3

How we Conduct Research

Contents

> 'What is truth? The negation of lies? Or the statement of a fact? And if the fact is a lie, what then is the truth?'
>
> Yennefer, in Andrej Sapkowski, *Sword of Destiny*

Introduction

Conducting research is simple, right? You just choose a topic and get going. Well, sort of. There are lots of smaller steps, and getting this right is one of the most important things we will cover. Getting the research design correct means that the data-handling and statistics portion can become far easier. We want to determine which statistical methods we can use (or if we do not need to use them at all!) to answer the questions we have about our world. This next point is one I want you to embed:

> It is the questions we are asking of our data that are important. We *never* approach a piece of research with 'I want to run a particular test'.

In Chapter 1, we covered the two main types of research – incremental and explorative research. Now, we specify *how* we conduct that research using the scientific method, be it hypothesis or research question driven.

Chapter Objectives

- To consider the questions we can ask using statistical analysis
- To learn how we structure research
- To introduce the concept of statistical significance
- To consider how to conduct research ethically

The Questions we can ask Using Statistics

It is important to start with the questions we can ask of our datasets before progressing into our statistical testing. Generally, the questions we can ask of our data revolve around four areas in Figure 3.1:

- Differences
- Association
- Relationships
- Patterns:
    - in opinions
    - in time (temporal patterns)
    - in space (spatial patterns)

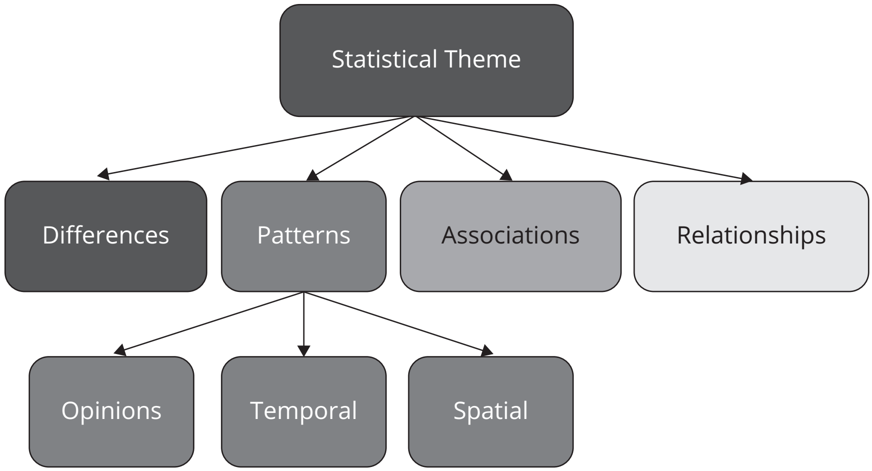

Figure 3.1 The questions we can begin to answer by using statistics in our research

In each of the above instances, there are some go-to statistical tests we can use. These aren't exclusive, but they are the major tests you will come across throughout your exploration of social science research. This book will focus on these, when is the correct time to use them and how we then interpret the results. In summary:

- Are there any differences between different variables?
- Do they have a relationship (can we estimate one from the other)?
- Are they simply associated (linked), or do they have any patterns, including temporally or spatially?
- How can we begin to investigate our theories by using these questions and associated analytical or statistical techniques?

Differences

One of the most common questions we can ask of our data is whether there is any difference in one group of observations compared to another group of observations. Perhaps we are interested in whether people earn different amounts of money in two regions. This question could be posed over a range of scenarios, including spatial scales, for instance at the continent or country level. We could ask whether there is a difference between average income in South America and Africa. When we start thinking about differences, our statistical tests fall into a group of 'significant difference' tests, such as ***t*-tests** and **ANOVA** (covered in Chapter 9).

Associations and Relationships

We could then think about adding additional variables to our analysis of income that allows us to ask a different set of questions – on **associations** and **relationships**. We could start looking at associations between average income in each country and other parameters such as infant **mortality rate**. An association tells us that there is some statistical link between the data but not about the underlying cause, which could be related to a different variable entirely. This is something we will cover in detail when we consider tests of **correlation** (Chapter 10), our measure of association, remembering that correlation does not equal causation! We can then move on using the same variables or more to think about relationships, where we can use **regression** (Chapter 11) to build predictive statistical models helping us understand, for example, the degree of influence there may be on income and infant mortality rate.

Patterns

Our final set of questions relate to patterns we can find in data. Perhaps there are patterns in opinions – for example, what is more important, the environment or economic stability? Are there changes in opinions based on which political party you vote for? We can begin to use tests like **chi-square** (Chapter 12) to look for patterns in frequencies of datasets. Questions can begin to link together – for example, are there patterns spatially, do people in a certain region favour the importance of the economy over stability? Have people's opinions on this important issue changed through time? You can see how very quickly we can ask a range of different questions using similar data – if the data exists of course. Spatial techniques can include investigation of **clustering** (Chapter 13) – are there an increased number of cancer cases near nuclear power stations? The statistics we can conduct with temporal data can range from regression to look at trends through to **autocorrelation** to look at cycles in data (Chapter 14). We can also go from spatial to temporal datasets very quickly – taking satellite data to quantify the urban spread of cities, for example.

How We Conduct Research – To Hypothesise or Explore?

Anytime you read a report or academic journal article one of the first things you will read is the introduction – occasionally also referred to as the background or rationale. We must first highlight what we already know about a subject and what others have already covered before we address the main **research question** or questions that will be investigated. This is a vital part of the scientific method (Figure 3.2) that, despite being developed centuries ago, is still a solid framework for research design.

Let us follow the scientific method adjusted for the social sciences, to see how things should be done:

1. *Initial idea (theory)* – is the idea we want to test grounded in some understanding of the world around us? If yes, then we can begin to ask questions or hypothesise. In this step, we will need to conduct background reading to verify that our idea is sound.
2. *Research questions and/or hypotheses* – what do we specifically want to test about our theory? We design research questions and/or associated hypotheses that we will use in

our statistical testing. Sometimes we will just pose questions of data as we just don't know what we will discover.

3 *Research design* – how will we test our hypothesis or hypotheses? We will do this through designing an experiment, we use our *when, where, what, and who* to help.
4 *Get* **ethics** *approval* – is your research ethically sound (see the next section)?
5 *Conduct research* – we have come up with an effective research plan during our design process, so now we can go out and conduct our research if it's primary analysis or obtain our data from any secondary sources.
6 *Data handling and statistical analysis* – we safely store data, interrogate it using data presentation methods and decide whether we need to perform statistics or other data-processing techniques on it to help us assess and evaluate our hypotheses.
7 *Revisit theory and research questions* – possibly the penultimate step, we evaluate our statistical results considering our theory. Does everything add up? Do we have the necessary statistical significance, if applicable? If our theory holds up, we can report this; if not, we may need to think again and revisit our initial idea!
8 *Make conclusions* – draw all the evidence together in a summary, balancing with what other people have also discovered on similar themes – do you agree or disagree? Why might that be the case?

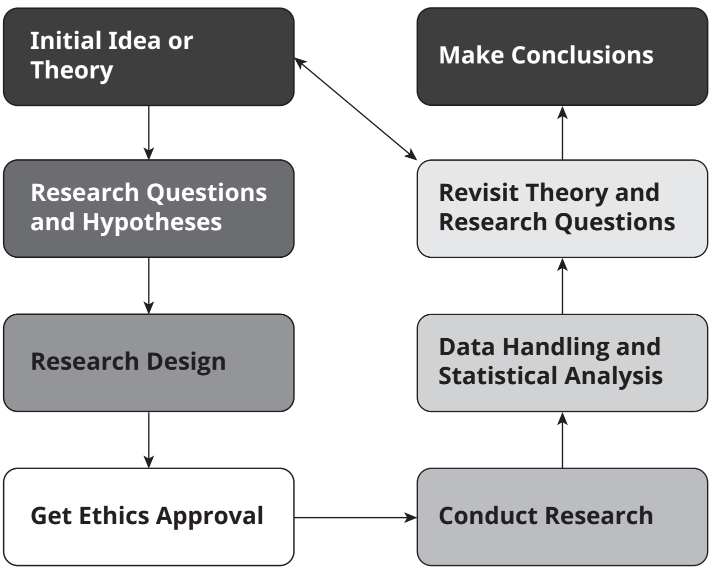

Figure 3.2 The steps of the scientific method, adapted to include crucial elements of social science research, such as ethical approval

You may have also heard of the problem, plan, data, analysis, and conclusion framework which is particularly popular in data science (Wolff et al., 2016). The difference with our social science framework is that we have additional important steps in ethics approval that require us to have a thorough plan in place for approval, outlined in Figure 3.2.

Example 3.1: Framing a Hypothesis

Let us consider mental well-being – we have a **theory** that participating in gardening improves our mental well-being. This theory will be based on something we know about our world, based on some existing research, and passes the common-sense test. The example is taken from research by Fjaestad et al. (2023) who surveyed 4,919 adults aged 46–80 years. We have our theory. We then come up with our **hypothesis**:

> Those who garden have an improved mental well-being over those who do not.

In general, a hypothesis describes a possible link, association, or relationship between two (or occasionally more) variables. Hypotheses are usually then directly testable using statistical analysis, related to our questions of difference, associations, relationships, and patterns. In the meantime, we try and keep our hypotheses as simple as possible. If a hypothesis is too complicated, then it will be challenging to assess. For example, we would not frame a hypothesis like this: 'Those who garden have an improved mental well-being if they conduct more than 2 hours per week and are aged over 63.' This is framed much more like the end results of a study, which would be based on the results of multiple tests and hence assessed using multiple hypotheses.

To help assess the results of any study, we use at least two hypotheses:

- The null hypotheses (H_0).
- The research (or alternative) hypothesis (H_1).

The **null hypothesis**, commonly denoted by H_0, is our default position. We always assume that there is no difference, association, relationship, or pattern, unless we can prove otherwise. If we do prove otherwise, through statistical analysis, then we can generally accept our **research hypothesis**, generally that there is a difference, association, relationship, or pattern (Glass and Hall, 2008). For our mental well-being example this would be:

- H_0 = Those who garden don't have an improved mental well-being.
- H_1 = Those who garden do have an improved mental well-being.

We could then build on these with more specific questions we ask of data:

- H_0 = There is no significant difference in the effect of gardening on mental well-being between those aged over 63 and those under 63.
- H_1 = There is a significant difference in the effect of gardening on mental well-being between those aged over 63 and those under 63.

We might test the above scenario using our significant difference tests that are looking for whether one group has a measurable difference in well-being compared to one or more other groups (Chapter 9). We must then perform multiple experiments and studies to prove that gardening and mental well-being are not linked. This might seem slightly counter-intuitive, but the only way of proving our ideas is repeatedly proving that the null is not the case. We therefore tend to accept our research hypothesis (H_1) and reject our null hypotheses (H_0). We may, of course, need to alter or amend our theory as new discoveries are made. This is the nature of scientific progress. For example, we know the importance of gardening in mental well-being but see an improved result in those of an older age. In the example from Fjaestad et al. (2023), after going through the steps of research design, ethics, conducting the research, statistical analysis, and finally evaluating their theory, they discovered that 'those who gardened for ≥150 min per week were more likely to report better mental well-being', noting that 'these effects were stronger for participants aged 64 years and older'.

Accepting or Rejecting Hypotheses

We now need to turn to wrongly accepting or rejecting a hypothesis based on the results of a study. We have a theory:

> All mountains above 3 km in height have snow on them.

For this theory we could create two separate hypotheses:

- H_0 = There are no mountains above 3 km in height that have snow on them.
- H_1 = All mountains above 3 km in height have snow on them.

To test our associated hypotheses, we design an experiment that samples 20 mountains worldwide. All but one of these have snow on them (Figure 3.3). We now have three possible conclusions:

- Our theory is incorrect, it is not true that there are no mountains above 3 km in height that have snow on them – we falsify the theory (Popper, 1963) and create a new one.
- The one mountain has snow on it, it was just covered by wind-blown material. We accept H_1.
- The mountain is not a mountain, it is just a mounded hill above 3 km in height, in which case we accept H_1.

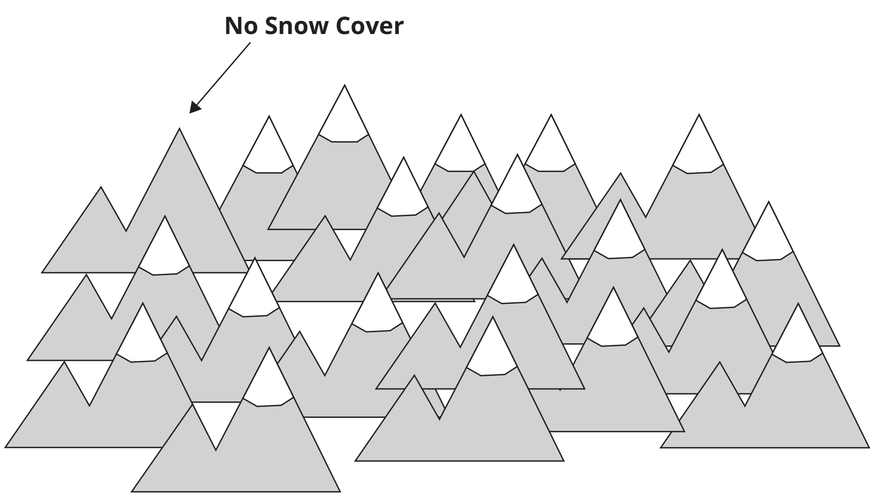

Figure 3.3 A visualisation of 20 sampled mountains for snow cover of which only one has no snow cover

But this problem becomes very complex if we break down some questions:

- What constitutes snow cover? Is this an area-based definition, for example 1 km^2 of snow, or does it just mean the presence of any snow regardless of area covered?
- What was the definition of a mountain used?
- Is a sample of 20 mountains appropriate?
- What time of year were measurements taken? Was it the same season in each case? Think about different hemispheres here!
- Was our theory a reasonable one in the first place?

The nature of hypotheses can be different from discipline to discipline (Donovan et al., 2015). One specific solution to a very hypothesis-driven way of thinking, that is being adapted given the increasingly negative outlook on using significance in our research that tends to naturally come with hypothesis testing (e.g., Amrhein et al., 2017), is through the usage and design of our studies around initial research questions. Suppose instead we asked:

Is there a link between mountain height and snow cover?

We can then avoid the multitude of research design sins that can occur when solely using hypotheses. One of these is creating expectations of what we may find and then becoming disappointed when we don't find it. Emotional investment in our results might sound strange, but tunnel vision when it comes to research is real. None of this excludes the fact that it's a bit of a silly research question in the first place. There is therefore a balance between using hypotheses and research questions – a balance which can court quite a bit of controversy (Gorard, 2016, 2017; Spreckelsen and Van Der Horst, 2016; Nicholson and McCusker, 2016).

If we were to turn our snowy mountain problem into a rigorous study, we would reframe the whole idea around the following structure:

An aim:

- To investigate the conditions that promote snow on mountains.

Research questions:

- Is altitude a contributing factor to snow cover?
- Is the position of the mountain important?

Note that it can sometimes be beneficial to combine research questions and hypotheses, but not always. You should ask yourself, does this help me with my aim?

During the evaluation process when we may want to reconsider the theory, we want to be careful that we do not overcomplicate things. There is a specific term for this that you have probably heard of: **Occam's razor**, whereby the simplest solution is probably the correct one. Taken at face value, if we have multiple possible explanations for a particular result, we take the one that is simplest or has the fewest possible variables to explain, which is also known as the **principle of parsimony** (Sober, 1981). This is a good up to a point. Unfortunately, we do sometimes need complicated theories to fit the facts. This is why we invoke statistics and our next subtheme, *significance*, within our work. Statistics essentially give us a mathematical way of saying:

> Yes, we have enough proof to accept this particular theory

or

> No, we don't yet have enough proof to accept this particular theory.

But as we have seen, this is not always enough. We need to keep throwing numerous studies at our theories and ideas! When designing a study and incorporating the effective use of statistics you need to (remember the four 'U's!):

- Always start from a hunch
- Decide on the variables you need to assess your theory and gather the relevant data carefully (understand your dataset)
- Choose the correct methods for the job (understand your methods and statistical analysis)
- Interpret the results carefully (understand what this tells you)

The best research should always be grounded in something that we have some understanding of already. As a psychologist, you would not then go and study particle physics, it would probably not end well!

Designing Ethical Research and Asking Appropriate Questions

Whenever we conduct research, whether it's an undergraduate dissertation or multi-million-pound study, we our bound by ethical obligations. Before conducting any research involving human participants or anything that could have any impact on a community or group of individuals, however small that impact may be, we must get ethical approval for our research (Israel, 2014). At universities and at any large organisation, dedicated teams will assess and help ensure that studies uphold the highest level of ethical principles.

Before we assess some of the common ethical considerations, we must first ask ourselves whether we should be asking the question in the first place. Here is an extreme example: you should never assess the efficiency of torture in retrieving information from suspects. Ethically, harming someone in the course of research (and in general) is an obvious no.

Informed Consent

Some disciplines do then walk a fine line between ethics and making statistically rigorous discoveries, but for the right reasons. Clinical trials are a big part of this. It is common for large trials of new drugs, for example, to be split in half, where one set of participants are given the new drug and the other a placebo. Is it ethical to withhold a potentially life-saving drug from half the participants? In this instance, such a setup is vital to control for the 'placebo effect' (Finniss et al., 2010). Is the result of improvements in health a result of the drug or something else? When asking people to agree to take part in a trial it is important to explain the purpose of the study to them and to make them aware that they may be taking either a placebo or the new drug – this is **informed consent** (Aldridge, 2014; Israel, 2014).

Vulnerable Groups

As researchers we should be aware of who the vulnerable groups are. Children are always seen as vulnerable. Refugees, dissidents, sex workers, and those who are unable to give informed consent are also considered vulnerable. When including vulnerable groups of individuals in our research, it is important that we consider whether we could obtain the same research outcomes on different groups of non-vulnerable people. Our research should never prey on that vulnerability or make it worse in any way (Aldrige, 2014, 2015; Limantė and Tereškinas, 2022; Kluczyńska et al., 2024).

With some research into vulnerable groups, such as sex workers or people who use illicit drugs, it may be tempting to attempt **covert research**. This is where the participant is unaware of the research and has not given informed consent. Not only do you put the participant in danger – even the most anonymised of datasets can reveal characteristics of individuals or groups of individuals – but also, if observing criminal activity, there may be an obligation on the researcher to report said crime. That being said, the usage and issues of covert research remain debated (Homan, 1980; Herrera, 2003; Calvey, 2008; Spicker, 2011).

Social Media

It can be very tempting to use data that is easily accessible and freely available on social media. Unfortunately, it can be very easy to disassociate between data held on social media and the real people behind them. Even the most carefully designed studies that may paraphrase content could lead to identifiable traits. Quite simply, just because the data is there, it does not mean that it should be used. However, we should also acknowledge that social media resources are huge troves of information – they are sources of 'big data'. Again, we should be asking ourselves the key questions, is there a risk of individuals being identified and a risk of harm? Was the information intended for a private audience? Such questions become very blurry when using open comment type sites such as X (Townsend and Wallace, 2017; Taylor and Pagliari, 2018; Chen and Quan-Haase, 2020).

Data Management

Within all of the above points, appropriate data management is important. Participants must give consent to data being stored, which must not be indefinitely and must be securely. As a final note, it is never OK to change or deviate from your approved ethics plan – even the slightest change could have dramatic implications and should go through ethics approval processes again. These procedures are put in place to protect researchers as much as they are the participants. When considering ethics and approval the main considerations are of informed consent, openness where appropriate, but vitally that no psychological, physiological or emotional harm could occur to the participant or indeed the researcher (Israel, 2014).

How Important is Statistical Significance?

The ***p*-value** or **probability value** and significance are one of the major pillars of data analysis and statistics which we unravel in detail in Chapter 7. We have some statistical result, but does it actually mean anything? This is what significance and the p-value try to help us with, but it is fraught with misunderstanding. Many statistical tests come with associated levels of significance provided in the form of a p-value, which in the social sciences is commonly set at $p = 0.05$. If we run a test and get a p-value which is lower than or equal to $p = 0.05$ we can reject our null hypothesis; however, the $p = 0.05$ level is somewhat arbitrary. We are going to cover this more than once in this book, because it is important that we get this right! Essentially, we want to be confident that we correctly accept or reject a (null) hypothesis. We want to be certain that our result is *significant*. In most social science studies, we should now start to specifically use the phrase *statistically significant*, as realistically without this statistical backup the term 'significance' is meaningless. Overall, an emphasis on displaying your data graphically and appropriately can help interpretation (Loftus, 1993).

Example 3.2: *p*-Value Problem

The p-value can often be fraught with controversy through practices such as ***p*-hacking**. p-hacking is the process of changing or altering the way you analyse data to affect the resultant p-value that is produced with the goal of reducing it below $p = 0.05$. The end result is that a study can appear *statistically significant* even though it is not (Head et al., 2015; Wicherts et al., 2016; Stefan and Schönbrodt, 2023)!

Let us consider whether someone earns more or less than the average in an area. One person earns £27,000 a year and the average income in their area is £24,000. These two numbers are certainly different, but are they *significantly different*? We would need more context to answer that specific question, but what we need to avoid straight away is the use of this style of terminology without evidence. Another favourite of mine is the use of words like *extreme* or *extremely*. For example, 'the wages in Sheffield are extremely high'. What are you measuring this extremeness against? What period? One person's high is another's low – that phrase would therefore be understood differently by different groups.

Our concept of statistical significance is therefore important. We commonly assess whether something is statistically significant or not using the p-value – the probability value. In social sciences, you may commonly see the result of a test described as 'statistically significant at the $p = 0.05$ level'. This $p = 0.05$ means that we are essentially OK with a 5% chance that we have wrongly rejected our null hypothesis. Other disciplines will use other levels of significance, with particle physics requiring particularly low p-values to be absolutely certain the discovery is not some natural statistical variation. It is also advisable to report the exact p-value rather than 'the result was statistically significant at $p < 0.05$'.

The major criticism of the *p*-value is that it is based on a somewhat arbitrary level. If we strictly follow the *p*-value approach, we would accept the null hypothesis of no significant difference at $p = 0.051$ but reject it at $p = 0.049$. This is such a small difference in *p*-value, what actual difference in reality does this represent? Nevertheless, *p*-values continue to play an important role in our research to this day, particularly in the world of medical trials, although modifications are often made to avoid the small *p*-value change scenario. Some have even suggested changing the threshold to $p = 0.005$ (Benjamin and Berger, 2019; Ioannidis, 2019).

We always refer to the null hypothesis when statistically assessing our hypotheses – remember we are trying to continually prove ourselves wrong! It is OK if we cannot statistically accept our theories – this is part of the scientific method. This is a common issue for those who really want to find something significant in their results when actually the non-significant result can be more interesting. We also need to remember that just because we get a significant result it does not mean that we are correct – we should always refer to the theory. Sampling, error, and the *p*-value are something we will come back to throughout this book. Overall, the key to using *p*-values is to use them correctly and understand their actual value (Lakens, 2021). What is truth?

Summary

In this chapter, we have covered the core questions we can ask of social science datasets, including those of differences, associations, relationships, patterns in opinions, in space, and in time. We have outlined the scientific method for conducting research from idea to research design and back again. Our journey starts with the theory, our idea about how the world works. We want to know whether this a reasonable theory. Can it be tested? The best analysis comes up with an idea and does its utmost to prove it is wrong, using the scientific method. It is only then that we can begin to consider a theory as accepted, but we must continually test and address it potentially using statistical significance in association with probability values. We covered some of the key ethical issues that we can come across in social science research, highlighting the importance of informed consent. We finished with a word of caution around the usage and true meaning of *p*-values.

Key Take-Home Messages

1. We can ask statistical questions on differences, associations, and relationships.
2. We can ask statistical questions on patterns in opinions, time, and space.
3. We use the scientific method to design our research.
4. Hypotheses can help us assess the results of statistical analyses, particularly when used in conjunction with research questions.
5. Acquiring ethical approval is an important element of research.
6. *p*-values can be useful but should be used with caution.

Test Your Knowledge

Test your knowledge of Chapter 3 by answering these questions – all the information needed is either in the text directly or uses content that can be applied to solve a problem.

Q1: What are the four core statistics questions we can ask of our datasets?

Q2: There are eight key steps in our scientific method for the social sciences. Which step requires us to assess our initial idea or theory?

Q3: What is a hypothesis?

Q4: With reference to our key statistical questions, what does the null hypothesis mean?

Q5: Can we have more than one research (alternative) hypothesis?

Q6: Why might it be preferable to use research questions in lieu of hypotheses?

Q7: What is Occam's razor and how do we apply it in the social sciences?

Q8: When is it key to gain ethical approval of research?

Q9: What is the *p*-value and its purpose?

Q10: What is a major pitfall of the *p*-value approach?

Further Reading

On designing research in the social sciences:

Blair, G., Coppock, A. and Humphreys, M., 2023. *Research Design in the Social Sciences: Declaration, Diagnosis, and Redesign*. Princeton University Press.
Maggetti, M., Radaelli, C. and Gilardi, F., 2012. *Designing Research in the Social Sciences*. Sage .
Mertens, D.S. and Ginsberg, P.E., 2009. *The Handbook of Social Research Ethics*. Sage. https://doi.org/10.4135/9781483348971

On research ethics:

Israel, M., 2014. *Research Ethics and Integrity for Social Scientists*, 2nd edn. Sage.

On working with vulnerable groups:

Aldridge, J. 2015. *Participatory Research: Working with Vulnerable Groups in Research and Practice*. Policy Press. https://doi.org/10.1332/policypress/9781447305644.001.0001