

International Program Evaluation

Public Affairs 974 (3 Credits), Spring 2019

University of Wisconsin-Madison

Instructor: Christopher McKelvey

E-mail: cmckelvey@wisc.edu

Course Overview

This course provides an overview of program evaluation, the practice of assessing the efficacy of policy initiatives, with an emphasis on international applications. A primary concern when evaluating a program is ensuring that estimates reflect the causal impact of the program, so we will discuss various strategies for addressing this challenge, including: randomized clinical trials, fixed effects, pre-post designs, difference-in-differences, instrumental variables, regression discontinuity, and propensity score matching. Along the way we will explore the history and ethics of program evaluation, study design, challenges that arise when working in an international setting, the use of cost benefit analysis to interpret results, and common problems that compromise the validity of the results.

Class Meetings and Office Hours

Classes will be held Fridays from 9:55 to 11:50 a.m. in Social Science 6112.

My office hours will be Thursdays from 10 to 11:30 a.m. in Social Science 7321.

Textbook

There is no required textbook for this class. Instead, weekly readings will be available to download from Canvas. The following two general resources on program evaluation are available for free online and may serve as useful references:

- Paul Gertler, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). Impact Evaluation in Practice, Second Edition. World Bank Training Series. <http://hdl.handle.net/10986/25030>
- Shahidur Khandker, Gayatri Koolwal, and Hussain Samad (2009). Handbook on Impact Evaluation: Quantitative Methods and Practices. World Bank Training Series. <https://openknowledge.worldbank.org/handle/10986/2693>

The first provides an excellent introduction to program evaluation, while the second tackles the subject in greater depth.

Evaluation

Your overall grade for the course will be based on:

- **Class Participation (20%):** You are expected to attend class and participate in class discussion. Classes will have a lecture format, but these lectures will be far more interesting if everyone comes to class prepared to discuss the topic at hand.
- **Weekly assignments (50%):** Beginning the second week of class, there will be a weekly assignment. You will submit your answers in Canvas each Thursday by 8:00 PM. There will be two main components of the assignments:
 1. To promote discussion, it is important for everyone to complete assigned readings before lecture. As a policy practitioner, it is also vital to be able to read a program evaluation and assess its quality. Consequently, the assignments will often ask you to comment on a reading and assess the strengths and weaknesses of its methodology.
 2. Everyone should come away from this class with a working knowledge of how to conduct a program evaluation. As a result, after we have introduced each program evaluation methodology, you will be given a data set and asked to implement the method in Stata. In this case, you'll be asked to submit your Stata output, interpret results, and discuss their plausibility.
- **Final Exam (30%):** There will be a take-home final examination at the end of the term. We will discuss the exact timing on the first day of class.

Learning Outcomes

- Students will demonstrate understanding of major current and past policy debates, research findings, and analytical methodologies in international program evaluation
- Students will demonstrate critical thinking skills. They will retrieve and examine the policy literature and evaluate evidence for and against hypotheses, identify knowledge gaps, strengths and weaknesses in existing literature, synthesize knowledge, and develop conclusions
- Students will read, comprehend, and effectively summarize policy research and policy-relevant academic research
- Students will identify and appropriately respond to scenarios involving the ethical and professional issues that arise in the analysis of public programs
- Students will demonstrate the ability to maintain fidelity to objective social science-based research methods.

Credits

The credit standard for this course is met by an expectation of a total of 135 hours of student engagement with the courses learning activities, which include weekly lectures, reading assignments, and writing assignments, as described on this syllabus.

Course Outline

This section provides a tentative roadmap of the topics we'll be covering each week. Assigned readings are marked with an asterisk.

Week 1 - Introduction

Structure of the course

What is the goal of program evaluation and why is it important?

*Banerjee, Abhijit (2006). "Making Aid Work." *Boston Review*, 31(4):7-9.

Center for Global Development (2006). "When Will We Ever Learn? Improving Lives through Impact Evaluation." Washington, DC: Center for Global Development.

Easterly, William (2009). "Can the West Save Africa?" *Journal of Economic Literature*, 47(2): 373-447. (We'll take a look at Table 3 in class.)

Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). *Impact Evaluation in Practice*. World Bank Training Series. (Chapter 1)

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). *Handbook on Impact Evaluation*. Washington, DC: World Bank. (Pages 7-21 discuss evaluation terminology.)

Week 2 - Ethics of Program Evaluation

Early abuses, current oversight, and lingering controversies

Grant, Ruth and Jeremy Sugarman (2004). "Ethics in Human Subjects Research: Do Incentives Matter?" *Journal of Medicine and Philosophy*, 29(6): 717–738. (One perspective on a lingering controversy in human subjects research.)

*Hesse-Biber, Sharlene and Patricia Leavy (2010). *The Practice of Qualitative Research*. Cambridge: MIT Press. (Chapter 4: Ethics of Social Research.)

Radiolab, "Trust Engineers", <http://www.radiolab.org/story/trust-engineers/>

*Ravallion, Martin (2014). "Taking Ethical Validity Seriously." World Bank Development Impact blog post. <http://blogs.worldbank.org/impac evaluations/taking-ethical-validity-seriously>

Week 3 - Review and Basic Concepts

Formal treatment of our goal (i.e., intent-to-treat vs treatment-on-the-treated)

Primer on regression and statistical concepts (e.g., point vs. interval estimates, causal inference, counterfactual, identification, validity, significance, heterogenous treatment effects, dose response, significance, data mining)

*Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). *Impact Evaluation in Practice*. World Bank Training Series. (Chapter 3.)

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). *Handbook on Impact Evaluation*. Washington, DC: World Bank. (Pages 22-23 discuss the counterfactual.)

*Manski, Charles (1995). Identification Problems in the Social Sciences. Cambridge: Harvard University Press. (Introduction discusses causal inference and identification.)

Manski, Charles (2011). “Policy Analysis with Incredible Certitude.” The Economic Journal, 121(August): F261–F289.

Week 4 - Regression Analysis and Difference Estimators

Using a difference estimator to identify a treatment effect

What can go wrong? What strategies can be used to address this?

*Berk, Justin and Achyuta Adhvaryu (2012) “The impact of a novel franchise clinic network on access to medicines and vaccinations in Kenya: a cross-sectional study” BMJ Open 2012(2).

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). Handbook on Impact Evaluation. Washington, DC: World Bank. (Pages 23-29.)

Week 5 - Randomized Clinical Trials (RCTs)

History of Randomized Clinical Trials

Why can't we just do RCTs all the time and call it a day?

Can anything go wrong, or is an RCT bulletproof?

Duflo, Esther, Rachel Glennerster and Michael Kremer (2007). “Using Randomization in Development Economics Research: A Toolkit.” Working paper.

Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). Impact Evaluation in Practice. World Bank Training Series. (Chapter 4.)

*Glennerster, Rachel and Kudzai Takavarasha (2013). Running Randomized Evaluations: A Practical Guide. Princeton: Princeton University Press. (Chapter 7 discusses what can go wrong with an RCT.)

Jameel Poverty Action Lab, “Case Study: Deworming in Kenya”

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). Handbook on Impact Evaluation. Washington, DC: World Bank. (Chapter 3.)

Week 6 - Propensity Score Matching

What is propensity score matching and how can it help with omitted variable bias?

Can anything go wrong, or is propensity score matching bulletproof?

Heinrich, Carolyn, Alessandro Maffioli, and Gonzalo Vázquez (2010). A Primer for Applying Propensity-Score Matching. Washington, DC: Inter-American Development Bank.

*Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). *Impact Evaluation in Practice*. World Bank Training Series. (Chapter 8.)

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). *Handbook on Impact Evaluation*. Washington, DC: World Bank. (Chapter 4.)

Week 7 - Difference in Difference (DnD) Estimators

What is a natural experiment?

What are the advantages of DnD (and DnDnD) over a difference estimator?

How do we go about picking suitable control groups?

Can anything go wrong, or is a DnD approach bulletproof?

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4):795-813.

*Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). *Impact Evaluation in Practice*. World Bank Training Series. (Chapter 6.)

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). *Handbook on Impact Evaluation*. Washington, DC: World Bank. (Chapter 5.)

Week 8 - Fixed Effects (FE)

What are fixed-effects and how can they help with omitted variable bias?

Can anything go wrong, or is a fixed-effects approach bulletproof?

Frankenberg, Elizabeth, Rayan Suriastini, and Duncan Thomas (2003). "Can expanding access to basic healthcare improve children's health status? Lessons from Indonesia's 'midwife in the village' programme." *Population Studies* 39(1):3-19.

*Galiani, Sebastian, Paul Gertler, and Ernesto Schargrotsky (2005). "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113(1):83-120.

Rangel, Marcos (2014). "Is Parental Love Colorblind? Human Capital Accumulation within Mixed Families." *Review of Black Political Economy* July 2014.

Week 9 - Instrumental Variables (IV)

What is an instrumental variable and how can it help with omitted variable bias?

What qualities are we looking for in an instrument?

Can anything go wrong, or is an instrumental variables approach bulletproof?

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). *Handbook on Impact Evaluation*. Washington, DC: World Bank. (Chapter 6.)

*Ravallion, Martin (2001). "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation." World Bank Economic Review 15(1):115-140.

Week 10 - Regression Discontinuity (RD)

*What is regression discontinuity and how can it help with omitted variable bias?
Can anything go wrong, or is a regression discontinuity approach bulletproof?*

*Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). Impact Evaluation in Practice. World Bank Training Series. (Chapter 5.)

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). Handbook on Impact Evaluation. Washington, DC: World Bank. (Chapter 7.)

Week 11 – Distributional Impacts of a Program

*Why are we so fixated on average effects?
Can quantile regression help?*

*Banerjee, Abhijit, Esther Duflo, Rachel Glennerster and Cynthia Kinnan (2015). "The Miracle of Microfinance? Evidence from a Randomized Evaluation." American Economic Journal: Applied Economics, 2015, vol. 7, issue 1, pages 22-53

Khandker, Shahidur, Gayatri Koolwal, and Hussain Samad (2010). Handbook on Impact Evaluation. Washington, DC: World Bank. (Chapter 8.)

Week 12 - Cost Benefit Analysis (CBA)

*Why are program evaluation and CBA such important complements?
Common assumptions, oversights, and mistakes when performing CBA
Subjecting CBA to robustness checks*

Asian Development Bank (2013). Cost Benefit Analysis for Development: A Practical Guide. Washington, DC: Asian Development Bank.

*Lomborg, Bjorn (2005). Global Priorities Bigger than Climate Change. TED Talk. https://www.ted.com/talks/bjorn_lomborg_sets_global_priorities

World Bank (2010). Cost-Benefit Analysis in World Bank Projects. Washington, DC: World Bank.

Week 13 - Qualitative Analysis and Study Design

*Qualitative and mixed-method studies
Instrument design and selecting a sample size
Selecting outcome variables*

Baker, Judy (2000). *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*, Washington DC: LCSPP/PRMPO, The World Bank. (Chapters 1-2 discuss practical considerations that are often overlooked.)

Bamberger, Michael (2000). "The Evaluation of International Development Programs: A View from the Front." *American Journal of Evaluation*, 21:95-102.

*Bamberger, Michael, Vijayendra Rao, and Michael Woolcock (2010). "Using Mixed Methods in Monitoring and Evaluation." World Bank Policy Research Working Paper 5245.

Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura Rawlings, and Chistel Vermeersch (2016). *Impact Evaluation in Practice*. World Bank Training Series. (Chapters 2 & 9-14)

Glennerster, Rachel and Kudzai Takavarasha (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press. (Chapter 5 discusses outcomes and survey instruments, Chapter 6 discusses statistical power)

Week 14 - Methodological Debates

Why have RCTs been so controversial in the development community?

Given everything we have discussed, is there any value in descriptive studies?

Banerjee, Abhijit and Esther Duflo (2009). "The Experimental Approach to Development Economics." *Annual Review of Economics*, 1:151-178.

*Deaton, Angus (2010). "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*, 48(2): 424-55.

Duncan, Greg (2008). "When to Promote, and When to Avoid, a Population Perspective." *Demography* 45(4): 763-784.

Imbens, Guido (2010). "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature*, 48(2): 399-423.

Ravallion, Martin (2009). "Should the Randomistas Rule?" *Economist's Voice*.

Rodrik, Dani (2008). "The New Development Economics: We Shall Experiment, but Shall We Learn?" Working Paper.