

**Graduate School of Public Health
Department of Human Genetics
HuGen 2080
Statistical Genetics**

Monday 1:00-2:30 PM
Thursday 12:30-1:55 PM
A425 Crabtree Hall
3 Credit Hours
Spring 2020

Contact Information

Daniel E. Weeks, Ph.D.
Professor of Human Genetics and Biostatistics
Department of Human Genetics
University of Pittsburgh
Public Health 3119
130 DeSoto Street
Pittsburgh, PA 15261
USA

Work: 1 412 624 5388
Email: weeks@pitt.edu
Web site: <https://watson.hgen.pitt.edu>
Twitter: @StatGenDan

Office hours: by appointment (or drop by).

Faculty Availability

I welcome your questions at any time. Please feel free to drop by my office, or set up an appointment. E-mail is also an excellent way to reach me. However, since I get so many e-mails, please use an informative subject line, starting with "HuGen 2080: ". And feel free to send a reminder e-mail.

Course Description

AN ADVANCED COURSE WHICH DISCUSSES THE PRINCIPLES AND PRACTICE OF STATISTICAL GENETICS IN THE AREA OF GENETIC EPIDEMIOLOGY OF HUMAN DISEASES AND TRAITS. THE COURSE WILL COVER STATISTICAL MODELING AND METHODOLOGY IN FAMILIAL AGGREGATION, LINKAGE

ANALYSIS AND ASSOCIATION ANALYSIS; THE COURSE INCLUDES HANDS-ON EXPERIENCE WITH CURRENT COMPUTER PROGRAMS USED IN THESE RESEARCH AREAS.

The objective of this class is to introduce students to advanced topics in genetic epidemiology, specifically related to human diseases and traits, with an emphasis on statistics and the mathematics behind the statistics. As such, the course requires familiarity with basic probability theory and statistical principles, as well as population genetics. After finishing this course, the student should understand the assumptions, strengths, and weaknesses of the various statistical tests, as well as be well prepared to undertake a statistical genetics analysis, including study design, phenotype definition and modeling, estimation of heritability, linkage analysis, and association analysis.

Learning Objectives

Upon completion of this course the student will be able to:

- Explain the assumptions, strengths, and weaknesses of the various statistical tests used in genetic epidemiology.
- Devise testable hypotheses in human genetics and identify appropriate study designs to test these hypotheses.
- Describe preferred methodological alternatives to commonly used statistical methods when assumptions are not met.
- Apply the statistical models presented in analyzing and interpreting real data sets.
- Apply genetic analysis programs and interpret the results.
- Read and evaluate current literature in statistical genetics.

Required Textbooks

Andreas Ziegler, Inke R. König (2010) A Statistical Approach to Genetic Epidemiology: Concepts and Applications, Second Edition

Full text available online: <https://ebookcentral.proquest.com/lib/pitt-ebooks/detail.action?docID=708055>

E-learning platform:

https://application.wiley-vch.de/books/geneticepidemiology/IMS_LO/onCampusBook/65270/index.html

Supplemental Readings/Bibliography

Henry Stewart Talks: <https://info.hsls.pitt.edu/updatereport/?p=4167>

There are multiple methods for locating the Henry Stewart Talks:

1. All talks are catalogued in [PITTCat](#); search for **Henry Stewart Talks** under “Title Begins with” for an alphabetical list or to search by individual title.
2. Click on the “Videos” tab on the [MolBio home page](#).
3. Visit [Henry Stewart Talks: Online Seminars by Leading World Experts](#).

Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. New York, Springer Science.

DOI: 10.1007/978-1-4419-7338-2

Web access: <http://www.springerlink.com/content/978-1-4419-7337-5>

Statistical Human Genetics

Methods and Protocols

Editor: Robert C. Elston

New York: Humana Press; Springer, 2017.

Web access: <https://link.springer.com/book/10.1007/978-1-4939-7274-6>

Handbook on Analyzing Human Genetic Data

Computational Approaches and Software

2010

Web access: <https://link.springer.com/book/10.1007%2F978-3-540-69264-5>

Current Protocols in Human Genetics. New York, John Wiley

DOI: 10.1002/0471142905

Web access: <http://onlinelibrary.wiley.com/book/10.1002/0471142905>

Genetic Mapping Section: [https://currentprotocols.onlinelibrary.wiley.com/doi/toc/10.1002/\(ISSN\)1934-8258.GeneticMapping](https://currentprotocols.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1934-8258.GeneticMapping)

Stram, Daniel O (2014) Design, analysis, and interpretation of genome-wide association scans

Web access: <https://link.springer.com/book/10.1007%2F978-1-4614-9443-0>

Electronic Supplementary Material:

http://www.springer.com/cda/content/document/cda_downloaddocument/Electronic+Supplementary+Material%2C+Ch.2-8.zip?SGWID=0-0-45-1430364-p176269527

CourseWeb/BlackBoard and GitHub Classroom Instruction

This course will extensively use the University's CourseWeb site <http://courseweb.pitt.edu/> [also known as BlackBoard]. To login, you must have a Pitt account. Your login ID is the same as your login ID for your Pitt account and your password is the same as for your Pitt account. This will be augmented with GitHub Classroom.

Each lecture will be accompanied by supporting material and further reading, all of which will be made available around the time of the lecture. It is the student's responsibility to check for, and read, this material.

Discussion topics related to the course may also be posted on CourseWeb, and, for the purpose of determining a student's grade, participation in these discussions will be considered as equivalent to participation in class discussion.

The instructors will use the CourseWeb site and GitHub Classroom as the primary means of communicating with the students, who are expected to check these sites on a regular basis throughout the semester.

Class Expectations/ Behavior and Ground Rules

Attendance and active participation are expected.

Please be on time and turn off your cell phone. While laptops may be used to complete computer exercises or to take notes, please do not use them during class time for non-class purposes. It is expected that you will have read all the assigned readings prior to class. To facilitate our use of the readings, please bring an interesting discussion question from the readings to class with you. If you will miss a class, please let us know in advance.

Statement on Classroom Recording

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Grading Scale

| Grades scored between | will equal |
|-----------------------|------------|
| 97% and 100% | A+ |
| 94% and less than 97% | A |
| 90% and less than 94% | A- |
| 87% and less than 90% | B+ |
| 84% and less than 87% | B |
| 80% and less than 84% | B- |
| 77% and less than 80% | C+ |
| 74% and less than 77% | C |
| 70% and less than 74% | C- |
| 67% and less than 70% | D+ |
| 64% and less than 67% | D |
| 60% and less than 64% | D- |
| 0% and less than 60% | F |

Student Performance Evaluation (Assessments and Weights)

Evaluation will be based on the following components:

Attendance and Participation

Attendance and active participation are expected.

Homework (50% of final grade)

The homework assignments will comprise problems that extend the in-class activities and complement the lectures.

Students will read the weekly assigned scientific paper and summarize their understanding of the main points of the assigned paper.

A least one of the homework assignments will consist of reviewing a preprint. After the reviews are completed, students will be divided into two groups, and each group will

consolidate their set of reviews into one combined polished review, which will then be submitted to AcademicKarma as a formal review.

Paper Presentation (10% of final grade)

All students will give a 30 minute presentation on an assigned paper in statistical genetics.

Mid-term exam (20% of final grade)

A written in-class mid-term exam will be given.

Final exam (20% of final grade)

A written in-class final exam will be given.

Schedule of Sessions and Assignments

The detailed schedule is provided at the end of this document.

Accommodation for Students with Disabilities

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services, 140 William Pitt Union, 412-648-7890 as early as possible in the term.

Academic Integrity Statement

All students are expected to adhere to the school's standards of academic honesty. Cheating/plagiarism will not be tolerated. The Graduate School of Public Health's policy on academic integrity, which is based on the University policy, is available online in the Pitt Public Health Academic Handbook www.publichealth.pitt.edu/home/academics/academic-requirements. The policy includes obligations for faculty and students, procedures for adjudicating violations, and other critical information. Please take the time to read this policy.

Plagiarism

Integrity of the academic process requires that credit be given where credit is due. Accordingly, it is unethical to present as one's own work the ideas, representations, words of another, or to permit another to present one's own work without customary and proper acknowledgement of sources.

A student has an obligation to exhibit honesty and to respect the ethical standards of the profession in carrying out his or her academic assignments. Without limiting the application of this principle, a student may be found to have violated this obligation if he or she:

1. Presents as one's own, for academic evaluation, the ideas, representations, or words of another person or persons without customary and proper acknowledgment of sources.
2. Submits the work of another person in a manner which represents the work to be one's own.

See: <http://www.bc.pitt.edu/policies/policy/02/02-03-02.html>

To avoid plagiarism, you must give “customary and proper acknowledgment of sources” by appropriately and clearly identifying which thoughts are yours and which are others, and appropriately citing your sources.

Sophisticated plagiarism detection software will be used in this course. If plagiarism is detected, you will automatically receive a grade of zero for that assignment and the incident will be reported, as required, to your Dean.

Sexual Misconduct, Required Reporting and Title IX Statement

The University is committed to combatting sexual misconduct. As a result, you should know that University faculty and staff members are required to report any instances of sexual misconduct, including harassment and sexual violence, to the University's Title IX office so that the victim may be provided appropriate resources and support options. What this means is that as your professor, I am required to report any incidents of sexual misconduct that are directly reported to me, or of which I am somehow made aware.

There are two important exceptions to this requirement about which you should be aware:

A list of the designated University employees who, as counselors and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here: www.titleix.pitt.edu/report/confidentiality

An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as part of an academic project, classroom discussion, or course assignment, are not required to be disclosed to the University's Title IX office.

If you are the victim of sexual misconduct, Pitt encourages you to reach out to these resources:

- *Title IX Office: 412-648-7860*
- *SHARE @ the University Counseling Center: 412-648-7930 (8:30 A.M. TO 5 P.M. M-F) and 412-648-7856 (AFTER BUSINESS HOURS)*

If you have a safety concern, please contact the University of Pittsburgh Police, 412-624-2121.

Other reporting information is available here: www.titleix.pitt.edu/report

Diversity Statement

The University of Pittsburgh Graduate School of Public Health considers the diversity of its students, faculty, and staff to be a strength and critical to its educational mission. Pitt Public Health is committed to creating and fostering inclusive learning environments that value human dignity and equity. Every member of our community is expected to be respectful of the individual perspectives, experiences, behaviors, worldviews, and backgrounds of others. While intellectual disagreement may be constructive, no derogatory statements, or demeaning or discriminatory behavior will be permitted. If you feel uncomfortable or would like to discuss a situation, please contact any of the following:

- *the course instructor;*
- *the Pitt Public Health Associate Dean for Diversity at 412-624-3506 or nam137@pitt.edu;*
- *the University's Office of Diversity and Inclusion at 412-648-7860 or <https://www.diversity.pitt.edu/make-report/report-form> (anonymous reporting form).*

Copyright Notice

Course material may be protected by copyright. United States copyright law, 14 USC section 101, et sec., in addition to University policy and procedures, prohibit

unauthorized duplication or retransmission of course materials. See [Library of Congress Copyright Office](#) and the [University Copyright Policy](#).

Schedule of Sessions and Assignments

Note: This schedule is, of course, may be updated as we proceed through the course.

Background reading:

Ziegler and König - Chapter 1: Molecular Genetics

Ziegler and König - Chapter 2: Formal Genetics

1/6/2020: **Models, Maps, and Markers**

Active Learning: Intro to Unix & PLINK

Required Reading:

Ziegler and König - Chapter 3: Genetic Markers

Ziegler and König - Chapter 5: Genetic Map Distances

Henry Stewart Talk:

Introductory genetics for statisticians/ Robert C. Elston.

Supplementary reading:

Elston RC (2000) Introduction and overview. Statistical methods in genetic epidemiology. Stat Methods Med Res 9:527-541

Chapters 1, 2, and 3: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Assignment:

Homework 1

Learning objectives:

To review basic genetic models

To learn about genetic markers

1/9/2020: **Study Design Overview**

Active Learning: Student Presentation

Learning objectives:

- To learn the basic principles of study design for genetic studies
- To understand the vital importance of phenotype definition
- To understand the best sample selection strategies

1/13/2020: **Familial Aggregation: Recurrence Risk Ratios, Heritability**

Active Learning: SOLAR heritability computer lab

Required Reading:

Ziegler and König - Chapter 6: Familiality, Heritability, and Segregation Analysis

Henry Stewart Talk:

Heritability and its uses/ Doug Speed.

Supplementary reading:

Chapter 4: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Assignment:

Homework 2

Learning objectives:

- To learn aggregation analysis
- To learn how to estimate recurrence risk ratios
- To review the concept of heritability

1/16/2020: **Familial Aggregation: Segregation Analysis, Ascertainment**

Active Learning: Student Presentation

Required Reading:

Ziegler and König - Chapter 6: Familiality, Heritability, and Segregation Analysis

Learning objectives:

- To learn about segregation analysis
- To understand how to take ascertainment into account in the segregation models
- To formulate testable hypotheses about genetic models

1/20/2020: **No class - Martin Luther King Day**

1/23/2020: **LOD scores: Model-based Linkage Analysis**

Active Learning: Scientific Writing: Review/Response

Required Reading:

Ziegler and König - Chapter 7: Model-based Linkage Analysis

Henry Stewart Talk:

Linkage and sequence analysis in families/ Christopher Amos.

Supplementary reading:

Chapters 5 and 6: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Assignment:

Homework 3

Learning objectives:

- To learn how to compute LOD scores
- To learn about different map functions, and the distinction between genetic and physical maps
- To formulate testable hypotheses about linkage

1/27/2020: **Non-parametric methods**

Active Learning: Merlin computer lab

Required Reading:

Ziegler and König - Chapter 8: Model-free Linkage Analysis for Dichotomous Traits

Supplementary reading:

Shih MC, Whittemore AS (2001) Allele-sharing among affected relatives: non-parametric methods for identifying genes. Stat Methods Med Res 10:27-55

Learning objectives:

- To learn how to carry out non-parametric linkage analyses
- To understand the motivation behind non-parametric linkage analysis approaches

1/30/2020: **Association: Case/Control & Quantitative Traits**

Active Learning: Student Presentation

Required Reading:

Ziegler and König - Chapter 10: Fundamental Concepts of Association Analysis

Ziegler and König - Chapter 11: Association Analysis with Unrelated Individuals

Henry Stewart Talk:

Introduction to genetic association analysis/ Jenny Barrett.

Supplementary reading:

Chapter 7: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2:91-99.

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781-791

Assignment:

Homework 4

Learning objectives:

To formulate testable hypotheses about association
To understand and apply various case/control association tests
To understand allele-based and genotype-based association tests, and trend tests.

2/3/2020: **Association: Family-based and Haplotype-based**

Active Learning: PLINK computer lab

Required Reading:

Ziegler and König - Chapter 12: Association Analysis in Families

Supplementary reading:

Chapter 9 & Chapter 10, Section 2: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Ott J, Kamatani Y, Lathrop M (2011) Family-based designs for genome-wide association studies. Nat Rev Genet 12:465-474

Learning objectives:

To learn how to analyze family data for association
To learn how to test haplotypes for association
To understand sparsity issues involved in haplotyped-based tests

2/6/2020: **Multiple testing**
Active Learning: Student Presentation

Required Reading:
Ziegler and König - Chapter 14, Section 14.4: Multiple Testing

Supplementary reading:
Chapter 10, Section 1: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Assignment:
Homework 5

Learning objectives:
To understand how to adjust for multiple testing

2/10/2020: **Power to detect Association: Linkage vs. Association**

Supplementary reading:
Clerget-Darpoux F, Elston RC (2007) Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. Hum Hered 64:91-96

Learning objectives:
To learn how to compute power for detecting association
To compare and contrast linkage and association
To understand the relative strengths and weaknesses of linkage and association tests

2/13/2020: **Rare variants**

Active Learning: Student Presentation

Supplementary reading:

Asimit J, Zeggini E (2010) Rare variant association analysis methods for complex traits. Annu Rev Genet 44:293-308

Bansal V, Libiger O, Torkamani A, Schork NJ (2010) Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 11:773-785

Assignment:

Homework 6

Learning objectives:

To learn how to test rare variants for association

To learn about burden tests, collapsing or grouping tests, weighted sum tests, and variable threshold tests.

2/17/2020: **Methods for correlated data: LME, GEE, Score**

Learning objectives:

To learn about linear mixed effects models, generalized estimating equations, and score tests

To learn how to properly model relatedness while testing genetic hypotheses

2/20/2020: **Bayesian Methods in Human Genetics**

Active Learning: Student Presentation

Assignment:

Homework 7

Learning objectives:

To learn about Bayesian methods in human genetics

To understand Bayesian principles

2/24/2020: **Gene x Gene interaction**

Supplementary reading:

Cordell HJ (2009) Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 10:392-404

Gilbert-Diamond D, Moore JH (2011) Analysis of gene-gene interactions. Curr Protoc Hum Genet Chapter 1:Unit1 14

Learning objectives:

To learn how to test for gene x gene interaction

To formulate testable hypotheses about gene x gene interaction

2/27/2020: **Gene x Environment interaction**

Active Learning: Student Presentation

Henry Stewart Talk:

Statistical issues in epidemiologic studies of gene-environment interaction/ Peter Kraft, Donna Spiegelman.

GxE interactions in genome-wide association studies/ David V. Conti.

Supplementary reading:

Chapter 10, Section 3: Laird NM, Lange C (2011) The Fundamentals of Modern Statistical Genetics. Springer.

Ottman R (1990) An epidemiologic approach to gene-environment interaction. Genet Epidemiol 7:177-185

Thomas D (2010) Gene-environment-wide association studies: emerging approaches. Nat Rev Genet 11:259-272

Learning objectives:

To learn how to test for gene x environment interaction

To formulate testable hypotheses about gene x environment interaction

3/2/2020: **Review for Mid-term exam**

3/5/2020: **Mid-term exam**

Assignment:

Mid-term Exam

3/9/2020: **Spring Recess**

3/12/2020: **Spring Recess**

3/16/2020: **Meta analysis**

Required Reading:

Ziegler and König - Chapter 14, Section 14.5: Accumulating Data from Genome-wide Association Studies

Henry Stewart Talk:

Winner's curse, replication and meta-analysis/ Frank Dudbridge.

Assignment:

Homework 8

Learning objectives:

To learn about the different types of meta analysis.

To understand the assumptions made by meta analysis.

3/19/2020: **Fine mapping**

Active Learning: Student Presentation

Learning objectives:

To learn how to carry out fine mapping

To understand and apply conditional tests of association

3/23/2020: **Methods for multivariate phenotypes**

Assignment:

Homework 9

Learning objectives:

To learn about methods for analyzing multivariate phenotypes

To learn how to properly account for correlation among phenotypes

3/26/2020: **Special Topic Lecture by Ron Yurko**

Required Reading:

Yurko R, G'Sell M, Roeder K, Devlin B. Application of post-selection inference to multi-omics data yields insights into the etiologies of human diseases. bioRxiv; 2019. DOI: 10.1101/806471

3/30/2020: **Heritability from GWAS**

Assignment:

Homework 10

Learning objectives:

To learn how to estimate heritability using unrelated samples.

To understand polygenicity.

4/2/2020: **Special Topic Lecture by Qi Yan**

Required Reading:

Yan Q, Weeks DE, Xin H, et al. Deep-learning-based Prediction of Late Age-Related Macular Degeneration Progression. medRxiv; 2019. DOI: 10.1101/19006171.

4/6/2020: **LDscore regression**

Active Learning: Student Presentation: Review/Response

Assignment:

Homework 11

Learning objectives:

To understand the principles of LDscore regression

To understand polygenicity.

4/9/2020: **Special Topic Lecture by Wei Chen**

Required Reading:

Sun Z, Chen L, Xin H, Jiang Y, Huang Q, Cillo AR, Tabib T, Kolls JK, Bruno TC, Lafyatis R, Vignali DAA, Chen K, Ding Y, Hu M, Chen W. A Bayesian mixture model for clustering droplet-based single-cell transcriptomic data from population studies. Nat Commun. 2019 Apr 9;10(1):1649. doi: 10.1038/s41467-019-09639-3. PubMed PMID: 30967541; PubMed Central PMCID: PMC6456731.

4/13/2020: **Mendelian Randomization**

Active Learning: Student Presentation

Henry Stewart Talk:

Causal inference in genetic epidemiology: Mendelian randomization and beyond / Krista Fischer.

Assignment:

Homework 12

Learning objectives:

To understand the basic principles of Mendelian Randomization.
To formulate testable hypotheses about causation using Mendelian Randomization approaches.

4/16/2020: **Special Topic Lecture by Jiebiao Wang**

Required Reading:

Wang J, Devlin B, Roeder K. Using multiple measurements of tissue to estimate subject- and cell-type-specific gene expression. Bioinformatics. 2019 Aug 10. pii: btz619. doi: 10.1093/bioinformatics/btz619. [Epub ahead of print] PubMed PMID: 31400192.

4/20/2020: **Genetic Risk Scores & Polygenic Risk Scores & Genomic Prediction**

Learning objectives:

To understand how to construct and use genetic risk scores.
To understand the limits of genomic prediction.

4/23/2020: **Final exam**

Assignment:

Final Exam

4/25/2020: **Spring Term ends.**
