Genetics, epigenetics, and Mendelian randomization: cutting-edge tools for causal inference in epidemiology

COURSE DURATION
This is an online distance learning course. Material will be available June 1st - 30th, 2019

INSTRUCTOR
Tomi Akinyemiju, PhD
tomiakin@uky.edu

Stella Aslibekyan, PhD
saslibek@uab.edu

Stephen Burgess, PhD
sb452@medschl.cam.ac.uk

COURSE DESCRIPTION
This course will provide participants with an in-depth understanding of Mendelian Randomization (MR) analysis, an approach developed to assess causal relationships in epidemiologic studies. We will teach participants the basis of instrumental variable analysis as well as the specifics on the MR approach, and provide practical skills for analyzing data and interpreting results within the MR context. A particular focus of the course will be on understanding how MR can be applied to consider the causal role of epigenetics.

Genetic susceptibility and epigenetic alterations have been implicated in a wide range of chronic diseases, including cancer, diabetes, Alzheimer’s and cardiovascular diseases. Epigenetic mechanisms have also been well studied as potential mediators of exposures on disease risk as well as downstream consequences of disease phenotypes. However, modifiable risk factors, including epigenetic and other biomarkers, are highly vulnerable to confounding and reverse causation. This course will provide in-depth training on the use of MR approaches to deal with confounding and reverse causation in epidemiologic studies, and outline various approaches to interrogate the causal relationships between exposures and disease outcomes.
PREREQUISITES
Previous training in biostatistics or statistics, including descriptive statistics and basic regression modeling, and basic genetics recommended. There are always exceptions, so feel free to contact us about your background to see if it is the right fit. Familiarity with R and/or STATA software will be helpful (although we will provide annotated programming codes for all exercises). We will use a lot of R in this class, which can be downloaded for free at http://www.r-project.org. R is an open source software and is available on all major platforms. You can find a virtually endless set of resources for R on the internet, including this Getting Started With R page. You may also be interested in using RStudio, an editor and development environment for R. If you are completely new to R, you should complete this online short course, Try R. An overview of genetics, including epigenetics, is available at this site.

COURSE OBJECTIVES

By the end of the course, participants will be able to understand:

1. The basic principles of causal inference in epidemiology and the role of confounding, reverse causation and measurement error
2. The basic principles of Mendelian randomization and the use of genetic variants as instrumental variables
3. How to perform Mendelian randomization analyses, understand the strength and limitations of genetic variants as instrumental variables, and interpret results
4. The basic concepts of epigenetic epidemiology, including the role of DNA methylation, histone modifications, and RNA-based mechanisms in disease processes
5. How to apply the principle of Mendelian randomization to evaluating DNA methylation as a causal risk factor
6. How to use publicly available Mendelian randomization tools and databases to conduct their own analyses

COURSE READINGS

There are no required readings but the following textbooks are highly recommended:

Causal Inference. Hernán, Miguel A. and James M. Robins. Forthcoming, Chapman & Hall/CRC Press (This book is still being written and draft PDFs are available at the linked page)


Introducing Epigenetics: A Graphic Guide. Cath Ennis and Oliver Pugh. Icon Books 2017

The following publications will be used for in-class discussion:

- Rothman and Greenland. Causation and causal inference in epidemiology


COURSE STRUCTURE
The online version of this course is meant to be a highly self-directed learning style that enables greater flexibility for course participants to complete the course objectives at the times and pace most conducive to the respective schedules of participants.

The course utilizes the learning management software, Canvas:  https://canvas.instructure.com/login

To get started, all registrants will receive an e-mail inviting them to join on the first day of the course offering. Upon receiving the e-mail, participants should follow the instructions to sign up for a Canvas account. This course is anchored around ~20 recorded lectures organized into three sections (Causal inference, Mendelian Randomization and Epigenetics). The recorded lectures are audio and screen recordings that will enable the instructor to teach key concepts, followed by applied sessions in the R and/or STATA software environment for applied examples of course concepts and hands on exercises. In general, each module will have a PowerPoint and video file, as well as annotated code files that can be used by participants and kept for future reference. There will also be paper discussions and an exciting 2-hour hackathon at the end of the course! We will alternate between structured lectures and hands-on applications of course material for each section.

COURSE OUTLINE

SECTION A- Causal Inference in Epidemiology (3 hours)

Module 1: Introduction
- Overview, course requirements, course outline
- Definition and measures of causal effect

Module 2: Randomized Experiments
• Randomized experiments
• Identification of Causal Effects under Randomization

Module 3: Issues in causal inference in observational studies
• Confounding
• Selection bias
• Measurement bias

Review and online discussion [1 hour]- Friday June 7

SECTION B: Mendelian Randomization (12 hours)

Module 1: Introduction to MR [2.5 hours]
• Motivation and introduction
• Estimation of a causal effect using individual-level data
• Some examples of MR
• Hands-on exercise: MR using individual-level data

Module 2: MR in practice [2.5 hours]
• Assumptions of MR
• Summarized data and two-sample MR
• Weak instruments and allele scores
• Hands-on exercise: MR using summarized data

Review and online discussion [1 hour]- Friday June 14

Module 3: Robust methods for MR [2.5 hours]
• Robust methods with summarized data
• Choosing instruments for MR
• Multivariable MR
• Hands-on exercise: Robust methods with summarized data

Module 4: Interpretation of findings [2.5 hours]
• Interpretation of MR estimates
• MR for drug development
• Paper discussion

Review and online discussion [1 hour]- Friday June 21

SECTION C: MR in Epigenetic Analyses (4 hours)

Module 1: Epigenetics and its discontents
• Intro to epigenetics
• What makes epigenetics a good (and bad!) fit for MR
• MR assumptions in the epigenetic context
• The need for the bidirectional approach

Module 2: Epigenetic markers as exposures or mediators
Module 3: Epigenetic markers as MR instruments
Module 4: MR resources
  • MR-base (http://www.mrbase.org/)
  • Choosing the right instrument for epigenetic studies: http://www.mqtldb.org/
  • Paper discussion- MR in the epigenetic context

Review and online discussion [1 hour]- Thursday June 27

SECTION D: Hackathon! (1 hour)-Friday June 28