to your research question, then your paper will never be published and your thesis may fail.

STATISTICAL METHODS

Proposal – broad outline only Protocol – detail
--

In positivist research, the outcome of your project may well depend on the application of statistical methods to your results. At this point in your proposal, you must outline the statistical methods you intend to use and estimate required sample size.

Important point

Students frequently make the mistake of thinking that all that is at stake here is getting the right words down in the protocol. This misses the point. We discussed measurement and metrics earlier. The metrics you choose, the statistical methods appropriate to their analysis, and the required sample size depend on each other. Ideally, you need to understand simple statistical techniques *before* you settle on the measurements you make. It is highly desirable that you have at least a basic understanding of statistics before choosing your measurements and methods (and certainly before you choose your sample size). It is very unfortunate to collect all your results and only then discover that they are inappropriate for statistical testing most likely to bring out the best in your results.

Think of the example given earlier. The choice between collecting ages as actual years or grouping them into age bands can mean the difference between a significant and a non-significant result. This

is why it is important to know something about statistics when you develop your proposal, and even more so when you develop your protocol. A knowledge of statistics is not something that applies only when you come to analyse your results. It is very useful when planning your project and writing the proposal.

A crash course in statistics

Have a discussion with ChatGPT about each of the following. These cover the most basic statistical contexts and how statistical methods are applied. If you engage for sufficient time and in sufficient depth (all the time following up each answer with questions about anything or any term in the answer), you will find that the same terms are coming up repeatedly, and it is becoming easier to see what is going on. Do not go into too much depth. You can always ask ChatGPT to simplify. The best way to learn is to request repeated examples of everything ChatGPT talks about. You can even test yourself by asking: *Am I right if I say that...*

A note of reassurance

Keep the following in mind and do not allow yourself to be overwhelmed by ChatGPT's information.

- The field of statistics has a highly technical vocabulary. Do not be disheartened by this. Pick out each unknown term in the response and query ChatGPT about it. If you keep doing this, such terms will become increasingly familiar and understood.
- You do not need to do any of the calculations or use the formulae that ChatGPT may suggest. Just instruct ChatGPT to leave these out. Concentrate on understanding the principles and how they can be applied in the context of your research.

Here are some suggested topics to start you off.

Requirement	Learning
Distributions	Ask ChatGPT about the difference between normal and skewed statistical distributions.
Describing your results	Ask Chat GPT how to describe data with normal and skewed distributions when you write up your results.
Variable	Ask ChatGPT about the different kinds of variables. Concentrate on continuous and categorical. Learn what is meant by a <i>factor</i> .
Comparing means of two variables which have (a) normal and (b) skewed distribution	Ask ChatGPT how to do this, and follow up on the statistical tests which emerge in the conversation. Ask for several examples of situations where these tests would be relevant. Also ask in reverse by describing a study question and asking what sort of statistical comparison would be most appropriate. Ask how your data should be presented for these tests.
Comparing means of three or more groups which have (a) normal and (b) skewed distribution	Ask ChatGPT how to do this, and follow up on the statistical tests which emerge in the conversation. Ask for repeated examples, so that you understand what sort of situation calls for these tests. Ask how your data should be presented for these tests.

Correlating two continuous variables	Ask ChatGPT how to do this, and follow up on the statistical tests which emerge in the conversation. Ensure that you follow up on every technical term that emerges in the replies so that you understand everything in that reply.
Comparing categorical variables	Ask ChatGPT how to do this, and follow up on the statistical tests which emerge in the conversation. Ask for several examples, so that you understand what sort of situation calls for these tests. Ask how your data should be presented for these tests.
Test yourself	Describe various types of results a researcher might generate in their study, suggest an appropriate statistical test for their analysis and see if ChatGPT agrees with you. Example: <i>I do a study where I compare the height of men and women. Am I right that ANOVA is the best test to look for a difference?</i> (Not completely right but not wrong either—ChatGPT will explain why and teach you much more besides.)

Consideration and reflection

Do you now have some idea how each of the observations you intend making in your study should be reported, and some understanding of what statistical tests will be appropriate for your data? You may not be particularly skilled or confident yet, but it is far better to have some grasp than none at all

SAMPLE SIZE

Proposal – estimate Protocol – detail
--

Choosing the correct sample size is critical. The knowledge of statistics you picked up through your previous ChatGPT conversation provides the background for understanding why this is so. Calculating sample size is very easy, and numerous online calculators are available to help you. The problem is not the calculation. Understanding why sample size is important and how to come up with the correct parameters to feed into the calculation is critical.

There is no *one-size-fits-all* answer to sample size for a single project. Sample size is specific to a particular analysis. So you might need one sample size to compare A with B, but a different sample size to calculate what proportion of your sample has characteristic C. The idea is to come up with a sample size sufficient to cover all the various analyses you will be performing on your results.

Nor is there a ready-made sample size for a particular study. (Qualitative research is an exception.) You will find that, typically, the sample size you need is entirely dependent on three parameters: *effect size*, or how big the effect is of whatever you are measuring (e.g. the difference between A and B), the *variability* in your sample (given by the standard deviation or variance) and the *power* you are prepared to accept for your study (which reflects your level of comfort with possibly not finding a significant result even when the effect is real, a “false-negative” outcome for your study).

This section is intended to provide you with just enough background knowledge that you have some idea of what is going on. A competent person or statistician must handle the actual calculation of sample size. The principle of *garbage in garbage out* applies here too. If you cannot explain to the

statistician precisely what is to be measured and what is to be compared, they may provide you with the wrong sample size.

Learning about sample size

Have a conversation with ChatGPT. You will find that a brisk discussion with ChatGPT will help you understand the concept of sample size fairly quickly. Start by prompting an introductory discussion on sample size. Then, as usual, follow up on anything you do not understand, and start asking pertinent questions (such as the effects of having too large or too small a sample) and ask for examples. Bring in your project and summarise what you intend to measure for ChatGPT. Discuss the factors relevant to the determination of sample size for your study. Along the way, you will come across some critical concepts that you will hear repeatedly in the future: significance, power, variance, effect size, accuracy, precision, and confidence intervals. Learn what these are.

Remember that integrated questions to ChatGPT lead to integrated answers, **and integration is the key to deep learning**. Instead of asking *What is sample size? What is precision?*, try asking *What has precision got to do with sample size?*

Consideration and reflection

You should now have some understanding of the factors which contribute to your required sample size. See your statistician. You will be able to explain exactly what you require of your statistician, and you will understand their answer. Once you have a sample size estimation, you must consider whether the study is feasible. **If the estimated sample size exceeds the time, number of participants or funding resources at your disposal, then you will have to redesign your project.**

Relax!

Do not be disheartened if the estimated sample size exceeds what is available to you. *It does not necessarily mean planning a whole new project.* Adjusting the aims and objectives is often all that is required. If whatever you now choose to measure has a larger effect size, or less variability in the sample, than what you initially planned, a smaller sample size may suffice.

TIMELINE AND STUDY DURATION

Proposal – broad outline, provisional Protocol – most realistic estimate, detail

Develop a timeline for the study, including follow-up periods. Writing this is usually straightforward. Create a Gantt chart to relate each stage of the project to a particular period of time. This helps plan the project and keeps you moving quickly once you have your ethics permission and start the project.

Gantt chart

A Gantt chart is a type of bar chart that illustrates a project schedule. It shows the start and end dates of the different tasks or activities that make up a project, and the overlap between them. Gantt charts are commonly used to plan and manage projects, as they visually represent the project timeline and dependencies between tasks.

In a Gantt chart, each task is represented by a horizontal bar, where the bar's length corresponds to the task's duration. The bars are typically arranged one above the other, with the tasks labelled down the side. The chart also includes milestones, key events or deliverables in the project, and dependencies between tasks, indicated by arrows. Some bars will overlap: you may, for example, be able to start collecting data from the participants you recruit early while continuing to recruit new participants. Others cannot overlap at all: for example, data collection cannot start until you have received ethics permission.

Gantt charts are not difficult. Do a Google search for descriptions and examples. Constructing them is fairly simple in PowerPoint: stack horizontal rectangles of different lengths on top of each other, aligning them with a horizontal time axis below. Then label each.

METHODOLOGY

Proposal – broad outline only, Do not go into any detail.
Protocol – detail required

Interventions

This will apply to any form of study in which an active intervention is made, such as a drug trial. It does not apply to observational studies. If applicable, provide a detailed description of any interventions that will be applied to the participants. Describe the timing, duration, and delivery of each intervention. If applicable, explain any blinding or randomisation procedures for interventions. Describe procedures for data entry, coding, and storage to ensure data security and confidentiality. Include any quality control measures, such as data checks or monitoring steps, to minimise errors.

A

Data collection procedures

Summarise all steps in collecting data, including the data types and timing. Describe the tools or instruments used (e.g., questionnaires, physical assessments, laboratory tests) and provide details on their validity and reliability if relevant. Specify who will collect the data and how it will be recorded or managed.

It is not necessary to specify every individual detail of your data. If you are administering a questionnaire of 56 items, you clearly do not want to list 56 items at this point in your protocol. Group them under broad headings and refer them to your data collection sheet or questionnaire, which you will add as an appendix to the protocol.

ETHICAL CONSIDERATIONS

Proposal – broad outline only
Protocol – detail

Informed consent

Describe how informed consent will be obtained from participants.

Risk/benefit analysis

Discuss potential risks to participants, the measures to minimise and mitigate them, and how you will respond to potential harm. (This applies particularly to studies with some form of intervention, such as a drug study where some patients may develop adverse effects.)

Confidentiality

Describe how participant data will be kept confidential. In general, deidentifying data is better than anonymising data for our studies. It allows researchers or other authorised people to return to the original patient data if necessary, making it difficult for unauthorised people to do so.

Learning about ethics considerations

Broaden your understanding of research ethics through a discussion with ChatGPT. Ask about the ethics specifically relevant to observational and interventional studies. Refresh your knowledge of informed consent through a conversation with ChatGPT. Ask it to suggest methods for keeping the data you obtain in your study confidential, and ask for examples. Learn about concepts relevant to safe data storage, such as anonymisation and deidentification..

Consideration and reflection

Decide on the steps you will take to ensure that your research follows ethical guidelines and commit them to the protocol.

DATA ANALYSIS AND STATISTICAL METHODS

Proposal – broad outline only Protocol – detail
--

Data management

Describe how data will be managed, analysed and stored.

Outcome measures

This refers directly back to the discussion earlier in this document about metrics. Apply what you learn there to answer this section. Define all outcomes (essentially, what you will measure), explaining how each will be measured. If appropriate, specify the time points at which outcomes will be assessed.

Statistical methods

Not necessary at proposal level

- Descriptive statistics: Explain how you will summarize data using means and standard deviations, or frequencies and percentages.
- Inferential statistics: Describe the statistical tests you will use to compare groups or assess associations (e.g., t-tests, chi-square tests, regression analysis).
- Handling missing data: State how you will handle missing data (e.g., imputation methods or sensitivity analysis).
- Significance level and software: Specify the significance level (e.g., 0.05) and any software used for analysis (e.g., SPSS, R, Stata).

SCIENTIFIC RIGOUR

Rigour refers to the careful design, conduct, and reporting of research such that findings are trustworthy and inspire confidence.

In positivist research, it is demonstrated through *reliability*, which ensures consistent measurement; *validity*, which confirms that the study measures what it intends and supports sound conclusions; *generalizability*, which allows the conclusions to be applied to populations beyond the sample, *objectivity*, which minimises bias; *reproducibility and replicability*, which allow results to be confirmed or repeated; and *transparency*, which enables verification through clear reporting.

In constructivist research, it is demonstrated through *credibility*, which reflects the believability and accuracy of interpretations; *dependability*, which ensures consistency and logical traceability of the research process; *confirmability*, which minimises researcher bias and shows findings are grounded in the data rather than personal preference; and *transferability*, which allows insights to be meaningfully applied in other contexts through thick description.

Consider the rigour of your proposed research

Stop at this point and take stock of your study as currently set out in your proposal or protocol. How does it measure up against each criterion appropriate to your type of research?

Positivist	Constructivist
<i>reliability</i>	<i>credibility</i>
<i>validity</i>	<i>dependability</i>
<i>generalizability</i>	<i>confirmability</i>
<i>objectivity</i>	<i>transferability</i>
<i>reproducibility and replicability</i>	
<i>transparency</i>	

Learning about scientific rigour

Have a conversation with ChatGPT about scientific rigour. Interrogate every one of the descriptors listed in the table. Whenever you come across a word or phrase in the reply that you do not fully understand, query it further until you do understand. After doing this repeatedly, **you should find the same concepts, which you are now beginning to recognise and understand, coming up repeatedly**, suggesting that you are indeed developing a sense of what constitutes a good study. Remember that you can test your understanding against ChatGPT, e.g., *I think that reproducibility and reliability are the same thing. Am I right?*

Consideration and reflection

Is your planned study of good scientific quality? Are you satisfied with its rigour? Is there any aspect you would like to change to increase rigour? For example, extending the age range of your subjects and recruiting subjects from several rather than a single hospital will improve generalisability. Always be prepared to go back and revise sections you completed earlier. Increasing the sample size

will increase reliability. Tightening up on inclusion criteria to exclude confounders will increase validity. Making such changes can significantly enhance the value of your research and increase the probability that it will find its way into a good journal.

REFLECTION, CONSIDERATION AND EDITING

The first draft of your proposal or protocol is now complete. Your next steps are as follows:

- Run it through Grammarly to improve spelling and grammar.
- Read through it repeatedly and revise as necessary. You know how to use ChatGPT to explain anything which you don't fully understand.
- Your supervisor must read your protocol carefully and discuss areas in which it may be improved with you.
- Have your protocol checked by external commentators this is invaluable in picking up problems and suggesting improvements, not only to the writing of the protocol, but for the benefit of the project itself.
- Comply with whatever proposal or protocol review process the school currently requires.
- Compile the final version of the proposal or protocol and check it again with Grammarly.

Submitting a protocol for ethics review

Use the RIG system to make an online ethics application. You will find that your protocol is written largely in the order in which information is required by the application. Cut and paste from your protocol into the appropriate sections of the online application. Upload the entire protocol as a single document using the upload facility provided by RIG.

CONCLUSION

If you have worked conscientiously through the learning suggested in this document and applied it intelligently to your research proposal and subsequent protocol, you should have avoided most of the common mistakes made by young researchers. Your planned project is likely to be of at least an acceptable standard. The final step is external review. You need to work through it carefully with your supervisor and any other knowledgeable person you can persuade to help you. Ideally, you should be subjected to a formal proposal defence.