

GRAD-E1340: Causal Inference and Machine Learning*Concentration: Policy Analysis*

Dr. William Lowe

1. General information

Class time	Tuesday, 16-18h
Course Format	This course uses a “flipped classroom” format and combines 50 minutes of pre-recorded material (audio or video) with a 50-minute interactive seminar. Students will use the pre-recorded material to prepare for the seminar. The seminar is taught onsite at the Hertie School, or online via the platform Clickmeeting, depending upon your location. For those attending the online seminar, Clickmeeting allows for interactive, participatory seminar style teaching.
Instructor	Dr. William E. M. Lowe
Instructor’s office	3.14
Instructor’s e-mail	lowe@hertie-school.org
Assistant	Email: adjunctsupport@hertie-school.org
Instructor’s Office Hours	Students can make individual or group appointments by email.

Link to Module Handbook [MIA](#) and [MPP](#)Link to [Study, Examination and Admission Rules](#)Instructor Information:

Dr. William Lowe is a political methodologist with interests in text analysis, causal inference, and machine learning. He has a B.A. in Philosophy and an M.Sc. and Ph.D in Cognitive Science and Natural Language Processing, but joined the Government Department at Harvard as a postdoc and has worked in Political Science almost ever since.

Course Contents and Learning ObjectivesCourse contents:

This course introduces students to contemporary approaches to causal inference as it is applied to policy questions.

Main learning objectives:

Conceptual understanding of causal identification strategies in policy; the ability to criticize existing and proposed studies on causal grounds and where possible to suggest improvements; a clearer appreciation of the relationship between statistical estimation and causal identification, as they apply to traditional econometric, machine learning, and epidemiological tools.

Target group:

2nd year MIA and MPP students.

The target audience for this course is interested in honing their research design and statistical skills. There is no restriction on the substantive domain of application and students are encouraged, though not required, to come with their own substantive applications. The course mixes methodological instruction and discussion with intermittent practical applications. These will take the form of paper critiques, design exercises, and short reports requiring some R coding.

Teaching style:

The course will consist of short pre-recorded lectures from the instructor and live discussion of these conducted either in person, remotely, or a mix of both. There are regular exercises to be submitted the week after they are set, and students will finish the course with an in depth analysis of existing substantively-focused instructor approved paper to be chosen by them.

Prerequisites:

Statistically, students should be familiar with fitting and interpreting linear models. This course will treat similar topics to Statistics 2, so having taken that course will be an advantage, but will focus more on conceptual issues and on literature outside political science, e.g. machine learning.

Practically, students should be competent, though need not be expert, with the sampling functions and simple data.frame manipulations in R.

Diversity Statement:

Causal inference is a topic of intense policy and ethical relevance because, as this course will emphasize, it is a core issue in issues of social welfare, development, and fairness. Wherever possible we will connect the apparently abstract topics of potential outcomes and randomization to their concrete policy consequences. Perhaps surprisingly, causal inference is a useful tool for thinking about all kinds of diversity and its consequences. In particular, contemporary causal inference focuses on the methodological consequences of population variation and heterogeneity of effects and mechanisms.

Grading and Assignments

Composition of Final Grade:

Assignment 1: Data analysis exercise (x5)	Deadline: One week after handout	Submit via Moodle	50% (10% each)
Assignment 2: Research paper	Deadline: End of exam week	Submit via Moodle	40%
Participation grade (if applicable)			10%

Assignment Details

Assignment 1

Data analysis exercises are guided practical exercises based around a data set or text collection, interspersed with conceptual questions about the tools being used, interpretation of results, etc. The exercises are designed to improve practical skills and test knowledge from lectures and reading. Grading is based on success in the practical components and the quality of written answers. Answers

to conceptual questions are intended to require *at most* several paragraphs of text. Note that different exercises have a different mixes of practical and explanatory / analytical requirements.

Assignment 2

For the final paper, students choose a substantively oriented paper on a policy topic of their choice, subject to instructor approval, and provide both a critique of its methods from a causal perspective, and motivate a set of proposals to remedy any identified defects (if that is possible), e.g. by providing alternative or additional research design or other analytical strategies to address the original research question. Note that there is no requirement that suggestions must use exactly the same data but suggestions for alternative data sources must be reasonable.

Participation grade

The participation grade is based on the assumption that students take part, not as passive consumers of knowledge, but as active participants in the exchange, production, and critique of ideas—their own ideas and the ideas of others. Therefore, students should come to class not only having read and viewed the materials assigned for that day but also prepared to discuss the readings of the day and to contribute thoughtfully to the conversation. Some course materials can be difficult, not least because we will often be reading recently written articles close to the state of the art, without the benefit of year of expository practice. Consequently, it is important to identify and be honest about what you have understood and what remains unclear as you participate in the class. In particular, the instructor does not expect, although would be delighted by, a complete technical understanding of every paper before class.

Participation is graded subtractively; students receive the full grade except to the extent they fail to take adequate part in the class. Participation is marked by its active nature, its consistency, and its quality, but note that it is both unnecessary and also unwise, to monopolize conversation in order to maximize participation grade. Participation that makes it harder for other class members to engage in discussion will lead to a lower grade, regardless of the quality of interventions.

Late submission of assignments: For each day the assignment is turned in late, the grade will be reduced by 10% (e.g. submission two days after the deadline would result in 20% grade deduction).

Attendance: Students are expected to be present and prepared for every class session. Active participation during lectures and seminar discussions is essential. If unavoidable circumstances arise which prevent attendance or preparation, the instructor should be advised by email with as much advance notice as possible. Please note that students cannot miss more than two out of 12 course sessions. For further information please consult the [Examination Rules](#) §10.

Academic Integrity: The Hertie School is committed to the standards of good academic and ethical conduct. Any violation of these standards shall be subject to disciplinary action. Plagiarism, deceitful actions as well as free-riding in group work are not tolerated. See [Examination Rules](#) §16.

Compensation for Disadvantages: If a student furnishes evidence that he or she is not able to take an examination as required in whole or in part due to disability or permanent illness, the Examination Committee may upon written request approve learning accommodation(s). In this respect, the submission of adequate certificates may be required. See [Examination Rules](#) §14.

Extenuating circumstances: An extension can be granted due to extenuating circumstances (i.e., for reasons like illness, personal loss or hardship, or caring duties). In such cases, please contact the course instructors and the Examination Office *in advance* of the deadline.

2. General Readings

There are no required textbooks, but students may find the following useful for reference:

- Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Pearl, Judea, Glymour, Madelyn, and Jewell, Nicholas 2016. *Causal Inference in Statistics: A Primer* Wiley.
- Angrist, Joshua D., Pischke, Jörn-Steffen 2009 *Mostly Harmless Econometrics*. Princeton University Press

3. Session Overview

Session	Session Date	Session Title
1	08.09.2020	Foundations: Mechanisms and Counterfactuals
2	15.09.2020	Experiments, Quasi-experiments, and Definitely not Experiments
3	22.09.2020	Stratification and Regression
4	29.09.2020	Regression, Matching, Machine Learning
5	06.10.2020	Difference in Differences and Synthetic Controls
6	13.10.2020	Regression Discontinuity
Mid-term Exam Week: 19.10 - 23.10.2020 – no class		
7	27.10.2020	Collider Bias in Theory and Practice
8	03.11.2020	Mediation in Theory and Practice
9	10.11.2020	Fairness and Bias in Algorithms and Humans
10	17.11.2020	Fairness and Bias: Paper analysis
11	24.11.2020	Special topics: Sensitivity Analysis
12	01.12.2020	Special Topics: Alternative approaches to Causal Inference
Final Exam Week: 14.12 - 18.12.2020 – no class		

4. Course Sessions and Readings

Most readings will be accessible on the Moodle course site before semester start. In the case that there is a change in readings, students will be notified.

Required readings are to be read and analysed thoroughly. Optional readings are intended to broaden your knowledge in the respective area and it is highly recommended to at least skim them.

Session 1: Foundations: Mechanisms and Counterfactuals

Learning Objective

- Review of the counterfactual theory of causation, the relationship between DAGs and potential outcomes.

	<ul style="list-style-type: none"> - What estimands are interesting for policy? - What can be a cause? Non-manipulable 'treatments'
Required Readings	TBA
Optional Readings	

Session 2: Experiments, Quasi-experiments, and Definitely not Experiments

Learning Objective	Advantages and disadvantages of randomized controlled experiments as a research design ideal
Required Readings	TBA.
Optional Readings	

Session 3: Stratification and Regression

Learning Objective	<ul style="list-style-type: none"> - The relationship between stratification, regression models, and weighting - When can we interpret control variables? - Problems of overlap due to covariate sparsity and their solutions
Required Readings	TBA
Optional Readings	

Session 4: Stratification, Regression, Machine Learning

Learning Objective	<ul style="list-style-type: none"> - Would more flexible function estimators help? - Double ML: Machine learning tools for causally interpretable regression models
Required Readings	TBA
Optional Readings	

Session 5: Difference in Differences and Synthetic Controls

Learning Objective	<ul style="list-style-type: none"> - Difference in difference designs when policies are rolled out over time. - The parallel trends assumption in policy context
Required Readings	TBA
Optional Readings	

Session 6: Regression Discontinuity

Learning Objective	
Required Readings	TBA
Optional Readings	

Mid-term Exam Week: 19 – 23.10.2020 – no class

Session 7: Collider Bias in Theory and Practice

Learning Objective	<ul style="list-style-type: none">- Collider bias as a unifying framework for non-confounding inference problems- Collider bias examples in administrative data
Required Readings	TBA
Optional Readings	

Session 8: Mediation in Theory and Practice

Learning Objective	<ul style="list-style-type: none">- Classical versus modern mediation analysis- Collider problems in mediation analysis- What are we talking about? Mediation estimands in policy context
Required Readings	TBA
Optional Readings	

Session 9: Fairness and Bias in Algorithms and Humans

Learning Objective	<ul style="list-style-type: none">- Counterfactual theories of fairness- Compare and contrast implications from other theories- Legal issues in counterfactual fairness
Required Readings	TBA
Optional Readings	

Session 10: Fairness and Bias: Paper analysis

Learning Objective	Worked example of a policy issue involving all the course topics so far: fairness in policing.
Required Readings	TBA

Optional Readings	
-------------------	--

Session 11: Special topics: Sensitivity Analysis

Learning Objective	<ul style="list-style-type: none"> - How bad would unmeasured confounding have to be to negate your results? - How non-random can your sample be before standard statistical error margins cease to hold?
Required Readings	TBA
Optional Readings	

Session 12: Special Topics: Alternative approaches to Causal Inference

Learning Objective	<ul style="list-style-type: none"> - Alternatives to counterfactual theories: Probabilistic causality, QCA - Process tracing as a qualitative causal approach
Required Readings	TBA
Optional Readings	

Final Exam Week: 14 - 18.12.2020 – no class