Research Procedures

3

Ivan Buljan (iD)

Abstract

This chapter offers a guide on how to implement good research practices in research procedures, following the logical steps in research planning from idea development to the planning of analysis of collected data and data sharing. This chapter argues that sound research methodology is a foundation for responsible science. At the beginning of each part of the chapter, the subtitles are formulated as questions that may arise during your research process, in the attempt to bring the content closer to the everyday questions you may encounter in research. We hope to stimulate insight into how much we can predict about a research study before it even begins. Research integrity and research ethics are not presented as separate aspects of research planning, but as integral parts that are important from the beginning, and which often set the directions of research activities in the study.

Keywords

Research plan · Research question · Study design · Sample · Measurement · Protocol registration · Reproducibility

What This Chapter Is About

This chapter offers a guide on how to implement good research practices in research procedures, following the logical steps in research planning from idea development to the planning of analysis of collected data and data sharing. This chapter argues that sound research methodology is a foundation for responsible science. At the

I. Buljan (✉)

Department of Research in Biomedicine and Health and Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia
e-mail: ivan.buljan@mefst.hr

beginning of each part of the chapter, the subtitles are formulated as questions that may arise during your research process, in the attempt to bring the content closer to the everyday questions you may encounter in research. We hope to stimulate insight into how much we can predict about a research study before it even begins. Research integrity and research ethics are not presented as separate aspects of research planning, but as integral parts that are important from the beginning, and which often set the directions of research activities in the study.

Case Scenario: Planning Research

This hypothetical scenario was adapted from a narrative about the process of poor research planning and its consequences. The original case scenario is developed by the Members of The Embassy of Good Science and is available at the Embassy of Good Science. The case is published under the Creative Commons Attribution-ShareAlike License, version 4.0 (CC BY-SA 4.0).

> Professor Gallagher is a leader of a research project on moral intuitions in the field of psychology. She is working on the project with Dr. Jones, a philosopher, and Mr. Singh, a doctoral student. Although she has little experience in the matter, Dr. Jones is put as the principal investigator in the study design and analysis of the two experiments, while Mr. Singh prepares materials and conducts the experiments.

> After the first experimental study, Mr. Singh sends the results to Dr. Jones for analysis. After some time, eager to enter the results in his thesis, Singh asks Dr. Jones about the results of the study. She admits that she forgot to formulate the hypothesis before data analysis, and now the results can be interpreted as confirmatory, regardless of the direction. They decide to formulate a hypothesis that will result in a positive finding.

> Mr. Singh and Dr. Jones present the results to Dr. Gallagher, who is satisfied and proceeds with paper writing. In the second study, Dr. Jones formulates multiple hypotheses before the study begins. Mr. Singh conducts the study and sends the results to Dr. Jones. She performs the analysis by trying to find only significant differences between groups. Finally, to achieve significance, she excludes participants over 60 years from the analysis and while presenting the results, admits that to Prof Gallagher. Prof Galagher is happy about the results and proceeds with the paper writing, while Mr. Singh enters the results in his dissertation.

Before Mr. Singh has the public defense of his dissertation, one of the internal reviewers notices that some data has been excluded from the second study and only significant results were reported. She invites Mr. Singh for an examination board meeting during which MR Singh admits that the data has been excluded and that in the first study hypothesis was formulated after the results were known.

Questions for You

1. Why is hypothesizing after the results are known, as described in the first study, considered problematic?
2. What was wrong about reporting only significant results in Study 2?
3. How would you improve the entire research process described in the scenario?

Good research practice from the European Code of Conduct for Research Integrity:
Researchers take into account the state-of-the-art in developing research ideas.
Researchers make proper and conscientious use of research funds.

What to Do First When You Have an Idea?

It is difficult to come up with a good research idea, and if you struggle to come up with a new research direction, that is perfectly fine. Creative processes are the highest form of learning and developing an idea requires significant cognitive effort. In some cases, you may have an epiphany, where you would suddenly come up with a great idea for your research project. This is something popularized by stereotypes about scientists as eccentric figures who come up with brilliant ways of tackling things using only their intelligence and intuition. However, scientific work resembles ore mining. It takes a tremendous effort to read relevant scientific literature, communicate with your peers, plan, and, in some cases, attempt and fail before you even start digging for gold. As in a mine, you will need to dig a lot of rocks before you come across diamonds and gold.

Usually, the most important decisions are made before digging even begins. To decide where you will start mining, you start with the exploration of the terrain. In research, this means knowing your field of study. You may read an interesting piece in the scientific literature or listen to a presentation at a conference and then

think of a hypothesis whose testing will answer an interesting and important question in your research field. On the other hand, sometimes you have to adjust your research interest so that they fit the specific aims of grant funding calls. It does not matter what the source of the idea is, there are always two things to consider when developing research ideas: the current state of the field and the resources available to you. Good research practice is to consider the state of the art in developing your research ideas and make the proper use of research funds. This does not mean that you are not allowed to develop research ideas if they address a research topic that has been neglected. It is the responsibility of a researcher to combine the best of the "old" evidence with new research developments. It is important to keep in mind that research is not performed in a vacuum and that the funds and resources provided by public or private funders are given with an expectation of an honest answer to a specific research question. The main responsibility for the proper use of research funds is on the researcher, and this is overseen by funders during and at the end of the proposal. Another recommendation refers to the use of state-of-the-art information as a basis for your research. The control system in this case is other scientists who read or evaluate your research, and who will recognize outdated research results.

Let's get back to the analogy of the mine for a moment. If you are paid to dig in the mine, you are expected to find important ore. In our case, a research funder is an employer, and the researchers are workers who need to go down the mine and get their hands dirty in the search for new true information. If you are set to dig a deep hole in the ground with the possibility of finding gold and diamonds, but you do not get any guarantee that you will find them unless you chose an appropriate place in a specific period, you would probably spend a lot of time planning and trying to decide where to start digging, what to do when specific problems arise and to avoid ending with a huge number of worthless rocks instead of gold and diamonds. The process is similar to research planning since a significant amount of the research process can be defined before data collection begins. As valuable as it can be, a research idea is just a thought which needs to be translated into research practice to gain its full impact.

How to Formulate a Good Research Question?

Research is performed to answer a specific question. The research process can be observed as a complex tool that, if used properly, can give a clear answer to a posed question. The research question is the compass of the research process (or the mine if we continue with our mine analogy) since it determines the steps of the research process. It translates into specific research aims and, consequently, into testable research hypotheses. Formulation of a research question is a skill that develops over time, a skill that can be learned. Your research question should have a FINER structure, which stands for: **F**easible, **I**nteresting, **N**ovel, **E**thical and **R**elevant.

Although initially developed as a set of recommendations for quantitative research, FINER recommendations can be applied to formulating a research question in any given field of science.

The feasibility of a research aim is often defined by time restrictions and funding because research is often burdened by deadlines and output requirements set by the funders. **F**easibility is also affected by the availability of technology, geographical restrictions, availability of participants, or availability of collaborators. If one considers all those factors, it is obvious that research interests play only a small part in the formulation of a research question. Ask yourself: What research can be published in an excellent journal if you have limited funds and only 1 year for research, with limited access to a specific technology? (Today, highly specialized experts may be a greater problem than the technology in question). You might experience that the formulation of the research question is mostly defined by non-research factors, because, in the end, it is better to have a completed than never-finished research.

There are other elements of the research question that are as important as feasibility. The first one to consider is **E**thics, which affects all parts of the research process due to its broad nature. If research is not ethical, then it should not be conducted. In a mining analogy, ethics is training and safety, which helps you to protect others and yourself during the entire process. To get back to the best research practices, researchers should make proper use of research funds and fulfill the basic research aim – the benefit to society. This also implies treating members of that society with respect, respecting their privacy and dignity, and being honest and transparent about the research process and results. Therefore, when determining the feasibility of a research study, ethics aspects are the first to consider, along with the objective factors of time, cost, and manpower.

Interest, **N**ovelty, and **R**elevance from the FINER guidance are the elements of the research question that increase the chances of getting funding or the chances for a journal publication, and they are closely aligned. Regardless of the audience (researchers, publishers, non-experts), research should be new to be interesting and relevant. However, doing research just for the novelty's sake is analogous to the digger who starts digging a new mine every couple of days. It gives you the thrill of a new beginning, but you have not dug deep enough to get to the real results. Relevance, defined in this context as a significant add-on to the current knowledge, can be assessed with a high probability of success by a thorough search for available evidence. The main aim of that process is to identify research or practice gaps that can be filled to improve general knowledge.

Interest is related to the principal internal motivation of an individual to pursue research goals. The interest to pursue research aims is difficult to assess. When planning research, do you consider that research is interesting to you, your peers, potential users, or all three? Probably the last, but here is the catch. Interest is the most subjective part of research planning. Research planning could be understood as a balance between your interest and all other factors that affect the research outcome.

A good research idea is often the compromise between objective possibilities and a desire to make a research discovery. If the research idea is interesting but extremely difficult (or even impossible) to conduct in given circumstances, you will end up frustrated. On the other hand, if you decide to perform research based solely on convenience (because it is something for which is easy to get funded or someone is offering you a research topic you are not interested in), it will be very difficult to stay motivated to complete the study.

The more structured your research question is, the easier it is to determine which research design is best to test the hypothesis and statistical analysis is more straightforward. Let's look at several examples of research questions in biomedical research: Are psychedelics more effective in the treatment of psychosis than the standard treatment? What are the opinions of young fathers on exclusive breastfeeding of their spouses? Which percentage of the population has suffered from post-COVID-19 syndrome? Intuitively, for each of posed research questions, we would try to find answers differently. In cases of comparison of treatment methods and assessment of population percentage, we could express the results quantitatively, e.g., we could state explicitly how much the psychedelics treatment is better compared to standard methods in terms of days of remission or everyday functionality or an explicit number of people in the sample who had COVID-19-related symptoms. On the other hand, the answers to the question about the opinions of young fathers about exclusive breastfeeding are not straightforward or numerical, but more textual and descriptive. It is an example of the research question that would be more suitable for qualitative research. Qualitative and quantitative study designs answer different types of research questions and are therefore suitable for different situations. It is important to carefully consider and choose the most appropriate study design for your research question because only then can you get valid answers.

To conclude, research question development is the crucial factor in setting research direction. Although framed as a single sentence, it defines numerous parts of the research process, from research design to data analysis. On the other hand, non-research factors also have an equal role in research questions and need to be considered.

Literature Search

In a literature search, researchers go through the relevant information sources to systematically collect information, i.e. foreground knowledge, about a specific research phenomenon and/or procedure. While research information is readily available online not only to researchers but to the whole public, the skill of systematic literature search and critical appraisal of evidence is a specific research skill. A literature search is closely tied with the development of the research aim, because you may want to change it after you read about previous research.

When doing a literature search, you must be careful not to omit previous studies about the topic. Here we have two directions that must be balanced. The first one is to do a very precise search to find specific answers, and the other one is to perform a wide, sensitive search that will include many synonyms and combinations of words to discover articles that related to a specific term. Both of those approaches have their advantages and disadvantages: a precise search is less time-consuming and retrieves a small number of studies. However, it may omit important results, so you may end up performing studies for which we already have established conclusions. This creates waste in research because you will spend time and resources, and involve participants in unnecessary work, which would be unethical. You may also miss citing important studies. On the other hand, if you perform a search that is too wide, you will spend a lot of time filtering for useful articles, which leaves less time for doing research.

> **Good research practice from the European Code of Conduct for Research Integrity:**
> **Researchers design, carry out, analyze and document research in a careful and well-considered manner.**
> **Researchers report their results in a way that is compatible with the standards of the discipline and, where applicable, can be verified and reproduced.**

What Is the Optimal Study Design for My Research?

Study designs are one of the main heuristics related to the reader's perception of the credibility of research information. Also, different study designs give answers to different research questions. It is intuitively easy to understand that different approaches should be taken if the question is about the percentage of infected people in the population vs about which drug is the most effective in the treatment of the disease. The roughest categorization of the study designs is observational and experimental (Box 3.1). However, in different scientific areas, even that type of categorization is not enough, since study designs can be theoretical, as in physics or mathematics, or critical, as in humanities, and those types of research will not be covered in this chapter.

For some research areas (e.g. health sciences, social sciences), there is another type of research often referred to as evidence synthesis, or literature review. The literature review is a review of evidence-based on a formulated research question and elements. They differ in their scope and methodology (Box 3.2).

Box 3.1 Types of Study Designs
Observational study designs:
Case study/case series/qualitative study: All three types of study designs take
    into account a small number of participants and examine the phenomenon
    of interest in-depth but cannot make generalizations about the entire
    population.
Case-control study: Individuals with a certain outcome or disease are
    selected and then information is obtained on whether the subjects have
    been exposed to the factor under investigation more frequently than the
    carefully selected controls. This approach is quick and cost-effective in
    the determination of factors related to specific states (e.g., risk factors),
    but it relies too much on records and/or self-report, which may
    be biased.
Cross-sectional study: Best study design for determining the prevalence and
    examination of relationships between variables that exist in the population
    at a specific time. Although it is simple to perform, and relatively cheap, it
    is susceptible to various types of bias related to participant selection, recall
    bias, and potential differences in group sizes.
Cohort study: Participants are followed over a certain period (retrospec-
    tively or prospectively) and data are compared between exposed and
    unexposed groups to determine predictive factors for the phenomenon
    of interest.

Experimental study designs:
Randomized controlled trial (RCT): Participants are allocated to treatment or
    control groups using randomization procedures to test the strength of the
    interventions.
Quasi-experimental trial: Participants are allocated to treatment or control
    groups to test the strengths of the interventions, but there is no randomiza-
    tion procedure.

How to Assess which Study Design Is Most Suitable for Your Research Question?

Based on the research aim, one may already get a hint about which study design will
be applied, since different study designs give answers to different research ques-
tions. However, very often a research question is not so straightforward. Sometimes
the research aim could be to determine whether category X is superior to category
Y, related to the specific outcome. In those cases, one must determine what the core
outcome of the study is (e.g., testing of the effectiveness of two interventions, the
scores on current differences between two groups, or the changes over time between

Box 3.2 Most Common Types of Review

Systematic review: A type of review that searches systematically for, appraises, and synthesizes research evidence, often adhering to guidelines on the conduct of a review.

Scoping review: Type of review which serves as a preliminary assessment of the potential size and scope of available research literature to identify the nature and extent of research evidence (usually including ongoing research).

Meta-analysis: Statistical synthesis of the results from quantitative studies to provide a more precise effect of the results.

Rapid review: A type of review that assesses what is already known about a policy or practice issue, by using systematic review methods to quickly search and critically appraise existing research to inform practical steps.

Umbrella review: Specific type of review that searches and assesses compiling evidence from multiple reviews into one accessible and usable document. Focuses on broad conditions or problems for which there are competing interventions and highlights reviews that address these interventions and their results.

different groups), and then it is not difficult to determine the study type in question. In principle, a single research question can be answered with a single study design. However, what we can also use are substitute study designs that can give approximate answers to the question we are asking but will never give as clear an answer as the appropriate design. For example, if we want to explore the reasons early-career researchers seek training in research integrity using a survey approach, we could list all possible answers and say to participants to choose everything that applies to them. The more appropriate study design would be to use a qualitative approach instead because in the survey approach the assumption is that we already know most of the reasons. The survey approach gives us the answer which answer is the most frequent of all. It is a subtle, but important difference. Similarly, although we can test causation using a cohort approach, the evidence for causation is never strong enough in a cohort study as it would be in an experimental study, simply because in a cohort study the researcher does not have control over the independent variable. For example, if we would test the effects of alcohol uptake on the occurrence of cancer, we would compare participants who drink versus those who do not drink to determine the incidence of cancer and make the conclusion about the association between alcohol and cancer. However, the true study design for testing the causation is the randomized controlled trial, where participants are randomized into the interventional and control group, the researcher can give an exact amount of alcohol based on persons' weight, over a specific period, and in the end, compare the incidence between two groups. However, that type of study would not be an ethical

study, so it is not possible to do it. So, there are subtle, but important differences which answer whether can specific and good formulated research questions can be tested and answered fully with only one study design, but due to the various reasons (time restrictions, ethics, cost-benefit analysis) we often use substitute study designs.

> **Good research practice from the European Code of Conduct for Research Integrity:**
> **Researchers design, carry out, analyze and document research in a careful and well-considered manner.**

When describing people involved in the research process, researchers often refer to them as "participants" or "respondents" (in the case of surveys). A more precise term would be to name the group based on the population they are drawn from (children, people with specific diseases, or people from a specific geographical area). The appropriate term to use would be "participants", since people are willingly involved in the research process, and the generation of new findings depends on them. Being a participant in a research process means that a person has willingly entered into a research, without any real or imagined coercion, possesses respect and interest for the research topic, with the understanding that positive aspects of research findings encompass the research situation and contribute to general knowledge. This would be a definition of an ideal participant and the researcher should avoid a situation where the participants are coerced to enter research, whether by situational factors or personal reasons because that will probably result in a decrease in motivation for participation and lower quality of research findings. To act ethically and to improve the quality of the research you have to inform participants about the reasons for the study, its purpose, research procedure, their rights, and expected outcomes. A potential pitfall in the research process can happen if all information were not given to participants at the beginning of a research. On the other hand, if a participant enters willingly into the trial, but possesses no real interest in the research topic, it may also affect the motivation for participation in research, because those participants may consider the topic irrelevant and not take the research process seriously (it is easy to imagine a situation where teenagers in a classroom willingly decide to take the survey and participate in research about personality traits, but quickly lose interest after the second page of the questionnaire). All those things are not reflected in the research report but may have an enormous influence on the research findings. Therefore, it is important to define the population of interest and try to motivate participants by providing them with all information before the research begins. Some additional ways to increase participant retention are financial rewards or similar incentives. There are several sampling strategies used when approaching participants for a study (Box 3.3).

It is difficult to give clear criteria on when to stop collecting data. In the case of pre-registered studies, the stopping rule is defined in the protocol. Examples include time restrictions (e.g. 1 month), or the number of participants (e.g. after collecting

Box 3.3 Most Common Sampling Methods

Simple random sampling: Each member of the defined population has an equal chance of being included in the study. The sampling is often performed by a coin toss, throwing dice, or (most commonly) using a computer program.

Stratified random sampling: The population of interest is first divided into strata (subgroups) and then we perform random sampling from each subgroup. In this way, the sample with better reflects the target population in specific (relevant) characteristics.

Cluster random sampling: In cluster sampling, the parts of the population (subgroups) are used as sampling units instead of individuals.

Systematic sampling: Participants are selected by equal intervals set before the data collection begins (e.g., every third of every fifth participant who enters the hospital).

Convenience sampling: Participants are approached based on availability. This is perhaps the most common sampling method, especially for survey research.

Purposive sampling: This is the most common approach in qualitative study designs. Researchers choose participants (or they define their characteristics in detail), based on their needs since participants with those special characteristics are the research topic.

data on 100 participants). If the research protocol has not been pre-registered, then the stopping rule should be explained in detail in the publication, with reasons. In the latter case, it is never completely clear if the stopping happened after researchers encountered the desired result or if it has been planned. The practice of stopping after you collect sufficient data to support your desired hypothesis is highly unethical since it can lead to biased findings. Therefore, the best way of deciding to terminate the data collection is to pre-register your study, or at least define the desired number of participants by performing sample size calculation before the study begins and pre-registering your study. More about pre-registration and biases which it eliminates will be said later in the chapter.

Ethics of the Sample Size: Too Small and Too Big Samples

A common problem in sampling is that researchers often determine the desired number of participants in a study. The problem is that the response rate is always lower than 100% (in survey research it is often around 15–20%), and a certain percentage of participants drops out of research, resulting in a sample size significantly lower than initially planned. The sample used in research can be too small, and there

is a possibility that you will not find a true effect between groups, and in that case, you would make a type II error. The reason is that in small-scale studies the error margin is big, and you would need an extremely large effect size to reach statistical significance. On the other hand, in cases of a big sample, the problem is different. If you have big samples, even small effects will be statistically significant, but the effect size may be negligible. The reason is that within big samples, the margin of error is small, and consequently, every difference is statistically significant. Once again, the proper solution (practically and ethically) for this issue is to calculate the minimum sample size needed to determine the desired difference between groups to avoid the issues with small samples and report effect sizes also, to avoid issues related to (too) big samples.

What We Can and What Cannot Measure?

When it comes to measuring in research, that part is mostly associated with statistical analysis of research data. The principal thing in statistical analysis is to determine the nature of the main outcome variable. In qualitative research (e.g. interview, focus group) or a systematic review without meta-analysis, statistical analysis is not necessary. On the other hand, for quantitative studies (a term often used for mostly case-control, cross-sectional, cohort, and interventional studies) the most important part of the research plan is to define the outcome which can be measured.

In general, there are two types of variables: qualitative and quantitative. When it comes to statistical analysis of qualitative variables (in a statistical context you will encounter the terms nominal and ordinal variables), we can do only basic functions, like counting and comparing the proportions between different groups, but we are not able to calculate mean or standard deviation, because those variables do not possess numerical characteristics. Examples of qualitative variables in research can be the number of surviving patients in a group at the end of the trial, self-reported socioeconomic status as a demographic characteristic, or any binary (yes/no) question in a questionnaire. In some cases, qualitative variables may be coded with numbers, but that does not make them quantitative. A good example is jersey numbers where numbers serve only as a label and not as a measure of quantity (e.g. if you have team player numbers 2, 4, 6, you probably will not state that the average jersey number is 4 because the very concept of the "average" jersey number is absurd). On the other hand, for quantitative variables, differences between numbers indicate the differences in value (e.g. if you say that person X is 1.80 m high, you know that that person is taller than person Y who is 1.70 m tall). You can also calculate different statistical parameters, like mean and median, and dispersion measures, which gives you a more flexible approach in the choice of statistical tests, especially those tests for differences between groups. On the other hand, applying statistical tests would mean that you are more

familiar with statistics, which sometimes may present a problem for less (and more) experienced researchers.

When Is the Time to Consult with a Statistician (and Do You Have to)?

Some (lucky) researchers possess sufficient knowledge to perform data analysis themselves. They usually do not need to rely on somebody else to do the statistical analysis for their study. For everybody else, statistical analysis is a crossroad where one needs to decide on including a person with statistical knowledge in a research team or to learn statistical analyses by themselves. The usual process is that the research team defines the research aim, spends time collecting data, collects data, and then tries to find a statistician who will analyse the data. If we keep in mind that research often has high stakes (e.g. doctoral diploma) and researchers are under a great time and financial pressure, the decision to include a statistician is sound and logical, but is it really necessary? The inclusion of a statistician in research when the data are already collected is similar to the situation when you give a cook an already finished stew and ask him/her how it can be improved. The cook may help with the decorations and give some spice which would make the food look and taste better but cannot change the essence of the food since it is already cooked. It is the same with data. The golden rule of statistics is "garbage in, garbage out", referring to a situation where poorly collected data or data of poor quality will give rise to wrong conclusions. Researchers should know statistics, not only because of the statistical analysis but because statistical reasoning is important in the formulation of measurable research aims. Therefore, statistical analysis is an important part of responsible research and begins with the formulation of the research aim. Statistical experts should be included in the study at that point.

Statisticians usually analyse data based on the initially set research aim. They send back the results of the data analysis to the research team, and they all together (in an ideal scenario) write the manuscript. The dataset remains in the possession of the principal researcher and the paper is published in a journal. Many journals and funders require that the data are publicly available so that anyone can use it, respecting the FAIR principles. Keeping that in mind, the process when somebody else is doing statistical analysis for you requires an enormous level of trust for statisticians, because they can do analysis wrong but you may never know it. Unless, of course, someone else analyses publicly available data and sees the error. In that case, you are also responsible for the analysis because it does not matter that you did not perform it. In some cases, this may lead to the retraction of the paper, which consequently may lead to certain consequences for you (especially if the articles are the basis for a doctoral thesis). If you are willing to put trust in someone to do data analysis, that is perfectly fine, just be aware of this risk, and remember that people make mistakes, very often unintentionally, and therefore a double check by a third party would be recommended.

On the other hand, if you are willing to learn how to do statistical analysis, the good news is that today there are lots of resources to help you. The first thing about statistics you need to know is that you do not need to know all statistics to do statistics. The only knowledge about statistics and statistical programs you need is the one that would help you do the analysis of your research aim and test the research hypothesis. To do that, you will have to see the data you have and search online for ways to analyze a specific problem. You can use tutorials of the statistical program that simultaneously give instructions about the statistical principles and procedures for analysis. Today, most of those programs have online videos and detailed tutorials. Some of those programs are user-friendly and free (e.g., JAMOVI or JASP), some are commercial (e.g., SPSS, Statistica), and some are less user-friendly but free and available (e.g., R programming language). If you are a beginner, use a more user-friendly program that has detailed instructions and try to do the statistical analysis by yourself. It is expected that you will make errors, so it would be good if someone more experienced looked at the results and provides feedback on your first attempts.

There are many tutorials on how to do statistical analysis, but far less on how to do proper data entry, which is the preparation of data for statistical analysis. Usually, the data entry table is made in a computer program that provides a tabular view of the data (e.g., Microsoft Excel). The golden rule is that each column represents a variable collected in research, by the order it was collected in the research and that each row represents the unit of the analysis (usually participant, text, article, or any other unit). In a separate sheet or a document, there should be a codebook that contains information about each level of each variable in the dataset, in a way that a person who is not familiar with research can understand the nature of the variable. The codebook should always accompany the dataset, so if the dataset is shared publicly, the codebook should also be shared. The rule of thumb for the data entry is that textual variables are entered as texts and quantitative variables as numbers, and textual variables can later be coded with numbers if necessary. The table for data entry should be made before the research begins, and it is good to seek help from a statistician when defining that, too.

> **Good research practice from the European Code of Conduct for Research Integrity:**
> **Researchers publish results and interpretations of research in an open, honest, transparent, and accurate manner, and respect the confidentiality of data or findings when legitimately required to do so.**

Preregistration of Research Findings

Pre-registration refers to the presentation of the research plan before the research begins. This process serves as the quality control mechanism because it prevents a change in the research hypothesis and methodology to fit the data collected.

Pre-registration of research findings should be done after the research has been approved by the ethics committee. There are various registries, some of which are more discipline-specific (e.g., ClinicalTrials.gov for clinical studies) while others are open to different disciplines and study designs (e.g., Open Science Framework). For the pre-registration of a study, one should clearly define all steps related to the research aim, methods, planned analysis, and planned use of data. Pre-registration of data is nothing more than the public sharing of a research plan. However, even that relatively simple procedure helps eliminate specific biases and decreases the probability of unethical behavior. Pre-registration eliminates the problem of hypothesizing after the results are known (so-called HARKing) because you need to state your hypothesis publicly before the research begins. Pre-registration should be done before the actual research begins, since you may have already collected the data and modified your hypothesis so that it fits your data (this is called PARKing –preregistering after the results are known), which should be avoided since it is not a true pre-registration.

Why is pre-registration good for research? When a study is pre-registered, researchers will follow the research plan and planned analysis and will not alter the study protocol and statistical analysis unless there is a valid and strong reason for protocol modification. Many journals today require that studies are pre-registered and that research data are shared. It is recommended to pre-register not only the study aim and methods, planned analysis, but also planned impact, data use, and authorship. When pre-registering authorship, you make clear from the beginning of the study the roles and expectations of each member of the research team. If during the research process some changes happen with the study protocol, those should be clearly explained and pointed out in the final publication, because deviations from the protocol can sometimes bring suspicion in the interpretation of the results if they are not reported. Pre-registration can be peer-reviewed and some problems, which would affect the final interpretation of the results, can be addressed even before the study begins. Finally, when pre-registered, you have the evidence that it was you who came up first with a specific research idea.

One problem that pre-registration cannot prevent is research spin or exaggeration in the scope of study results. Even if data have been carefully collected and properly analyzed, the interpretation of the results is up to the researcher. You should be honest (and modest) when interpreting the results of your study, by stating the true magnitude of your results and putting them in the context of the previous studies.

After the research has been published, the data used in research should be made available to everyone who wants to use them, since data sharing helps research replication and evidence synthesis. You can read more about data sharing in the chapter on Data Management and the chapter on Publication and Dissemination.

With this knowledge in mind, how would you improve the research procedure from the case scenario at the beginning of this chapter?

If You Want to Learn More

The Embassy of Good Science

Replicability
AllTrialscampaign:https://embassy.science/wiki/Theme:0bb5e4f7-9336-4ca8-92e3-c506413d1450
Forensic statistics to detect data fabrication: https://embassy.science/wiki/Theme:467f5cf6-d41f-42a0-9b19-76556579845d
Pre-registration of animal study protocols
Prospective registration of clinical trials
Statistical pre-registration
Data driven hypothesis without disclosure ("HARKing")
Insufficiently reported study flaws and limitations
Spin of research results.

Published Articles

Cooke A, Smith D, Booth A (2012) Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 22(10):1435–1443. https://doi.org/10.1177/1049732312452938

Grant MJ, Booth A (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J 26(2):91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Kerr NL (1998) HARKing: hypothesizing after the results are known. Personal Soc Psychol Rev 2(3):196–217. https://doi.org/10.1207/s15327957pspr0203_4

Lipowski EE (2008) Developing great research questions. Am J Health Syst Pharm 65(17):1667–1670. https://doi.org/10.2146/ajhp070276

Onwuegbuzie AJ, Leech NL (2006) Linking research questions to mixed methods data analysis procedures. Qual Rep 11(3):474–498

Rau JL (2004) Searching the literature and selecting the right references. Respir Care 49(10):1242–1245

Richardson WS, Wilson MC, Nishikawa J, Hayward RS (1995) The well-built clinical question: a key to evidence-based decisions. ACP J Club 123(3):A12–A13

Yamada Y (2018) How to crack pre-registration: toward transparent and open science. Front Psychol 9:1831. https://doi.org/10.3389/fpsyg.2018.01831

Guidance

Hulley SB. Designing clinical research: Lippincott Williams & Wilkins; 2007.
University of Oxford. The Centre for Evidence-Based Medicine.
HealthKnowledge.

Cummings SR, Browner WS, Hulley SB. Conceiving the research question and developing the study plan. In: Designing Clinical Research. 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2013. p. 14–22.