Research Methods in Global Health

Duke University, Spring 2015

Course Code:	GLHLTH 371-01/PSY 309-01
Instructor:	Eric Green, Ph.D.
Contact:	eric.green@duke.edu / @ericpgreen
Office Hours:	Th 3-5 Trent 238 or by appointment
Classroom:	Trent 040
Time:	TTh 1:25-2:40
Web:	Sakai
TA:	Lauren Zalla / lauren.zalla@duke.edu / W 12-2, Trent Lounge
TA:	Amy Finnegan / amy.finnegan@duke.edu / by appt, Rubenstein

COURSE DESCRIPTION

This course will introduce you to research methods in global health. Global health is a multi-disciplinary field, so we will consider approaches common to the behavioral and social sciences, public health, and medicine. Our primary interest will be the study of causal inference. In global health, we are often interested in knowing what programs and interventions "work" and why. To answer questions of impact, we often turn to randomized controlled trials, a mainstay of medical research. As such, we will spend a good amount of time understanding the rationale, process, and limitations of field experiments. Randomization is not always possible or advisable, however, and researchers must build a causal argument using non-experimental methods. We will review several approaches, consider relevant threats to causal inference, and discuss how to improve non-experimental research designs. As we build up to this discussion, we will cover research basics, such as developing and testing theory, asking good questions, understanding variability, designing good measurement, and selecting research participants. In the latter part of the course, we will turn to more specialized topics in global health research, such as cost effectiveness, community based participatory research, research on humanitarian aid, and monitoring & evaluation.

LEARNING OBJECTIVES

We have two broad goals this semester:

1. to make you a better consumer of research, and
2. to prepare you to contribute to research teams at Duke and beyond.

Each class session has specific learning objectives to help you achieve these overall goals. Jump to the "Assignments" section to read the objectives of each session.

CLASS FORMAT

This class will incorporate pedagogical concepts and strategies from "flipped" classrooms and team based learning. After the first class, you will be assigned to a team of 4 to 6 students. You and your team will work

together on in-class assignments; expectations for team meetings outside of class sessions will be kept to a minimum.

You will be asked to prepare for class by completing assignments and reviewing pre-recorded lectures. At the beginning of most class sessions you will complete a short online assessment based on the assigned readings and lectures (closed book, closed notes). A typical assessment will consist of 5 to 10 multiple choice or short answer questions and last about 15 minutes.

Following these Individual Readiness Assessments, you and your team will complete the same assessment together as a team. We will then discuss answers as a group, and I may present additional material in class to clarify concepts. Class time will often be used to solve applied problems in teams. A link to "lab" assignments is here.

Students needing additional help understanding the material will be encouraged to attend office hours or set up individual or small-group meetings. We will regularly post post-lecture videos to clarify confusing concepts.

COURSE REQUIREMENTS

You will need to bring a laptop or a mobile device to every class to participate in the Individual Readiness Assessments that will be administered online via Sakai. You are responsible for ensuring that you can connect your device to the Duke network and access Sakai. Please review the Sakai Online Testing Guidelines prior to the second class session.

There are two required books for this course, both on 3-hour reserve at Perkins:

1. Leary, M. R. (2011). Introduction to Behavioral Research Methods (6th Edition). Boston: Pearson Education, Inc. (Amazon Kindle Version)
2. Glennerster, R. & Takavarasha, K. (2013). Running Randomized Evaluations: A Practical Guide. Princeton, NJ: Princeton University Press. (Amazon Kindle version)

Other resources will be marked Required or Recommended. Required means required. Recommended resources are optional. If there is a link to the resource, you can access it online. If not, it will be on 3-hour reserve at Perkins.

EVALUATION

Students should abide by the Duke Community Standard at all times. If a questionable circumstance arises, do not hesitate to seek my guidance (before is always better than after).

Any student with a documented disability needing academic adjustments or accommodations should speak with me during the first two weeks of class. All discussions will remain confidential. Students with disabilities will also need to contact the Student Disability Access Office.

Grading Scale

The grading scale for this course is as follows:

A+: 100-98	A: 97-93	A-: 92-90
B+: 89-87	B: 86-83	B-: 82-80

A+: 100-98	A: 97-93	A-: 92-90
C+: 79-77	C: 76-73	C-: 72-70
D+: 69-67	D: 66-63	D-: 62-60
F: 59 and below		

Final Grade

Your final grade will be a weighted average of your Individual Readiness Assessments (30%), Team Readiness Assessments (15%), application activities (15%), final exam (30%), and peer evaluation (10%). If you are in between grades, and if the difference is less than or equal to 0.5, your score will round up (e.g., 97.5 rounds up to 98).

Components of Final Grade

Individual Readiness Assessments (30%)

At the beginning of most class sessions you will complete a short online assessment based on the assigned readings and lectures (closed book, closed notes). A typical assessment will consist of 5 to 10 multiple choice or short answer questions and last about 15 minutes. There may be occasional opportunities for extra credit. Students carrying an average percentage correct of 93 percent or higher after the final in-class assessment will be exempted from having to sit for the final exam (a score of 100 percent will be awarded).

Missed classes will result in a score of 0 and there will be no make-ups, even if the absence is excused. Instead, I will drop your 4 lowest individual assessment scores. Plan to arrive to class on time; late arrivals will not be given additional time to complete assessments.

Missed classes	Lowest earned scores	Lowest graded scores
0	20,30,40,50,60,70,...	60,70,...
1	0,20,30,40,50,60,70,...	50,60,70,...
2	0,0,20,30,40,50,60,70,...	40,50,60,70,...

Team Readiness Assessments (15%)

Following the Individual Readiness Assessment, you and your team will complete the same assessment together as a team. Teams will use scratch off cards to mark answers.

Attempts Before Correct Answer	Points
1	4
2	3
3	2
4	1
5	0

Members who miss class will not receive a team score that day. For example, if there are 28 team assessments and a member misses 2 classes, this student's final team score will be an average of her other 26 team scores.

Application Activities (15%)

I may assign short assignments and ask you to complete them during or after class. If I indicate that the assignment will be graded, late submissions will be penalized 5 percentage points every 24 hours late. Students who miss class will be expected to complete the assignment.

Final Exam (30%)

The final exam for this course is scheduled for Monday April 27 from 9am-12pm ET. Students carrying an average percentage correct of 93 percent or higher after the final Individual Readiness Assessment will be exempted from having to sit for the final exam (a score of 100 percent will be awarded).

Peer Evaluation (10%)

Team members will complete confidential ratings of all other members at the end of the semester. Student ratings will be averaged to construct an overall score.

CLASS SCHEDULE

Date	Session	Date	Session
Th Jan 08	1. Introduction	T Mar 03	16. Randomized experiment 3
T Jan 13	2. On the nature of research	Th Mar 05	17. Quasi-experimental 1
Th Jan 15	3. Theory and questions	T Mar 10	Spring Break
T Jan 20	4. Variability	Th Mar 12	Spring Break
Th Jan 22	5. Measurement	T Mar 17	18. Quasi-experimental 2
T Jan 27	6. Selection 1	Th Mar 19	19. Systematic reviews
Th Jan 29	7. Selection 2	T Mar 24	20. Single case designs
T Feb 03	8. Threats to causal inference	Th Mar 26	21. Qualitative and mixed methods
Th Feb 05	9. How to read critically	T Mar 31	22. M&E and cost effectiveness
T Feb 10	10. Screen and diagnose	Th Apr 02	23. Implementaton science
Th Feb 12	11. Describe	T Apr 7	24. CBPR
T Feb 17	12. Correlate	Th Apr 9	25. Design thinking
Th Feb 19	13. Observe	T Apr 14	26. Conflict, disaster, and research
T Feb 24	14. Randomized experiment 1	Th Apr 16	27. Replication & reproducibility
Th Feb 26	15. Randomized experiment 2	T Apr 21	28. Research ethics

ASSIGNMENTS

Abbreviation	Title
BRM	Introduction to Behavioral Research Methods
RRE	Running Randomized Evaluations

Enrolled students will be able to access podcasts of pre-recorded lectures on Sakai (click on Warpwire), at https://dukestream.duke.edu/, or via the direct links on the syllabus. If going through Warpwire, look for GLHLTH 371 (2015:Spr). I will also post the slide decks from these podcasts, but not until after class. This is to encourage you to take notes while viewing the podcasts.

If I ask you to complete an assignment using the R statistical program, please go here and click on "Docker" to create your own personal RStudio environment.

1. Introduction

Welcome to Research Methods in Global Health! I'll introduce the course, explain the format, and answer your questions.

Required

Set up your laptop to connect to the wireless network and review the Sakai Online Testing Guidelines.

2. On the nature of research

It's likely that you come to this course with a good sense of the nature of social science and global health research. You are all consumers of research, and some of you might even be engaged in the research process already. So in this session we'll build a solid framework on this foundation. Our objectives are as follows:

- to explain the scientific approach
- to outline major research approaches (e.g., quantitative, qualitative, mixed methods) and how these relate to research designs, methods, and philosophical worldviews
- to identify how to match research goals to research designs

Required

1. Lecture
2. BRM 1, "Research in the behavioral sciences".
3. RRE 3, "Asking the right questions".
4. Creswell, J. W. (2014). *Research Design.* Thousand Oaks, CA: Sage Publications. Chapter 1, "The selection of a research approach".

Recommended

- Video: *What is Evaluation*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)

- White, H. (2009). Some reflections on current debates in impact evaluation. *3ie Working Paper 1*.
- Rogers, P. (2014). Overview of Impact Evaluation, *Methodological Briefs: Impact Evaluation 1*, UNICEF Office of Research, Florence.
- Video: *Building Blocks of Impact Evaluation*. UNICEF (2014).

3. Theory and questions

Some research aims to develop theory. Other research aims to test existing theory. Sometimes research does not do either one, when maybe it should. In this session we'll explore the use of theories and get specific about how researchers can develop theoretical frameworks. Our objectives are as follows:

- to state the components of a theory
- to describe how theory is often used in different research approaches
- to define what we mean by "theory of change" and "logic model", and to understand how these frameworks can help guide a research study

During the in-class activity you will create a logic model for a published impact evaluation and develop a theory of change for a new program.

Required

1. Lecture
2. Creswell, J. W. (2014). *Research Design*. Thousand Oaks, CA: Sage Publications. Chapter 3, "The use of theory".
3. W. K. Kellogg Foundation (2004). *Logic Model Development Guide*. Chapter 1, "Introduction".

Recommended

- White, H. (2009). Theory-based impact evaluation: principles and practice. *3ie Working Paper 3*.
- Rogers, P. (2014). Theory of Change, *Methodological Briefs: Impact Evaluation 2*, UNICEF Office of Research, Florence.

4. Variability

Various public opinion polls ask random samples of American adults if they approve or disapprove of the way presidents handle their role in office. If everyone surveyed said "approve", the politician would be thrilled, and we would learn that the public has a lot of confidence in her performance, but there would be no variability to explore for that particular snapshot (cross section) in time. No differences by age or gender. No differences by socioeconomic status or political party affiliation. There would probably be variability over time, however. Gallup estimated that President George W. Bush had a 90 percent approval rating in a poll conducted less than two weeks after September 11, 2001. By October 31, 2008, his approval rating was 25 percent in the same poll. There is also variability over units (e.g., presidents); the average approval rating in Gallup polls for #43 (49.4%) is more than 10 percentage points lower than his father's average approval rating (60.9%). Why? Maybe we think the answer has something to do with economic indicators, so we explore this relationship statistically. Humans are complicated, so a single variable is not going to explain all of the variance we observe in average approval ratings, but maybe there is some systematic variance between approval ratings and economic conditions that we can find. Hopefully the answer will inform a theory about presidential approval ratings that we set out to test. We'll explore the nature of variability in this class session. Our objectives are as follows:

- to recognize the importance of variability in the research process
- to calculate variance
- to compare the difference between systematic and random error variance
- to describe the concept of an effect size

During the in-class activity you will get an introduction to R, RStudio, literate programming, and the concept of variability.

Required

1. Lecture
2. BRM 2, "Behavioral variability and research".

5. Measurement

Often we design studies to determine whether X causes Y. X and Y always start out as constructs that need to be defined. For instance, we might conduct a randomized experiment to determine if assignment to the treatment condition, X, has an impact on depression symptoms, Y. Issues of measurement make us think through what we mean by Y in this case. What is depression? There is no blood test for depression (yet?), so we probably need to ask study participants (or maybe we could ask family members or somehow observe behaviors). But how? "On a scale of 1 to 10 where 1 is low and 10 is high, please rate your depression"? Or should we view depression as a larger idea that can't be measured directly (latent construct) and ask about individual experiences (symptoms) thought to make up this thing called depression? For instance, "In the past 4 weeks, have you had trouble sleeping?" "Have you cried frequently?" Etc. What if we combined all of these responses into a single score? Would it measure depression? We'll address these issues and more in this session. Our objectives are as follows:

- to list scales of measurement
- to identify how to evaluate the quality of a measure (reliability and validity)
- to define different approaches to measurement (observational, self-report, etc)

During the in-class activity you will review published articles to find at least one example of the following types of measurement: (i) observational; (ii) physiological; and (iii) self-report. Then you will explore a real dataset to determine the reliability of a measure of emotional distress.

Required

1. Lecture
2. BRM 3, "The measurement of behavior".
3. BRM 4, "Approaches to psychological measurement".

Recommended

- Video: *Measurement and outcomes*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)
- RRE 5, "Outcomes and instruments".
- White, H., & S. Sabarwal (2014).Developing and Selecting Measures of Child Well-Being, *Methodological Briefs: Impact Evaluation 11*, UNICEF Office of Research, Florence.

6. Selection 1

Now that you know what you want to measure and how you are going to measure it, you need to determine your targets of measurement. If you want to run a survey, who will you invite to participate? How will you find these people? We'll discuss these issues and more in this session. Our objectives are as follows:

- to compare probability and non-probability approaches to sampling
- to articulate the rationale behind different probability and non-probability sampling approaches (e.g., convenience sampling, stratified random sampling)

During the in-class activity you will prepare a slide deck that explains how you will recruit a sample for a nationally representative malaria survey in Liberia.

Required

1. Lecture
2. BRM 5, "Selecting research participants".

7. Selection 2

Whether you will use probability or non-probability approaches to sampling, you need to decide on how many people to include in your sample. If your goal is hypothesis testing, then considerations of sample size get us into a conversation about power. Our objectives are as follows:

- to explain the concept of statistical power to a lay audience
- to list the determinants of power
- to perform power calculations to determine sample size

During the in-class activity you will learn how to estimate the required sample size to detect a difference between two group means. In the process, you will review core concepts, such as hypothesis testing and statistical power.

Required

1. Lecture
2. RRE 6, "Statistical power".

Recommended

- Video: Sample size and power calculations. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)

8. Threats to causal inference

Many of the today's most pressing questions are causal. Does X cause Y? Stated differently, does Program X increase/decrease some target outcome? To answer questions of impact, we often turn to randomized controlled trials, a mainstay of medical research. Causal inference in RCTs is often straightforward, but not always. Sometimes randomization is "broken" by treatment non-adherence or attrition. Randomization is not always possible or advisable, however, and researchers must build a causal argument using non-experimental methods. Whether randomized or not (or "quasi randomized"), it is important to think through how to build a causal argument. Donald Campbell is known in psychology and education literatures for his approach to causal inference that focuses on threats to causal inference. Donald Rubin's causal model, which is mathematical in nature, is more well known in economics, public health, and medicine. We'll consider the Dons' perspectives and think about how different disciplines approach causality. Our objectives are as follows:

- to explain the importance and challenge of causal attribution
- to compare Campbell's and Rubin's perspectives on causal inference
- to describe common threats to causal inference

Required

1. Lecture
2. Video: *Strategies for Causal Attribution*. UNICEF (2014).
3. West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods, 15(1)*, 18-37.
4. RRE 7, "Threats".

Reading preview

Recommended

- Video: *Managing threats to evaluation and data analysis*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)
- Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15(1)*, 3-17.
- Imbens, G. W. (2010). An economist's perspective on Shadish (2010) and West and Thoemmes (2010). *Psychological Methods, 15(1)*, 47-55.
- Rogers, P. (2014). Overview: Strategies for Causal Attribution, *Methodological Briefs: Impact Evaluation 6*, UNICEF Office of Research, Florence

9. How to read critically

We are inundated with research results in our daily lives. Often we consume research findings through a bad game of telephone. A friend hears about a study from another friend who heard something on the news based on a story by a journalist who read a press release written by a university communications staffer based on the abstract of the peer-reviewed article published in a scientific journal. "We find modest evidence that the program boosts IQ by 1 point" becomes "Brain training will unleash your hidden potential and take you to new intellectual heights". Other times we actually sit down to read the primary source ourselves. If you think the game of telephone can confuse, try reading something written by an academic! Fortunately, there are strategies for sifting through the confusion, and we'll discuss them in class. Our objectives are as follows:

- to outline the key components of results needed to evaluate the voracity of the conclusions
- to critique scientific claims
- to explain the purpose of peer review

During the in-class activity you will get to practice your critical appraisal skills with several articles about deworming. You'll complete a CONSORT checklist, visit ClinicalTrials.gov, and consider threats to causal inference.

Required

1. Lecture
2. Young, J. M., & Solomon, M. J. (2009). How to critically appraise an article. *Nature Clinical Practice Gastroenterology & Hepatology, 6(2)*, 82-91.
3. Schulz KF, Altman DG, Moher D, for the CONSORT Group (2010) CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *PLoS Med 7(3): e1000251*. doi:10.1371/journal.pmed.1000251

Reading preview

Recommended

- Valentine, J. C. (2009). Judging the quality of primary research. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). *The Handbook of Research Synthesis and Meta-analysis*. Russell Sage Foundation.
- Ioannidis J. (2005). Why most published research findings are false. *PLoS Med 2(8): e124*. doi: 10.1371/journal.pmed.0020124.

10. Screen and diagnose

Mammography saves lives, but should women of all ages undergo preventative screening? False positives, a.k.a. the test suggests a patient has cancer when that's not the case, can mean women choose to get unnecessary mastectomies. This comes with obvious physical and emotional health risks, as well as costs to the health system. The American Cancer Society thinks that the benefits outweigh the risks and recommends routine screening for women ages 40 to 49. The U.S. Preventive Services Task Force disagrees, recommending that women discuss the issue with a medical provider before deciding to get screened. In many low-income settings globally, this particular issue is moot as there are few resources for preventative screening. But diagnostic instruments are still common in global health research. For instance, maybe we want to test a new treatment for depression. OK, but who is depressed? How do we know if there is not a blood test? Isn't this a "Western" concept? We'll discuss these issues and review the foundations of evaluating screening and diagnostic tests. Our objectives are as follows:

- Define what is meant by "gold standard"
- Calculate sensitivity, specificity, and pos/neg predictive value
- Interpret a ROC curve
- Contrast accuracy with precision

During the in-class activity you will learn about the search for a Rapid Diagnostic Test for the Ebola Virus Disease and consider how you would evaluate candidates.

Required

1. Lecture
2. Newman et al. (2007). Designing studies of medical tests. In Hulley et al. *Designing Clinical Research.* Philadelphia, PA: Lippincott Williams & Wilkins.
3. Angold et al. (1995). The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*, 237 - 249.

Reading preview

11. Describe

The first step in understanding and explaining is description. Sometimes the process stops at description, as is often the case with Demographic and Health Surveys and census reports. The goal is to characterize the population. X% are illiterate. Y% have an unmet need for contraception. Z% are HIV positive. Good description requires an understanding of basic statistics, so we'll review some basics in this session. Description can also be qualitative in nature. We'll dig into qualitative research later this semester, but I'll plant the concept of "thick description" now. Our objectives are as follows:

- to describe the purpose of descriptive research
- to list types of descriptive research
- to display frequency distributions
- to calculate central tendency
- to calculate measures of variability

During the in-class activity you will learn about how to estimate basic descriptive statistics for complex survey data.

Required

1. Lecture
2. BRM 6, "Descriptive research".

Reading preview

12. Correlate

Once you've described X and Y, the next step is often to determine whether these variables are associated. Whether they are related. Whether the covary, or change together. Whether they are correlated. Correlational research takes many forms, and we'll explore several approaches in this session. We'll also consider the limitations of correlational research. You've probably heard that correlation does not imply causation. Here's a good demo of that. Did you know there is a nearly perfect correlation between the per capita consumption of cheese and the number of people who have died by becoming tangled in their bedsheets? Our objectives are as follows:

- to describe the purpose of correlational research
- to define and calculate correlation coefficients and the coefficient of determination

- to explain the concept of statistical significance
- to describe the relationship between correlation and causality
- to explain how regression is used to predict one variable from others
- to describe the purpose of structural equation modeling, multilevel modeling, and factor analysis

Required

1. Lecture
2. BRM 7, "Correlational research".
3. BRM 8, "Advanced correlational strategies".

Reading preview

13. Observe

When randomization is not possible, researchers will sometimes turn to cohort, case-control studies, or cross-sectional studies (a.k.a. prevalence studies, discussed with descriptive research). You'll see these categorized as observational or non-experimental designs, and employed a lot by epidemiologists. Researchers use these designs to determine whether there is an association between some exposure and a disease. Prospective cohort designs are the best of the lot, but they can be quite expensive due to large sample size requirements and the need to wait a number of years to see who develops the disease. Emily Oster gives a good example of how this design can be impractical for determining whether exposure to cell phones leads to rare diseases like gliomas. The fall back in situations like this is often the case-control design. Find some people with gliomas ("cases") and some people without gliomas ("controls"), and ask them about their cell phone use patterns. We'll discuss the pros and cons of each design. Our objectives are as follows:

- to compare cohort and case-control studies
- to contrast prospective and retrospective cohort designs
- to explain the pros and cons of each design

During the in-class activity you will read about a prospective cohort study of postnatal depression and child development in South Africa and make links between this article and everything we've learned in class up until now.

Required

1. Lecture
2. Cummings et al. (2007). Designing a cohort study. In Hulley et al. *Designing Clinical Research.* Philadelphia, PA: Lippincott Williams & Wilkins.
3. Newman et al. (2007). Designing cross-sectional and case-control studies. In Hulley et al. *Designing Clinical Research.* Philadelphia, PA: Lippincott Williams & Wilkins.

Reading preview

14. Randomized experiment 1

UPDATE: This session will combine material from sessions 14 and 15 to enable us to cancel class on 2/26 to make it easier for students to attend the Josh Angrist talk.

The trilogy begins. We'll start with the "what" and "why" of randomized evaluations, a.k.a. randomized controlled trials, a.k.a. impact evaluations, a.k.a. social experiments, a.k.a. randomized field trials. Our objectives are as follows:

- to explain the rationale for experimental designs
- to evaluate whether or not experimental designs are the "gold standard"

Required

1. Lecture (in class)
2. RRE 1, "The experimental approach".
3. RRE 2, "Why randomize?".

Reading preview

Recommended

- Video: Michael Kremer, *The Origin and Evolution of Randomized Evaluations in Development*
- Video: *Why randomize?*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)
- Teele, D. (2014). *Field Experiments and Their Critics*. New Haven: Yale University Press.
- NIH (2008). *Clinical trial phases*.

15. Randomized experiment 2

In this session we'll move to a discussion of the "how". What are the basic issues to consider? Or, "how things can go wrong and what you can do to prevent this from happening." Our objectives are as follows:

- to contrast independent and dependent variables
- to explain procedures for randomizing
- to list how experiments can be compromised
- to describe external validity

Required

1. Lecture (in class)
2. BRM 9, "Basic issues in experimental research".

Recommended

- Video: *How to randomize I*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)
- Video: *How to randomize II*. Duflo, E., Glennerster, R., & Banerjee, A. RES.14-001 Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs 2009, Spring 2009. (MIT OpenCourseWare: Massachusetts Institute of Technology)
- Video: *Data Collection and Analysis*. UNICEF (2014).
- Duflo, E. et al. (2006). *Using randomization in development economics research: A toolkit*.
- RRE 4, "Randomizing".

16. Randomized experiment 3

Next we'll consider various flavors of experimental designs, such as factorial designs, cluster randomized trials, stepped wedge designs, and randomized encouragement designs. Factorial designs are often used when researchers want to estimate the marginal impact of program components. Cluster (a.k.a. group) randomized designs are handy when it does not make sense or is impossible to randomize at the individual-level. Often in global health it is necessary to roll out an intervention over time, rather than all at once in a parallel fashion, in which case the order can be randomized in a stepped wedge design. Other times we want to know the impact of programs that are universally available but not universally adopted. In these cases, we can randomize who is explicitly encouraged to participate and use this invitation as an instrumental variable (we'll see IV approaches again during a discussion of quasi-experimental approaches). Our objectives are as follows:

- to compare factorial designs and one-way designs
- to explain the rationale for cluster randomized trials
- to identify situations in which a stepped wedge design would be preferable
- to identify situations where randomized encouragement designs would be viable

During the in-class activity you will become more familiar with randomization strategies used in global health and international development by preparing summaries of studies that used either individual randomization, cluster randomization, randomized encouragement, factorial designs, or stepped wedge designs.

Required

1. Lecture
2. BRM 10, "Experimental design".

Reading preview

Recommended

- Hayes, R., & Moulton, L. (2009). *Cluster Randomised Trials*. Chapman and Hall/CRC.
- Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: a systematic review. *BMC Medical Research Methodology, 6(1)*, 54.
- Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28(2)*, 182-191.

17. Quasi-experimental 1

Experiments are the "gold standard" in the eyes of many people, but researchers are not always able to assign people or clusters to conditions or otherwise manipulate an independent variable. In these cases, researchers might rely on non-experimental designs commonly referred to as "quasi-experimental" designs. The name of the game in quasi-experimental research is to reduce threats to internal validity, something that randomization pretty much takes care of naturally. Beware: not all non-experimental designs are created equal. In this session we'll focus on pre/post designs (including differences-in-differences) and time series designs. Our objectives are as follows:

- to compare quasi-experimental and experimental designs
- to describe the logic of difference-in-differences
- to interpret an interrupted time-series analysis (segmented regression)

During the in-class activity you will go back to the days of John Snow and cholera in London and apply difference-in-difference estimation.

Required

1. Lecture
2. BRM 13, "Quasi-experimental design".

Reading preview

Print a copy of this matrix from JPAL

18. Quasi-experimental 2

We'll stick on this theme and discuss a few more non-experimental designs that fall under the "quasi" heading. For instance, what do you get when the world just happens to operate in a way that is approximately random? How about a natural experiment. Or what if the world is systematic, and people are exposed to a policy or intervention based on whether they are above or below some cutoff point, say a poverty line? In this case, regression discontinuity would compare outcomes for people just above and just below the cut point since they should be pretty similar, except that some were eligible for the program but others were not. Statistical matching is another way to create a comparison group. Alternatively, you might even be able to find an instrumental variable that predicts participation in a program but is uncorrelated with your outcome of interest (e.g., Vietnam War draft lottery numbers). Our objectives are as follows:

- to explain how natural experiments work
- to interpret the results of a regression discontinuity design
- to explain the purpose of matching
- to list possible instrumental variables for global health research

During the in-class activity you will explore the use of regression discontinuity in the wild.

Required

1. Lecture
2. Blattman, C., & Annan, J. (2010). The consequences of child soldiering. *The Review of Economics and Statistics, 92(4)*, 882-898.

3. West, S. G. et al. (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health, 98(8)*, 1359.

Reading preview

Recommended

- Evans, D. (2013). *Regression discontinuity porn*. Development Impact.
- White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, *Methodological Briefs: Impact Evaluation 8*, UNICEF Office of Research, Florence.

19. Meta-analysis and systematic reviews

We live in an era of evidence-based medicine. Data rule. As you should understand by this point in the semester, however, not all evidence merits the same consideration. This is where systematic reviews come in. The Cochrane Collaboration is the largest producer of systematic reviews. Systematic reviews are literature reviews on steroids. These reviews are meticulously planned and conducted. After defining a specific research question, review teams (often consisting of subject matter experts, research/clinical librarians, and statisticians) devise specific, reproducible search terms for major literature databases and often articulate a strategy for systematically searching other "gray" sources. It's not uncommon for these searches to return thousands of hits. It's also not uncommon for teams to screen abstracts and exclude all but a handful of eligible articles. When possible, teams often extract detailed statistics from each eligible study and conduct a meta-analysis that estimates an overall effect size. We'll spend this class session discussing the need for research synthesis, the basic steps in conducting a systematic review, and tips for reading and understanding the results. Our objectives are as follows:

- to explain the need for research synthesis
- to outline the steps of conducting a systematic review
- to interpret results from a meta-analysis

During the in-class activity you will run through the basic steps of a systematic review.

Required

1. Lecture
2. Cooper, H. & Hedges, L. V. (2009). Research synthesis as a scientific process. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). *The Handbook of Research Synthesis and Meta-analysis*. Russell Sage Foundation.
3. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. *Cochrane Database of Systematic Reviews 2013, Issue 12*. Art. No.: CD007458. DOI: 10.1002/14651858.CD007458.pub3.

Reading preview

Recommended

- Sackett, D. L., Rosenberg, W., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312(7023)*, 71-72.
- Murad, M. H., Montori, V. M., Ioannidis, J. P., Jaeschke, R., Devereaux, P. J., Prasad, K., ... & Guyatt, G. (2014). How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA, 312(2)*, 171-179.

20. Single case designs

N-of-1 trials. As in one person. In these within-subject designs, the patient under investigation serves as his own control, so there is no comparison group, unlike the randomized controlled trial. Also unlike RCTs that estimate average treatment effects, SCDs speak to causality for the single case. One class of SCDs is withdrawal or reversal designs. For instance, ABAB. Take a baseline, A, then introduce the treatment, B, then remove it for another baseline, then bring it back. Visual analysis should show that the target behavior responds during the treatment period and reverts during the baseline periods. In the multiple baseline design, at least three people (or behaviors or settings) are studied. One person gets the treatment at a time. The expectation is that patients will not show changes in target behaviors *until* they start the treatment. Graphs will be your friend here. I'll show you some. Our objectives are as follows:

- to explain the rationale for single case designs
- to interpret results from reversal and multiple baseline studies

Required

1. Lecture
2. BRM 14, "Single case research".

Reading preview

21. Qualitative and mixed methods

Qualitative research can be a powerful complement to a quantitative impact evaluation in addition to being valuable on its own as a method of developing and testing theory. We'll discuss specific qualitative approaches and analytic strategies and consider how to mix QUAL and QUANT approaches. Should QUAL come first? Second? Simultaneously? We'll discuss how to make these decisions. Our objectives are as follows:

- to compare qualitative methods to quantitative methods
- to list rationales for using qualitative methods
- to outline major qualitative approaches
- to explain the rationale for mixed methods research
- to outline various mixed methods designs

During the in-class activity you will explore how software can help you organize and analyze your qualitative data.

Required

1. Lecture
2. Padgett, D. K. (2011). *Qualitative and Mixed Methods in Public Health*. SAGE Publications. Chapter 1, "Introduction".
3. Padgett, D. K. (2011). *Qualitative and Mixed Methods in Public Health*. SAGE Publications. Chapter 2, "Choosing the right qualitative approach(es)".
4. Creswell, J. W. (2014). *Research Design*. Thousand Oaks, CA: Sage Publications. Chapter 10, "Mixed methods procedures".

Recommended

- Mack, N. et al. (2005). *Qualitative research methods: A data collector's field guide*. FHI360.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics 44(4)*, 823-830.
- McDonald, B., & P. Rogers, (2014). Interviewing, *Methodological Briefs: Impact Evaluation 12*, UNICEF Office of Research, Florence.
- Humphreys, M., & Jacobs, A. (2014). *Mixing Methods: A Bayesian Integration of Qualitative and Quantitative Approaches to Causal Inference*.

22. M&E and cost effectiveness

Spend more than 5 minutes in global health and you will hear someone talking about M&E. Monitoring and evaluation. What is it, and how does it relate to everything we've discussed so far? Also, let's say a program "works"? Should we use public funds to scale it up so that more people can benefit? Does cost matter? Our objectives are as follows:

- to compare and contrast research, monitoring, and evaluation
- to explain how to calculate cost effectiveness

During the in-class activity you will bring together several new skills you've learned this semester and think through the process that every investigator goes through when submitting a new grant.

Required

1. Lecture
2. Rossi, P. H. et al. (2003). Evaluation: A Systematic Approach. Thousand Oaks, CA: Sage Publications. Chapter 1, "An overview of program evaluation".
3. Rossi, P. H. et al. (2003). Evaluation: A Systematic Approach. Thousand Oaks, CA: Sage Publications. Chapter 6, "Assessing and monitoring program progress".
4. Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2012). *Comparative cost-effectiveness analysis to inform policy in developing countries: a general framework with applications for education.*

Recommended

- Frankel, N. & Gage, A. (2007). *M & E Fundamentals*. Measure Evaluation. MS-07-20.

23. Implementation science

Have a great research finding? Great. Wait a decade or two and you might see it influence policy or practice. Implementation science is the study of how to close this gap. We'll discuss why some efficacious interventions are not effective when studied in "real world" settings and why scale up is so hard. Our objectives are as follows:

- to explain the research-to-practice gap
- to compare efficacy and effectiveness
- to discuss why scale-up can be challenging

Required

1. Lecture (combined with 22)
2. Remme, J. H. et al. (2010). Defining research to improve health systems. *PLoS Medicine, 7(11)*, e1001000.

24. Community based participatory research

There is a growing recognition that community involvement can lead to better designed interventions and more successful evaluations. In CBPR, the research process starts and ends with the community. In terms of end goals, social change is at least as important as the production of generalizable knowledge. In this sense, CBPR is *action oriented*. CBPR is not a method, per se, but a orientation to research. We'll discuss what this means and looks like in practice. Our objectives are as follows:

- to compare CBPR approaches to more traditional, investigator-driven models of research
- to outline the core principles of CBPR

Required

1. Lecture
2. Minkler, M., & Wallerstein, N. (2010). Introduction to CBPR: New issues and emphases. In Minkler, M., & Wallerstein, N. (Eds.). (2010). *Community-based Participatory Research for Health: From Process to Outcomes*. John Wiley & Sons. Available as an ebook through the Duke Library.
3. Israel, B. A. et al. (2010). Critical issues in developing and following CBPR principles. In Minkler, M., & Wallerstein, N. (Eds.). (2010). *Community-based Participatory Research for Health: From Process to Outcomes*. John Wiley & Sons. Available as an ebook through the Duke Library.

Recommended

- Leung, M. W. et al. (2004). Community based participatory research: a promising approach for increasing epidemiology's relevance in the 21st century. *International Journal of Epidemiology, 33(3)*, 499-506.
- Guijt, I. (2014). Participatory Approaches, *Methodological Briefs: Impact Evaluation 5*, UNICEF Office of Research, Florence.

25. Design thinking

Before an intervention can scale, it needs to demonstrate efficacy and effectiveness.† And before efficacy comes design—and often enough, a lot of failure. Working in global health means failing a good bit, so we'll use this session to discuss methods for learning from failure. The buzz words of today's lesson are "design thinking". This approach is "human-centered" in that it emphasizes the importance of empathizing with the intended user in order to create something meaningful and relevant to the context (fail here and you can expect to fall flat trying to scale). Other features of this approach include rapid iterating and testing. Basically, create a prototype (rough is ok) and start measuring. Find out what does not work, and repeat. Our objectives are as follows:

- to explain the need for design thinking
- to list the stages of design thinking

† *This is not entirely true. Sometimes customers and revenue suffice for evidence prior to scaling up. "What works" can become "what sells".*

Required

1. Lecture [in class]
2. Brown, T., & Wyatt, J. (2010, Winter). Design thinking for social innovation. *Stanford Social Innovation Review, 8*, 31-35.
3. Frog Design. *A stronger start*.

Recommended

- IDEO (2011). *Human Centered Design Toolkit, 2nd ed.* San Francisco: IDEO. Available for free download here.
- Denend, L., & Lockwood, A. (2014, Spring). Meeting the challenges of global health. *Stanford Social Innovation Review, 12*, 36-41.
- Taylor, M. J. et al. (2013). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Quality & Safety, 0*, 1-9.
- Sandhu, J. S. (2013). Measure early, measure often: rapid, real-time feedback in design for social innovation. *PopTech.*

26. Conflict, disaster, and research

In this session we'll discuss how to evaluate humanitarian assistance programs. Currently, you find much more "M" than "E" when it comes to studies of aid during and after humanitarian crises. These are challenging settings, and until recently, little work had been done to think how to implement rigorous methods in this context. Some people balk at the idea of "experimenting" with such vulnerable populations (whether randomization is used or not), but others point to our duty to ensure that assistance is as effective as possible. Our objectives are as follows:

- to list designs that may be appropriate for these settings
- to describe the ethical challenges of this work

Required

1. Lecture
2. Puri et al. (2014). What methods may be used in impact evaluations of humanitarian assistance? *3ie Working Paper 22*.

Recommended

3. Clarke et al. (2014). What evidence is available and what is required, in humanitarian assistance? *3ie Scoping Paper 1*.

27. Replication and reproducibility

Replication is a core component of the scientific method. No one study rules the day. If the results of your study are robust, another research group should be able to follow your methods and replicate the findings. Replications are relatively rare, however. For one, there are often few resources for replicating studies, especially when it comes to big field experiments. Second, journal space is limited (especially if there is still a print version) and peer review takes a lot of resources. Journals want to use their space and resources to publish novel ideas. Unfortunately, novelty can sometimes mean small effects with a lot of noise that might fail to replicate.

A separate but related issue is reproducibility, the ability to generate a study's findings given the original dataset and (hopefully) the original analysis code. Think this is rare? Rarely studied, maybe. But the Quarterly Journal of Political Science took on the challenge and found that slightly more than half of their published empirical papers subjected to review had results that could not be replicated with the author's own code. Presumably most of the time this happens we can blame mistakes. But not always, as you will see. Our objectives are as follows:

- to describe the role of replication in the scientific process
- to contrast replication and reproduction
- to test drive some open source tools for reproducible science

Required

1. Lecture
2. Mayer, M. N. & Chabris, C. (2014). Why psychologists' food fight matters. *Slate*.
3. Peng, R. D. (2011). Reproducible research in computational science. *Science, 334(6060)*, 1226.
4. Video: *Deception at Duke*. 60 Minutes (2012).

Recommended

- "Starter Set" Materials for the Duke Saga

28. Research ethics

In our final session, we'll review how shameful practices in the name of science prompted the field to put in place a number of safeguards to protect the rights of research participants. We'll consider how Institutional Review Boards function and what they look for in modern study designs. Our objectives are as follows:

- to outline basic ethical guidelines
- to explain the principle of informed consent

Required

1. Lecture
2. BRM 15, "Ethical issues in behavioral research".

FREQUENTLY ASKED QUESTIONS

How can we contact you?

Posting to the Sakai Forum is almost always the best option for asking a class-related question. I'm likely to reply quickly because everyone benefits from the post. Otherwise, send an email; please copy the course TA and include "Methods" in the subject.

How will you announce changes to the syllabus?

I'll post any substantial changes as announcements on Sakai and mention in class. Minor changes will just be pushed to GitHub.