CLASS SESSIONS

3 lectures per week and 2 additional periods for recitations/computer labs.
Days: TBA

INSTRUCTORS

Cohort 1	Jeanine Genkinger, PhD	jg3081@columbia.edu
Cohort 2	Melissa Begg, ScD	mdb3@columbia.edu
Cohort 3	Martina Pavlicova, PhD	mp2370@columbia.edu
Cohort 4	Katherine Keyes, PhD	kmk2104@columbia.edu

COURSE DESCRIPTION

This course will provide an integrated approach to the disciplines of Biostatistics and Epidemiology. It will have a large group didactic component, smaller breakout groups for recitation as well as for hands on data analysis. We start by discussing the complexities of public health challenges, as introduced in the Systems Thinking module. In that module, first year MPH students will be presented with a public health (as compared to a medical/individual level) model of interventions and health; reflecting that each phenomenon is related to a web of interconnected elements, at the genetic, individual, family, neighborhood, community, environmental, governmental, and societal levels, each with dynamic feedback mechanisms. If Public Health is not a simple, reactive, "take the pill three times a day" solution, but a purposeful approach to preventing disease and promoting health, then being able to document, measure and understand all the consequences becomes imperative. Quantitative methods introduced in this course begin to provide some of the tools necessary, in addition to those introduced in the Qualitative Foundations module, to help estimate the relationships between the smaller pieces that comprise the complex and dynamic web of systems in Public Health.

REQUIRED TEXTS

Gordis, L. Epidemiology, fourth edition. 2009. Saunders/Elsevier. IBSN: 1416040021.
Pagano, M, Gauvreau, K. Principles of Epidemiology, Second Edition. Duxbury. IBSN: 0534229026.

RECOMMENDED TEXTS

Gonick L, Smith W. The Cartoon Guide to Statistics. 1993. Harper Peerennial. IBSN: 0062731025.

OVERALL QUANTITATIVE COURSE LEARNING OBJECTIVES

Through a balanced mixed of didactic and interactive lecturing, small group discussion of issues and examples, hands-on use of analytic software, and targeted assignments, by the end of the Quantitative Core Course, students will be able to:

- Describe the roles that the quantitative disciplines of Epidemiology and Biostatistics serve in biomedical research.
- Calculate and interpret measures of health and disease including incidence and prevalence.
    - Compute and evaluate disease parameters and trends at a population level, as well as basic definitions such as epidemic, pandemic, and endemic.
    - Compute, describe, and compare joint, marginal, and conditional probabilities of measures of disease burden.
    - Distinguish between and calculate basic measures of disease frequency, including: incidence rate, prevalence, cumulative incidence, mortality rate, and case fatality rate.
- Apply principles of screening and diagnostic tests, and recognize their strengths and limitations.
    - Describe how screening tests impact on outcome assessment at the population level.
    - Calculate and interpret properties of screening tools such as sensitivity, specificity, positive and negative predictive values (PPV).

    - Describe the effect that the underlying prevalence of the disease has on the performance of these properties.
- Translate research objectives into clear, testable hypotheses.
    - Specify the correct form of hypotheses, and name the key elements of scientific hypotheses: the independent variable, dependent variable, and population of interest.
    - Describe the specific and precise operationalizations of those elements.
    - List the limitations that inevitably result from the choice of those operationalizations.
- Apply general principles of study design to be able to evaluate and select appropriate designs, including randomized controlled trials, cohort, case-control, and ecological studies, to best address selected hypotheses.
    - Distinguish between the application of "best case" study design, and those that are most feasible.
    - Identify the time, resources, and efforts needed to implement, monitor and complete each type of design.
- Describe the limitations of various study designs, and assess each design's ability to allow us to infer causation.
    - Identify the elements of each study design that allow or preclude the ability to infer causation and to calculate various estimates of the population risk and of association.
- Examine the sources of bias within the context of human population studies
    - Define and describe selection and information bias
    - Define and describe confounding
- Assess reliability, validity, and suitability of data sources needed to address specific testable hypotheses.
    - Calculate and interpret measures that inform the reliability or reproducibility of each variable or measure.
    - Describe the factors that improve or reduce the accuracy of each variable or measure.
- Apply numerical, tabular, and graphical techniques to characterize and summarize public health data.
    - Prepare and interpret summary measures and graphs using frequencies, means, medians, counts, standard deviations, and other measures of central tendency and dispersion.
- Describe basic principles of statistical inference, including Type I and Type II Errors, Power, and detectable effect sizes.
    - Illustrate the relationship between these components, particularly between changing sample size and detectable effect size.
    - Define the difference between statistical significance and clinical importance.
- Apply appropriate bivariate inferential statistics (e.g., Chi-squared tests and t-tests), interpret the results from these methods, and recognize the limitations.
    - Select a correct analytic method based on the type of data, and assess if assumptions have been satisfied.
    - Describe the implications of violating the assumptions that underlie various methods.
    - Determine subsequent analytic steps based on results of bivariate tests.
- Calculate, apply, and interpret measures of association between exposures and outcomes including the rate ratio, risk ratio, odds ratio, risk difference, attributable risk, mean difference, and population attributable risk.
    - Demonstrate ability to assess the design circumstances under which it is appropriate to calculate each measure.
- Describe the use and applicability of statistical models for describing relationships among multiple variables.
    - Describe the advantages and limitations of statistical models.
    - Distinguish between the appropriate applications of linear versus logistic regression models.
- Employ computer software to conduct basic analyses, and produce and interpret output.
    - Demonstrate ability to move from data entry to data analysis.
    - Create and read statistical output, and interpret results.
- Critique published public health research, and identify strengths, weaknesses (including bias and confounding) and limitations to interpretation and inference.
    - Integrate accumulated course knowledge to organize and assess information as presented in relevant peer reviewed publications.
    - Characterize the uses and application of epidemiology, biostatistics, and quantitative analysis in setting public health policy.

COURSE STRUCTURE

Integrated quantitative concepts related to the disciplines of Biostatistics and Epidemiology will be presented, followed by applied Public Health examples, related readings, and appropriate homework. Homework will include the use of statistical software STATA. Students will be responsible for becoming familiar with STATA through the Computer Lab sessions. STATA is available for student purchase at a significant discount; visit http://www.stata.com/order/new/edu/gradplans/gp-direct.html for more information. Note that "Small STATA" is sufficient for the needs of this course. It is also available for use on PCs in the library and in the Student Learning Center (LC 17-107).
The recitation component of the course reviews material discussed in lecture. Recitations are highly effective in helping students to understand concepts and apply them to problems.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Academic Integrity
Students are required to adhere to the Mailman School Honor Code, available online at http://mailman.columbia.edu/honorcode.

Disability Access
In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu. If you have already registered with ODS, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Lillian Morales (lm31@columbia.edu), the School's liaison to the Office of Disability Services.

COURSE SCHEDULE

PLEASE READ THE LECTURE SECTION OF COURSEWORKS TO DOWNLOAD THE READINGS, EXAMS, AND LECTURE SLIDES.

WEEK OF CORE	DATE	Topic	Readings	Assignments
3	**9.17.12**	Framing concepts of epidemiology and biostatistics; History/notable events in epidemiology.	• Gordis chapter 1 • Gerstman 1.1 • Gonick Chapter 1	
	9.19.12	Measures of disease occurrence and frequency (incidence, mortality, fertility, prevalence, cumulative incidence, survival time or time-to-event); disease surveillance.	• Gordis chapters 2 and 3	
	9.21.12	Types of data (categorical, continuous, count, rate, person-time, time-to-event) and properties of measurements (reliability, validity); Descriptive Statistics (numerical: mean, median, sd and graphical: histogram, boxplot, scatterplot).	• Gerstman chapters 1.2-1.4, 3 and 4 • Gonick Chapter 2	
4	**9.24.12**	Intro to probability, definition of conditional probability, law of total probability, Bayes theorem	• Gerstman chapters 5.1, 5.2 and 5.5 • Gonick Chapters 3 and 4	
	9.26.12	Epidemic modeling, reproductive rate, herd immunity, and methods in outbreak investigations.	• Giseke readings*	Homework 1 due
	9.28.12	Specification of a research hypothesis, and key elements of the hypothesis, including the independent variable, dependent variable, and population of interest; Distinctions among the study population, target population, and study sample. Operationalization of each element of the hypothesis.	• Gerstman chapters 9.1, 2.1 • Gonick Chapter 6	
5	**10.1.12**	What is a cause? Introduction to the sufficient-component cause framework and Hill's causal criteria	• Gordis chapter 14	
	10.3.12	What gets in our way of detecting causes? Introduction to confounding and bias.		Homework 2 due

	10.5.12	Introduction to Distributions: Binomial, Poisson, Normal, Standard normal distribution, effect of standardization	• Gerstman chapters 5.3, 5.4, 6 and 7 • Gonick Chapter 5	
6	**10.8.12**	Randomized Clinical Trials: design, purpose of a placebo and understanding comparison groups, blinding, ethics, sample size.	• Gordis chapter 7 • Gerstman 2.2	
	10.10.12	Potentials for bias in a clinical trial, non-compliance, intention to treat analysis, generalizability, and single vs multi-site trials.	• Gordis chapter 8	Homework 3 due
	10.12.12	Central limit theorem, normal approximation to binomial distribution, normal approximation of X-bar, standardization of X-bar, standardization of p.	• Gerstman chapter 8	
7	**10.15.12**	Concepts of screening (natural history of disease, critical point, screening vs. surveillance, simultaneous vs. sequential screening); sensitivity, specificity, positive predictive value, negative predictive value.	• Gordis chapter 5	
	10.17.12	Introduction to confidence intervals inference, confidence interval for proportion, mean with known and unknown variance, t-distribution.	• Gerstman 10.1-10.3, 11.4, 16.5 • Gonick Chapter 7	Homework 4 due
	10.19.12	Observational Cohort Studies (design features); Exposure construct (when you cannot randomize); measures of association (risk ratio and risk difference, rate ratio and rate difference)	• Gordis chapters 9 and 11	
8	**10.22.12**	Occupational epidemiology as a case study illuminating principles of cohort studies, choosing comparison groups. Direct and indirect standardization		
	10.24.12	Introduction to hypothesis tests, philosophy behind hypothesis testing, steps of hypothesis testing, conclusions. Hypothesis testing for 1-sample with binary and continuous data.	• Gerstman 9.2-9.5, 11.1-11.3, 16.1-16.3 • Gonick Chapter 8	Homework 5 due
	10.26.12	Extend hypothesis testing, Errors in hypothesis testing, Neyman-		

		Pearson criterion, Power and computation of power with graphical interpretation.		
9	**10.29.12**	Cross-sectional and ecological studies		
	10.31.12	Comparing 2 or more groups with respect to a binary variable (2x2 table, measures of association, and testing for association). Alternative specifications of exposure and outcome variables (review of continuous, ordinal, and categorical measures). 2-sample proportion test. Confidence interval for risk difference.	• Gerstman 17.1-17.4 • Gonick Chapter 9	Homework 6 due
	11.2.12	Comparing 2 groups with respect to a continuous variable (Measure of association, test for association) 2-sample t-tests with equal and unequal variances.	• Gerstman 12.1-12.5	
10	**11.5.12**	Discuss the definitions and interpretations of population parameters, sample estimates, and confidence intervals; one-to-one mapping between CI and HT results and its violation.	• Gerstman 8.1, 10.4	
	11.7.12	Review of measures of occurrence and frequency, epidemic modeling and outbreak investigation, screening, RCTs and cohort studies.		Homework 7 due
	11.9.12	Review of probability and biostatistics concepts, confidence intervals, hypothesis testing in one and two sample settings.		
11	**11.12.12**	Introduction to the Case-Control design; Case ascertainment; Incident versus prevalence cases; Selection/sampling of controls	• Gordis chapter 10	
	11.13.12	EXAM		
	11.14.12	Selection Bias in a Case-Control study (selection of controls); Information Bias in a Case-Control study (recall bias).	• Gordis chapter 15	
	11.16.12	Odds ratios vs. relative risks, and the rare disease assumption; Cases vs. Controls; Introduce analytic tools for statistical comparison and "adjustment".	• Gordis chapter 15 • Gerstman 18.2-18.3. 18.5	
12	**11.19.12**	Formalizing assessments of	• Gordis chapter 15	

		confounding in the design and analysis stage of the study	• Gerstman 19.1-19.2	
	11.21.12	Concepts in effect measure modification; assessing effect measure modification through stratified data analysis	• Gordis chapter 15 • Gerstman 19.4	Homework 8 due
	11.23.12	NO CLASS THANKSGIVING HOLIDAY		
13	**11.26.12**	Introduction to ONE-WAY analysis of variance, ANOVA table F-distribution, F test statistic, Bonferroni	• Gerstman 13.1-13.5	
	11.28.12	Extension to TWO-WAY ANOVA, ANOVA tables without and with interaction, visualization of interaction, understanding steps when testing for significant interactions		Homework 9 due
	11.30.12	Assessing association using a scatter-plot; Correlation.	• Gerstman 14.1-14.3	
14	**12.3.12**	Introduction to Simple Linear Regression (Continuous outcome variable).	• Gerstman 14.4 • Gonick Chapter 11	
	12.5.12	Extending the chi-squared test for 2x2 tables to rxc tables.	• Gerstman 18.3	Homework 10 due
	12.7.12	Person-years, life table and Kaplan Meier methods. Brief introduction to other types of regression analysis (logistic regression for binary outcomes and Cox regression for survival outcomes).	• Gordis chapter 6	
15	**12.10.12**	Study planning: principles of sample size justification (computing power, minimum required sample size, or minimum detectable difference).	• Gerstman 9.6, 12.6, 17.6	
	12.12.12	Exercises for integrating and applying new knowledge in quantitative methods	• Gordis chapter 19	Homework 11 due
	12.14.12	Final exam review lecture		
	12.19.12	FINAL EXAM		