

**MOLECULAR AND GENETIC EPIDEMIOLOGY II**  
**EPI 219 / SPRING 2006 / 1 UNIT**  
**COURSE DIRECTOR: SAUNAK SEN, PHD**

**OBJECTIVES**

This course will discuss selected statistical topics in genetic and molecular epidemiology. The focus will be on specialized statistical procedures and concepts applied more often in genetic studies.

Students will be able to appreciate the statistical principles underlying genetic analysis, and anticipate practical issues in data analysis. Topics covered will range from basic science (Array CGH) to implementation (DNA identification). Examples with real data will be provided.

Students are encouraged to utilize their own projects to motivate discussion and to suggest additional topics of interest.

**PREREQUISITES**

Designing Clinical Research (Epi 180.04), Molecular and Genetic Epidemiology I (Epi 217) and possession of a MD, PhD, DDS or PharmD or equivalent doctoral degree. Exceptions to these prerequisites may be made with the consent of the Course Director, space permitting.

**FACULTY**

Course director: Saunak Sen, PhD  
Phone: 415-514-8035  
Email: [sen@biostat.ucsf.edu](mailto:sen@biostat.ucsf.edu)  
Website: <http://www.biostat.ucsf.edu/sen>

Lecturers:	Jane Fridlyand, PhD	Elad Ziv, MD
	Email: <a href="mailto:janef@cc.ucsf.edu">janef@cc.ucsf.edu</a>	Email: <a href="mailto:elad.ziv@ucsf.edu">elad.ziv@ucsf.edu</a>

**FORMAT**

**Time/Place:** The class meets on Tuesdays from 8:45a-10:15a at China Basin Landing, Room 6702. The first session is on Tuesday, March 28, 2006, and the last session is on Tuesday, May 30, 2006.

**Lectures:** The first 8 sessions will be devoted to lectures. Students should complete the assigned readings before class, and complete exercises after class.

**Student presentations:** The last 2 sessions will be devoted to student presentations. Groups of two students will pick a research paper of their choice to present to the class. Alternatively, they can present a research project they are interested in. The presentation topic should be decided by the seventh week (May 9, 2006).

**GRADING**

Grades will be based on class participation and student presentations.

## A LITTLE EXPERIMENT

Before the first class, can you perform a little coin-spinning experiment? We will use the experiment results of the class to illustrate statistical ideas.

Find a quarter that is not a state quarter. Note the minting year printed on the “heads” side of the coin. Stand the coin on its edge on a table. Hold it vertical on the table pressing lightly with a finger. Then with your other hand, flick the coin sharply so that it starts spinning on its edge. Wait till the coin falls on one of its sides. Note whether it is “heads” (H) or “tails” (T). Repeat this process 10 times. Your data will look like this:

1997      HHTHTTHTTT

There are two pieces of information – the date of minting and the sequence of heads and tails.

One more thing. Practice spinning the coin a few times before you start collecting the data. The whole experiment should not take more than 10 minutes of your time. Thank you!

## LECTURE OUTLINES

Below, please find an outline of what we plan to cover in the course. The *lecture notes* distributed with the class are “works in progress.” I hope you find them useful, and welcome feedback on them (or any other aspect of the course) so that they can be improved for future students.

### 1. Likelihood ratios and permutation tests.

After outlining the course, we will review two simple but important statistical ideas that are used extensively in genetics. The first idea, likelihood ratios, are an example of “parametric” statistics, while the second idea, permutation tests, are an example of “non-parametric” statistics.

- (a) Lecture notes: “Likelihood ratios.”
- (b) Lecture notes: “Permutation tests.”

### 2. Genetic mapping in model organisms

Genetic mapping in model organisms such as mice provide valuable clues for the study of human disease. Genetic crossing experiments are of independent interest, but can also be viewed as simplified versions of linkage studies, and association studies in humans. We will discuss genome scans, searching for epistasis, multiple comparisons, and selective genotyping.

- (a) Lecture notes: “Genetic mapping in model organisms.”
- (b) Korstanje R, Paigen B (2002). From QTL to gene: the harvest begins. *Nature Genetics*. 31:235-6.

### 3. Admixture mapping

We will review the impact of population structure as a confounding factor in association studies, and methods for circumventing the problem. This will be followed by examining the use of recently-admixed human populations for genetic mapping.

- (a) Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics*. 65:220-228.
- (b) Ziv E, Burchard EG (2003). Human population structure and genetic association studies. *Pharmacogenomics*. 4:431-441.

#### 4. Family-based association studies

The TDT tests for the absence of both linkage and association, and is robust to population structure. After reviewing its connections to a randomized trial, we will discuss its extensions to quantitative traits, and other family-based tests of association.

- (a) Lecture notes: "The transmission disequilibrium test."
- (b) Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*. 52:506-16.

#### 5. Population-based association studies

We will discuss population-based association studies including case-control studies. We will discuss methods for probabilistic reconstruction of haplotypes when parental data on subjects is unavailable, and see examples of how to implement such statistical analysis.

- (a) Lecture notes: "Statistical analysis of population-based candidate gene studies."

#### 6. Multiple comparisons

Genome-wide association and linkage studies have to grapple with the problem of multiple comparisons. After reviewing the classical approaches to controlling the family-wise error rate of a set of hypothesis we will discuss recent developments related to the false discovery rate (FDR).

- (a) Lecture notes: "Multiple comparisons and the False Discovery Rate."
- (b) Curran-Everett D (2000). Multiple comparisons: philosophies and illustrations. *Am J Physiol Regul Integr Comp Physiol*. 279:R1-8.

#### 7. DNA identification and genetic testing

With the growth of genomic information, DNA identification and genetic testing are becoming increasingly common. We will begin by reviewing basic ideas of Bayesian statistical inference in the context of diagnostic tests. These ideas will be used to examine forensic applications of DNA identification and genetic testing.

- (a) Jobling MA, Gill P (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*. 5:739-752.
- (b) Taylor MRG (2001). Genetic testing for inherited breast and ovarian cancer syndromes: important concepts for the primary care physician *Postgraduate Medical Journal*. 77:11-15.

#### 8. Array CGH

Microarray-based Comparative Genomic Hybridization (Array CGH) is a technique that measures DNA copy number changes, and localizes them on the genome. Such copy number aberrations are common in cancer and in many developmental abnormalities. After outlining the technology, we will discuss statistical methods currently used for their analysis, and future directions.

- (a) Pinkel D, Albertson DG (2005). Array comparative genomic hybridization and its applications in cancer. *Nature Genetics*. 37Suppl:S11-17
- (b) Willenbrock H, Fridlyand J (2005). A comparison study: applying segmentation to array CGH data for downstream analyses. *Bioinformatics*. 21:4084-91.

#### 9. Student presentations 1

#### 10. Student presentations 2