Graduate Group in Epidemiology and Biostatistics
Advanced Topics in Epidemiologic Research (EPID 702)
Spring 2021
Syllabus

Course Description
The overarching goal of this course is to expose doctoral students in epidemiology to advanced epidemiologic and statistical research methods and theories that are limitedly or not otherwise covered in courses available in the curriculum. Topics that will be covered include reporting guidelines and best practices for reporting statistical methods and results, handling missing data, purposeful selection and application of propensity scores, selected topics in longitudinal and clustered data analysis, contemporary topics in statistical inference and use of p-values and other Frequentist statistical methods, Bayesian theory and inference, and topics selected in collaboration with students and the Graduate Group in Epidemiology and Biostatistics (GGEB) each term. This course is intended for doctoral students in the PhD program in Epidemiology. However, students from other graduate groups are welcome, as long as they meet the pre-requisites; such students are welcome during any year of study.

Pre-Requisites
1. EPID 526 and EPID 527 or equivalent.
2. EPID 701.
3. Permission from the instructor.

Leaning Objectives
Three learning objectives have been developed for this course; (i) provide students with an understanding of modern and cutting-edge quantitative methods, advanced topics, and best practices in epidemiologic, statistical, and biomedical research; (ii) develop students competence and confidence in statistical programing to support accurate and reproducible epidemiologic and biostatistical analyses; (iii) improve the ability of students to make informed decisions regarding the selection of analytic methods in their individual and collaborative research projects.

Core Competencies
This course emphasizes the following core competencies: knowledge within program area (epidemiologic and biostatistical methods); research skills (study planning, critically appraising published research); quantitative and computational methodologies (data manipulation, data analysis, statistical coding and debugging, Bayesian inference, data visualization, purposeful statistical inference, and model selection).

Scientific Rigor and Reproducibility
Through technical lectures, reading of carefully selected peer-reviewed tutorials, critical appraisal of published research studies, and in-class statistical coding laboratory sessions, this course will provide instruction on rigorous and informed statistical model selection, estimation, and interpretation.

Learning Outcomes
After completing this course, students will be able to:
- Critically appraise the application of several advanced methods they will encounter in the epidemiologic and biomedical literature.
- Describe current approaches to epidemiologic research and statistical inference.

- Make informed decisions regarding the selection of analytic methods in their individual and collaborative research projects.
- Provide sophisticated peer-review assessments to leading medical and scientific journals.

Class Format
This is primarily a lecture course with in-class laboratory sessions. The course will be 1-credit and meet once a week for 3 hours for a total of 14 weeks. The course will be offered every spring semester with the class scheduled on Tuesday from 9:00am-12:00pm. For each week, selected readings will be provided by the instructor. A total of 9 homework assignments will be given. They will be graded. The purpose of the homework assignments is to acquire mastery of the statistical concepts used in research applications and the use of the computer software Stata and R for statistical calculations and presentation. The focus of the course will be on the use, interpretation, and concepts rather than on memorization of formulae. The final exam may include a take-home project and in-class presentation. Students are expected to attend all class sessions and actively participate and will be graded on their participation.

Rationale/Additional Comments for the Proposal for the Biomedical Curriculum and Student Academic Standards Committee – BCC
This course is being designed to expand the curriculum and methodologic exposure of PhD students in epidemiology and is thus specifically tailored to PhD students. The creation of this new course will provide curricular support specific to the PhD program in Epidemiology, increasing the quality and breadth of training for the PhD students and hopefully attracting future applicants seeking a more expansive PhD training program.

Participating Faculty
Course director and facilitator: Michael O. Harhay, PhD mharhay@pennmedicine.upenn.edu
Guest lecturers will be considered on an ad hoc basis.

Course Units
This is a 1.0 course unit course.

Contact Hours
The course will consist of one three-hour classroom session each week. In addition, a discussion board will be established on Canvas for out-of-class communication.

Course Structure
The course will consist of a series of weekly sessions, each of which will focus on a primary theme.

Grading
There will be a total of 10 problem sets that require the application of the method(s) or concepts taught in the weekly lectures, accounting for 7.5% each and 75% of the total course grade. The final 25% will reflect the student's completion of a final project. The final project will entail the students finding a method not covered in the course, providing a written statistical review of the method from 2 separate peer-reviewed published papers, and a class presentation at the end of the course that is meant to teach the method to the class.

Course and Topic Schedule

Lecture	Topic	Learning objectives

1	Statistical reporting and communicating statistics	- EQUATOR Network guidelines
		   o STROBE for observational studies
		   o RECORD statement for observational data
		   o PRIMSA for systematic reviews
		- PROSPERO
		- SAMPL (Statistical Analyses and Methods in the Published Literature) Guidelines
		- Elements of high-quality peer-review
		
		- **Laboratory 1**: Statistical peer review

2	Missing data 1	- Missing data patterns
		   o Missing completely at random (MCAR)
		   o Missing at random (MAR)
		   o Missing not at random (MNAR)
		- Best practices for MCAR and MAR
		- Visual approaches to assessing the quality of multiple imputation
		
		- **Laboratory 2**: Multiple imputation in Stata

3	Missing data 2	- Examples of MNAR in the epidemiologic literature
		   o Dropout, informative missingness, informative truncation
		- Approaches to dealing with MNAR data
		   o Principal stratification
		   o Inverse probability of censoring weighting
		   o Shared parameter (joint) models
		
		- **Laboratory 3**: Estimating joint models in Stata

4	Development and validation of prediction models	- TROIPOD statement and checklist for prediction model development and validation - PROBAST tool for risk of bias assessments of prediction models - Best practices for prediction model development - **Laboratory 4**: Implementation of the '*tripod'* command in Stata

5	Estimands from different propensity score methods	- Best practices for propensity score estimation
		- Empirical meaning of different propensity score methods
		   o matching (and types of matching)
		   o stratification
		   o inverse probability of treatment weighting
		   o doubly robust methods
		
		- **Laboratory 5**: Implementation of the '*teffects'* command in Stata

6	Advanced topics in survival models	- Competing risks models - Accelerated failure models

		- Multi-state models - Shared frailty models - Parametric survival models - **Laboratory 6**: Advanced survival modeling in Stata
7	Selected topics in longitudinal modeling	- Basic elements of random effects, fixed effects, sandwich estimators, and GEE models - Modeling change scores - Phenotyping similar trajectories - **Laboratory 7**: Implementation of the "traj" command in Stata
8	Selected topics in clustered data	- Assessment of clustering - Best practices to handling clustered or multisite data - Decomposing differences between sites - **Laboratory 8**: Calculating and displaying risk-standardized mortality by hospital/site in Stata
9	Modeling trends	- Direct and indirect standardization - Joint point analysis - Non-linear trends - Differences-in-differences - **Laboratory 9**: Modeling trends and differences-in-differences in Stata

10	Moving to a world beyond "$p < 0.05$"	- Best practices for inference using Frequentist statistics
		- *American Statistical Association* guidance on p-value use
		- Selected topics from the 2019 series on beyond the p-value published in *The American Statistician*
		    o P-value functions
		    o Second Generation p-Values

11	Bayesian inference 1	- Basic concepts in Bayesian inference - Bayes factors versus p-values - Best practices in prior selection
12	Bayesian inference 2	- Additional examples of Bayesian analysis using observational and randomized trial data - **Laboratory 10**: Implementation of a Bayesian analysis. Students will replicate Harper, 2019, *Epidemiology* "Would Stronger Seat Belt Laws Reduce Motor Vehicle Crash Deaths?: A Semi-Bayesian Analysis."
13	Student presentations	
14	Student presentations	