

# Course Syllabus for SPPH 500 – Analytical Methods in Epidemiological Research Section 007, January, 2022

-  
**Instructor:** Hubert Wong

**Email:** [hubert.wong@ubc.ca](mailto:hubert.wong@ubc.ca)

**Office - UBC:** SPPH 145

**Phone - St. Paul's Hospital:** 604-806-8443

-

## Calendar Description

Basic epidemiological designs as a framework for commonly used biostatistical techniques such as the Mantel-Haenszel, chi-squared, linear and logistic regression, and survival analysis. Computer packages will be available for computation of assignments.

-

## Learning Objectives

This course is intended to provide a solid understanding of the concepts, principles, and interpretation of multivariable regression modeling. These key principles will be covered using a combination of examples, review of journal articles and relevant case studies. Students will be required to effectively communicate statistical concepts, interpret statistical models, critically evaluate the application of these models, and present findings. On completion of this course the student should be able to:

1. Distinguish between different objectives of a regression analysis;
  2. Understand the principles of model development for each objective;
  3. Relate regression models (linear, logistic, Poisson, survival) to appropriate types of health outcome data;
  4. Interpret parameters of each regression model and summarize results for scientific journals;
  5. Critically evaluate the analytic methods sections in health care literature;
  6. Communicate effectively with statistical practitioners on regression methods.
- 

## Pre-requisites

SPPH 400 (or SPPH 567), SPPH 502, or permission of the Instructor

-

## Course Format

In-class seminars Mondays 1:00pm – 4:00pm from January 10, 2021 to April 4, 2021 excluding February 21.

The typical session will comprise: (1) presentation of new topics by the instructor with interactive discussion, (2) small-group work applying the concepts, and (3) share-back by groups and summarizing take-home points.

Statistical software will be used to support illustration of the modeling process; students are encouraged to use statistical software to facilitate learning but assignments will not require its use.

### Reference Texts

Vittinghoff E, Glidden D, Shiboski S and McCulloch C. *Regression Methods in Biostatistics, 2nd ed.* Springer 2011. (primary course text; available online through the UBC library)

Rothman, KJ, Greenland S and Lash TL. *Modern Epidemiology, 3rd ed.* Wolters Kluwer 2008.

Gelman A, and Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge 2007.

### Assessment, Evaluation and Grading

**Class participation (17%):** Students are expected to have read the designated literature prior to class and be prepared to contribute actively to the discussion. Note that there is no such thing as a “stupid question” – if something is unclear, students are encouraged to speak up – the instructor will re-direct if it is a topic that is best dealt with outside of class time.

**Class summaries (8 summaries x 6% each = 48%):** Following each class, students will prepare a summary (no more than 1000 words, say) of the most important 3 to 5 “take-home points” from the class. Each point should be accompanied by a short explanation of the underlying rationale. For example, the point “Do not choose adjustment variables for a causal model based on p-values” should be accompanied by an explanation such as “The p-value does not tell you whether bias is introduced by including/excluding the variable.” **These summaries are due 11:59pm on the Thursday following the class (the aim is to have the summaries marked and returned prior to the next class). Late assignments are not accepted. Please submit through Canvas in a Microsoft Word compatible format to enable addition of comments.** Assessment of these summaries will be based on accuracy of points made and clarity of exposition.

**Final project (presentation 10% + write-up 15% + peer-critique 10% = 35%):** The final project will be a critique of a published journal article (chosen by students working in pairs; approved by the instructor) that involved a substantial amount of regression modeling. **It is comprised of a presentation during one of the last two classes, a write-up due one week after the last class, and a written critique of one submission from your peers.**

## Course Topics and Required Readings

### Session 1: (Jan 10)

Introduction: Course outline, objectives

Review of basic concepts and definitions in statistics (population, sample, parameter, statistic/estimator, etc)

Key objectives of statistical modeling in health research (data summarization, prediction, causal effect estimation)

Review of basic epidemiologic concepts and measures (RD/RR/OR, confounding, effect measure modification, direct/indirect effects, etc)

General approach and principles in data analysis, role of design and sampling

### Session 2: (Jan 17)

Simple and multiple linear regression, basics

Ref: Vittinghoff 4.1 – 4.2

Analysis of variance and analysis of covariance

Ref: Vittinghoff 4.3

### Session 3: (Jan 24)

Modeling variable interactions/effect modification

Ref: Vittinghoff 4.6

Data transformations; guidelines and impact on model interpretation

Ref: Vittinghoff 4.7

### Session 4: (Jan 31)

Principles of variable selection for estimating causal effects: confounding bias, causal graphs, adjusted effect, multicollinearity considerations, direct vs indirect effects

Ref: Vittinghoff 10.2 – 10.6, Rothman 12

Principles of variable selection for prediction: prediction errors, over-fitting and bias-variance trade-offs, AIC, cross-validation, practical considerations

Ref: Vittinghoff 10.1, 10.2 – 10.6

Interpreting and reporting linear regression models

### Session 5: (Feb 7)

Logistic regression, basics

Ref: Vittinghoff 5.1 – 5.4, 5.8

Regression coefficients; relationship to contingency tables

Assumptions and categorical variables

Ref: Vittinghoff 5.1

### Session 6: (Feb 14)

Logistic regression case study – estimating a causal effect, omitted variable bias

Logistic regression case study – developing a predictive model, sensitivity, specificity, ROC curves

Ref: Vittinghoff 5.2.5 – 5.2.6

**Session 7:** (Feb 28)

Matched case-control studies; conditional logistic regression

Ref: Vittinghoff 5.3

Poisson regression

Ref: Vittinghoff 8.1 – 8.3

**Session 8:** (Mar 7)

Time to event data, censoring, hazard rates & ratios, Kaplan-Meier method

Ref: Vittinghoff 3.5

Cox proportional hazards model

Vittinghoff 6.1, 6.2, 6.4

Time-dependent exposures, more on censoring

Ref: Vittinghoff 6.3, 6.6 – 6.9

**Session 9:** (Mar 14)

Accounting for non-simple random sampling

Ref: Vittinghoff 12.1 – 12.6

Modeling hierarchical/clustered/longitudinal data, basics

Ref: Vittinghoff 7.1 – 7.12, Gelman Ch 12

**Session 10:** (Mar 21)

Approaches to missing data

Ref: Vittinghoff 11.1 – 11.9

Causal inference (by guest lecturer)

**Session 11:** (Mar 28)

Student presentations

**Session 12:** (Apr 4)

Student presentations

.

**UBC's values and policies**

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious and cultural observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest

academic standards in all of their actions. Details of the policies and how to access support are available at: <https://senate.ubc.ca/policies-resources-support-student-success>