

# Research Design

Master in Computational Social Science  
Universidad Carlos III de Madrid

Fall 2025

Francisco Villamil  
[francisco.villamil@uc3m.es](mailto:francisco.villamil@uc3m.es)  
Office: 18.2.A34 (Getafe)  
Office hours by appointment

**Tuesdays 18h-21h (see exceptions)**  
Room 2.A.04 (Puerta de Toledo)  
September 16th - October 20th

## 1 Description

This course provides an introduction to research design in the social sciences. The goal is to equip students with the skills necessary to evaluate and develop strategies to answer empirical questions with data. We will not cover statistical techniques or advanced data analysis. Rather, our focus will be on the logic of empirical comparison, particularly applied to quantitative data.

## 2 Requirements

We meet once a week for a total of seven sessions. During the first four weeks, we will review the main topics in research design. Every session will include a lecture followed by a discussion of a research paper. During our fifth session we will cover more advance topics, give an overview of the course, and reserve time for questions or to review some topics. During the last session we will have a workshop where students present ongoing work on the final essay and comment on each other's projects. Although the final essay can be submitted individually, presentations will be done in groups. Attendance to all sessions is mandatory.

## 3 Materials

This course does not follow any textbook in particular. However, there are some books that were used to design the course and can be very useful to expand on what is covered in class (books by Huntington-Klein and Cunningham are available online for free):

- Nick Huntington-Klein, *The Effect: An Introduction to Research Design and Causality* (Chapman and Hall/CRC, 2021).
- Kosuke Imai, *Quantitative Social Science: An Introduction* (Princeton UP, 2017).
- Dimitar Toshkov, *Research Design in Political Science* (Palgrave, 2016)
- Scott Cunningham, *Causal Inference: The Mixtape* (Yale University Press, 2021).

## 4 Assessment

### Participation (15%)

Every student is expected to be an active participant in all sessions, asking questions and engaging in discussions, including during the lectures.

### Research papers reviews (15%)

In sessions 2–4, we will discuss a research paper in the second half of class. Students are expected to bring in or send before class a short commentary (no summary, just a few discussion points). Each one is 5%.

### Workshop presentation (20%)

The last session (Oct 20th, 3–9pm) will be a workshop where students present in group an ongoing project, corresponding to their final essay. This project should present a research question that can be answered empirically with quantitative data and a strategy to answer it. Rather than focusing on data analysis, the focus should be on the type of variation that will be exploited and how well it answers the question. All the aspects that we covered in class should be discussed. A potential structure could be:

- Present a topic in general terms, why is it relevant?
- Specific research question
- Empirical strategy:
  - What type of variation is going to be exploited? Unit of analyses
  - Data: source, variables, measurement
  - Empirical comparisons
- External validity: how do the findings travel to other contexts?
- How much more of the original topic we now know?

### Workshop feedback (10%)

Student will also comment on each other's project, looking for limitations and possible ways forward.

### Final essay (40%) – **deadline: October 28th, 23.59h**

The main assignment is a written essay developing a research design. This assignment can be thought of as a pre-analysis plan for some study or an overview of how to explore quantitatively a given topic, e.g. comparing two different strategies for the same research question. Both individual and group essays are allowed (same grade for all group members).

## 5 Course outline

### Session 1: Introduction to research design

During the first session we will cover the basics of designing research and the reason why we need it. We will review what types of research there are, what is unique about quantitative research, how to develop research questions, and give an overview to all the steps taken in the research process in order to bring some answers to a given problem or question using empirical evidence.

*Reading:*

- Hannah Fry. [What Data Can't Do](#). *The New Yorker*, 29/03/2021.

### Session 2: Elements of quantitative research

We will focus on the basic elements of quantitative research. We will review the types of empirical relationships there are, the differences between descriptive and explanatory approaches, and the building blocks of empirical analyses, which cover both qualitative and quantitative research (e.g. units, variables, processes). We will also review the importance of theory and talk about how to build mechanisms at different levels. The final part will cover the importance of concepts, and both their operationalization and measurement.

*Reading:*

- Carl Müller-Crepon, Philipp Hunziker, and Lars-Erik Cederman. [Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa](#). *Journal of Conflict Resolution* 65(2–3): 563–590.

### Session 3: Understanding causality

We will introduce the concept of causality and explanatory designs more generally, covering the potential outcomes framework. We will review the experimental ideal in the natural and social sciences and the basics of how to get closer with observational data, including how to use DAGs. We will discuss the potential limitations of both experiments and observational designs, covering the most common biases.

*Reading:*

- Andrew M. Guess *et al.* [How do social media feed algorithms affect attitudes and behavior in an election campaign?](#) *Science* 381(6656): 398–404.

#### **Session 4: Causal identification with observational data**

In this session, we will cover the basic strategies used to infer causal relationships from observational data, building on what we discussed the previous week and without using statistical analyses. We will cover both controlling and matching as well as more advanced methods that exploit exogeneity, such as difference-in-differences, regression discontinuity, or instrumental variables. If time permits, we will review issues related to external validity of the results, problems with the unit of analyses (e.g. ecological fallacy), and unaccounted for processes in the data generation process, such as diffusion or unit interdependence.

*Reading:*

- Francisco Villamil and Laia Balcells (2021) *Do TJ policies cause backlash? Evidence from street name changes in Spain*. *Research & Politics* 8(4).

#### **Session 5: Advanced methods, overview, questions**

In this last session we will review the topics we did not have enough time to discuss in class, as well as presenting some more advanced topics, such as extra methods for causal inference, robustness and inference tests (e.g. placebo tests), recent approaches to external validity, etc. We will also give time for questions and, if time permits, we will cover the basics of quantitative research workflow. To allow for more time for questions, we do not discuss any paper this session.

#### **Session 6: Workshop (double session, 3-9pm)**

A 6-hour workshop where students present research designs in groups and give feedback to each other. We will have around 12-15 slots. Details will be discussed in class.