SPPH 504 (Section 007): Application of Epidemiological Methods

Time and Place: Term 1, Tuesdays, 9 am-noon

Location: SPPH 143, School of Population and Public Health, University of British Columbia, 2206 East Mall, Vancouver BC

Instructors:

- Ehsan Karim, Room 145, 2206 East Mall, School of Population and Public Health, University of British Columbia, 604-682-2344 ext. 64251, ehsan.karim@ubc.ca
- Mieke Koehoorn, Room 293, 2206 East Mall, School of Population and Public Health, University of British Columbia, 604-822-5756, mieke.koehoorn@ubc.ca

Office Hours: (time/location TBD)

Teaching Assistant: TBD

Lab Hours: One-hour weekly lab (day/time/location TBD)

Pre-requisites: (currently restricted to SPPH PhD students only)

- SPPH400 Statistics for Health Research, and
- SPPH500 Analytical Methods in Epidemiological Research, and
- SPPH502 Epidemiological Methods I, and
- one of
    - SPPH503 Epidemiological Methods II, or
    - SPPH506 Quantitative Research Methods, or
    - SPPH519 Qualitative Methods for Health Research Design, or
    - SPPH530 Epidemiology of Occupational and Environmental Health

Course Overview: The purpose of this course is to provide students with learning opportunities to understand fundamental epidemiological concepts through the application of methods using population and public health datasets. The purpose is also to introduce students to emerging epidemiological methodologies that are frequently being applied to population and public health-related research questions in prestigious epidemiology journal publications.

Course Objectives: By the end of this course, students will develop practical skills to:

- Design an appropriate study for a population/public health research question that will help reduce bias;

- Create an analytic dataset from a complex population/public health survey to answer an epidemiological research question, including statistical programming/coding and data management/documentation skills;
- Recognize the scenarios where various data analysis methods are appropriate to answer a research question;
- Apply an appropriate data analysis method to answer a population/public health research question;
- Interpret estimates and conclusions from data analysis methods;
- Describe the assumptions and limitations of data analysis methods;
- Communicate an epidemiological study (design, analytic method, findings and discussion (including issues related to limitations, biases)) in a format suitable for submission to an academic health journal; and
- Present an epidemiological study and analytic method in a format suitable for an academic health conference.

Student Evaluation: The course will involve pre-class reading materials (journal article with short pre-class video introduction focused on the application of epidemiological methods), lectures based on discussion of the pre-class reading content, and in-class activities and course assignments focused on the application of epidemiological methods. Although students will gain some expertise in statistical computation and programming, this course is focused on the application of epidemiologic analytic methods. Students will be evaluated based on the following elements: (a) understanding of key epidemiologic concepts, (b) understanding of analytic approaches to reduce study biases, (c) the application of epidemiological methods to population and public health research questions, and (d) the appropriate interpretation of analytic estimates from analytic output.

• Pre-class survey of pre-reading materials	5%
• In-class data analysis activities	5%
• Assignments (3 in total; 15% each)	45%
• In-class presentation	10%
• Final Paper	35%

Description of the Assignments:

a) **Pre-class survey**: The surveys consist of several multiple choice questions assessing understanding of pre-class reading and video materials. Pre-class readings and videos are focused on the application of an epidemiological method to a population/public health issue/research question that will be explored further in class through lectures, instructor demonstrations, in-class student application activities.

b) **In-class data analysis activities**: Students will complete in-class data analysis exercises to apply concepts covered in pre-class reading materials/videos, and in instructor lectures/in-class demonstrations.

c) **Assignments**: The instructor will provide a list of potential research questions at the beginning of the course. Students will select one research question to work on during the term that will be the focus of the three course assignments and the final paper. Students may propose their own research questions with instructor approval. Students will be evaluated on the design of a study and choice of analytic method to answer their research question that reduces bias, the application of their analytic method to a conplex survey or epidemiological dataset, the interpretation of their analytic output, and the discussion of the assumption/limitations of their epidemiological method. In each assignment there will be one part pertaining to the final report. These parts will serve as inputs to the students' final report during the term, as follows: Assignment 1 – Introduction Section of Final Paper with a focus on the design of an appropriate study for a population/public health research question that will help reduce bias; Assignment 2 – Methods section with a focus on the selection of an appropriate data analysis method to answer a population/public health research question; and Assignment 3 – Results section with a focus on the interpretation of analytic outoput for their research question. The second half of each of the assignment will be related to the continuation of the in-class data analysis activities.

d) **In-class presentation**: Students will present the analytic approach for their research question with a focus on the justification of the choice of approach, in a format suitable for an academic health conference.

e) **Final Paper**: The Final Paper will have 5 sections: (i) objective and motivation of the study with a brief literature review, (ii) methods description, (iii) bivariable and multivariable results, (iv) discussion outlining strengths and limitations of the study, and (v) appendix with project /statistical analysis coding (maximum 3,000 words for sections i through iv).

<u>Assignment Due Dates</u>:

- Pre-class surveys are due by 4 pm the day before each class (1st and 7th classes do not have any pre-reading materials or surveys, worth 5% of the total grade).
- September 25th - Research proposal for instructor feedback (P/F)
- October 2nd - Assignment 1 (worth 15% of the total grade)
- October 15th - In-class Presentation slide submission
- October 16th - In-class Presentation (worth 10% of the total grade)
- October 30th - Assignment 2 (worth 15% of the total grade)
- November 20th - Assignment 3 (worth 15% of the total grade)
- December 11th - Final Report (worth 35% of the total grade)

Late Assignments: Pre-class survey answers, Assignments 1 to 3, and the Final Paper must be submitted via Canvas. Typically, no late pre-class survey answers or assignments will be accepted. Extensions of the due date for the assignments will be considered pending extenuating circumstances with the approval of the instructor. The instructor will require documentation of extenuating circumstances (medical certificates, etc.). Assignments submitted later than the due date will be penalized 10% of the possible grade for each day past due.

Plagiarism: Students are expected to review the Student Discipline section of the UBC Calendar, available online at http://www.calendar.ubc.ca/vancouver/index.cfm?tree=3,54,111,959 and know what constitutes plagiarism and academic misconduct, and that such activities are subject to penalty.

Grading *(from the UBC Department of Educational Studies, Graduate Course Grading Policy, D. Pratt):*

- **A Level (80% to 100%):** A+ is from 90% to 100%: It is reserved for exceptional work that significantly exceeds course expectations. Also, achievement must satisfy all the conditions below. A is from 85% to 89%: A mark of this order suggests a very high level of performance on all criteria used for evaluation. Contributions deserving an A are distinguished in virtually every aspect. They show that the individual (or group) significantly shows initiative, creativity, insight, and probing analysis where appropriate. Further, the achievement must show careful attention to course requirements as established by the instructor. A- is from 80% to 84%: It is awarded for the high quality of performance, no problems of any significance, and fulfillment of all course requirements.
- **B Level (68% to 79%):** This category of achievement is typified by adequate but unexceptional performance when the criteria of assessment are considered. It is distinguished from A-level work by problems such as one of the more significant errors in understanding, superficial representation or analysis of key concepts, the absence of any special initiatives, or lack of coherent organization or explanation of ideas. The level of B work is judged by the severity of the difficulties demonstrated. B+ is from 76% to 79%, B is from 72% to 75%, and B- is from 68% to 71%
- **C Level (55% to 67%):** Although a C+, C, or C- grade may be given in a graduate course, the Faculty of Graduate Studies considers 68% as a minimum passing grade for doctoral graduate students.

Course Materials:

- All course materials (lecture slides, pre-class readings and video, assignments) will be posted on the Canvas course website.
- Required pre-class readings (listed at the end of this document) can be downloaded from either the Canvas course website or from the UBC library website.

- Access to a computer with R software (either R or RStudio, free of charge) is recommended for all course work. Students are also recommended to bring their own laptop computer to class with the R software installed, if possible. Students may choose to use other software (such as SAS, STATA, etc.) for their coursework, assignments and final report, but course instruction will be provided in R.

Data Sources (public use population/public health survey and epidemiological datasets to be used in various parts of this course – demonstrations, in-class application exercises, student assignments/final paper):

- Canadian Community Health Survey - Annual Component (CCHS): http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226
- National Health and Nutrition Examination Survey (NHANES) https://www.cdc.gov/nchs/nhanes/index.htm
- National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 (ICPSR 21600) http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/21600
- Vanderbilt Biostatistics Datasets http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets
- Additional datasets https://vincentarelbundock.github.io/Rdatasets/datasets.html

Course Summary:

Week #	*Topic(s)*	*Pre-class Reading list*	*Assignment due*
# 1 (Sept 4)	• **Outline of the class and review** ○ Review of survey designs ○ Introduction to complex survey designs	•	
# 2 (Sept 11)	• **Hands-on data analysis: demonstration and practice - session 1** ○ Working with CCHS, NHANES dataset, sampling weights ○ Summaries, Graphs	Korn et al. (1991)	
# 3 (Sept 18)	• **Concepts of analyzing complex survey data 1: Linear regression** ○ Regression adjustments for continuous outcomes ○ Variable selection criteria ○ Machine learning methods: Least absolute shrinkage and selection operator (LASSO) / elastic net	Brookhart et al. (2010)	

Week #	*Topic(s)*	*Pre-class Reading list*	*Assignment due*
# 4 (Sept 25)	• **Hands-on data analysis: demonstration and practice - session 2** ○ Fitting linear regression using NHANES dataset ○ Fitting LASSO (continuous outcome) using NHANES dataset	Katz et al. (2003)	Selection of a final project topic.
# 5 (Oct 2)	• **Concepts of analyzing complex survey data 2: Logistic regression and categorical data analysis** ○ Regression adjustments for binary & categorical outcome	Peng et al. (2002)	Assignment 1 submission at the beginning of the class
# 6 (Oct 9)	• **Hands-on data analysis: demonstration and practice - session 3** ○ Fitting logistic regression using NHANES dataset ○ Fitting LASSO (binary outcome) on NHANES dataset	Hosmer et al (1991)	
# 7 (Oct 16)	• **Student Presentations** (related to final report) Presentation of research question with rationale for analytic approach, plus preliminary results (8-10 min each student)		Slide submission on Oct 15th
# 8 (Oct 23)	• **Concepts of reducing confounding in complex surveys** ○ Matching on variables ○ Propensity score methods ○ Disease Risk Score (DRS)	Austin (2011)	
# 9 (Oct 30)	• **Hands-on data analysis: demonstration and practice - session 4** ○ Matching using NHANES ○ Fitting PS and DRS on NHANES dataset	Brookhart et al. (2006)	Assignment 2 submission at the beginning of the class
# 10 (Nov 6)	• **Concepts of missing data analysis** ○ Chained equations for imputation ○ Tree-based approaches ○ Random Forest for imputation	Sterne et al. (2009)	

Week #	*Topic(s)*	*Pre-class Reading list*	*Assignment due*
# 11 (Nov 13)	• **Hands-on data analysis: demonstration and practice - session 5** o Applying various missing data imputation approaches using NHANES dataset	Shah et al. (2014)	
# 12 (Nov 20)	• **Concepts of survival analysis (time-dependent exposure and confounders)** o Immortal time bias o Marginal structural models (MSM)	Suissa (2007)	Assignment 3 submission at the beginning of the class
# 13 (Nov 27)	• **Hands-on data analysis: demonstration and practice - session 6** o Fitting Cox regression and MSM using NHANES o Open discussion (on topics voted by students)	Cole et al (2008)	
(Dec 11)	No class		Final Report

Required Re-Class Readings: (subject to change)

Papers are accessible via UBC Library or Canvas:

1) Korn, E. L., & Graubard, B. I. (1991). Epidemiologic studies utilizing surveys: accounting for the sampling design. American Journal of Public Health, 81(9), 1166-1173.
2) Brookhart, M. A., Stürmer, T., Glynn, R. J., Rassen, J., & Schneeweiss, S. (2010). Confounding control in healthcare database research: challenges and potential approaches. Medical Care, 48(6 0), S114.
3) Katz, Mitchell H. Multivariable analysis: a primer for readers of medical research. Annals of Internal Medicine 138, no. 8 (2003): 644-650.
4) Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. The Journal of Educational Eesearch, 96(1), 3-14.
5) Hosmer, D. W., Taber, S., & Lemeshow, S. (1991). The importance of assessing the fit of logistic regression models: a case study. American Journal of Public Health, 81(12), 1630-1635.
6) Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46(3), 399-424.

7) Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. American Journal of Epidemiology, 163(12), 1149-1156.
8) Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ, 338, b2393.
9) Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. American Journal of Epidemiology, 179(6), 764-774.
10) Suissa, S. (2007). Immortal time bias in observational studies of drug effects. Pharmacoepidemiology and Drug Safety, 16(3), 241-249.
11) Cole and Hernán. Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology, 2008;168(6):656-64.

Additional Readings:

There is no required textbook. The following textbooks are suggested for further reading (also available via UBC library):

1. Heeringa, S.G., West, B.T., Berglund, P.A (2017) Applied Survey Data Analysis, Second Edition, Taylor & Francis, Florida.
2. Lumley, T. (2010) Complex Surveys: A Guide to Analysis Using R, John Wiley & Sons, Inc., New Jersey.
3. Lewis, T.H. (2016) Complex survey data analysis with SAS, Taylor & Francis, Florida.