Biostatistics & Genetic Epidemiology

The mission of the Institute for Informatics, Data Science and Biostatistics (I^{2}DB) focuses on the informatics, data science, and biostatistics landscape at Washington University School of Medicine in order to transform research, education, and patient care by emphasizing precision medicine and efforts to improve the quality of health care and public health initiatives locally, nationally, and worldwide.

The Biostatistics education programs offered by I^{2}DB include two master's degrees and two graduate certificates: the Master of Science in Biostatistics (MSIBS), the Master of Science in Biostatistics and Data Science (MSBDS), the Certificate in Genetic Epidemiology, and the Certificate in Biostatistics and Data Science. Interested students may pursue individual courses offered by the division.

Washington University School of Medicine is known for being at the forefront of medical research and primary care; the school engages students in research and practical training so that they can contribute to improving health outcomes. Our programs train students as critical thinkers and collaborators in biostatistics, genetics, and data science. We seek those with undergraduate degrees in the quantitative and biomedical sciences, including fields such as Mathematics, Statistics, Computer Science, Informatics, and Biomedical Engineering.

Our programs are designed to teach students how to manage, analyze, and interpret health data using statistical and data science approaches. Internationally renowned faculty from multiple disciplines, including Biostatistics, Genetics, Informatics, Medicine, and Public Health, will train a new generation of quantitative scientists. The curriculum offers a unique training experience that combines core data science learning in statistical and computational methodologies with practical training in real-world data analysis of cutting-edge Biomedical and Genomics research.

Academic Calendar

The academic programs begin in early July each year. They start with preparatory workshops, which are followed by intensive summer semester courses. The program follows the Washington University Arts & Sciences academic calendar for fall and spring courses.

Location

The Biostatistics & Genetic Epidemiology program is located in the Institute for Informatics, Data Science, and Biostatistics, which is on the fifth floor of the Bernard Becker Medical Library (660 S. Euclid Ave., St. Louis, MO 63110), rooms 500 through 508.

Additional Information

Shelby Cripe, MA
Program Manager
Email: s.swanner@wustl.edu

Treva Rice, PhD
Interim Program Director
Professor of Biostatistics
Email: treva@wustl.edu

Lei Liu, PhD
Associate Program Director
Professor of Biostatistics
Email: lei.liu@wustl.edu

Washington University School of Medicine
Biostatistics Education Program
Institute for Informatics, Data Science and Biostatistics (I^{2}DB)
660 S. Euclid Ave., MSC 8067-0013-05
St. Louis, MO 63110-1093

Phone:	314-362-1384
Email:	OHIDS-Education@wustl.edu
Website:	https://i2db.wustl.edu/

Degrees & Offerings

- Master of Science in Biostatistics
- Master of Science in Biostatistics and Data Science
- Certificate in Biostatistics and Data Science
- Certificate in Genetic Epidemiology

Research

Master's students have multiple opportunities to engage in biomedical research. After completing the first summer semester, students in the MSIBS and MSBDS program are eligible to work as part-time research assistants. These positions are frequently available, both within the I^{2}DB as well as in other departments and research labs on the School of Medicine campus. In addition, depending on the degree program, students will intern and/or work on an independent mentored research project to hone their research skills, including study design, data analysis, and interpretation. GEMS students will work on a mentored research project to explore and characterize the interplay between genes and the environment that affects the biological processes underlying disease.

Mentored Research

All students enrolled in the Mentored Research course will complete a master's thesis, which may involve conducting and reporting on comprehensive data analysis or conducting research and reporting on a focused methodological problem; the latter may include a computer simulation approach to solve a problem, an in-depth review of available methods in a certain topical area, or the development of new methods. Each student will work closely with a nentor who has expertise in biostatistics or a related quantitative field. The grade for each student will be determined in consultation with the mentor.

Internship

The primary goal of the Internship program is for all students enrolled in Internship to acquire critical professional experience so that they will be well prepared to enter the job market upon graduation. This provides an opportunity for students to test-drive the job market, develop contacts, build marketable skills, and figure out likes and dislikes in the chosen field.

Faculty

Philip R.O. Payne, PhD, FACMI, FAMIA, FAIMBE, FIAHSI
Founding Director, Institute for Informatics, Data Science and Biostatistics (I^{2}DB)
Janet and Bernard Becker Professor
Associate Dean for Health Information and Data Science, School of Medicine
Chief Data Scientist, School of Medicine

Charles Goss, PhD
Associate Director, Center for Biostatistics and Data Science
Assistant Professor of Biostatistics

Lei Liu, PhD
Associate Program Director
Professor of Biostatistics

Treva Rice, PhD
Interim Program Director
Professor of Biostatistics

Zachary Abrams, PhD
Instructor in Biostatistics

Ling Chen, PhD
Assistant Professor of Biostatistics

William Dunagan, MD, MS
Professor of Medicine in the Division of Infectious Diseases

Rosie Dutt, PhD
Instructional Consultant

Robert Fitzgerald, PhD
Associate Professor of Psychiatry

Charles Gu, PhD
Associate Professor of Biostatistics

Aditi Gupta, PhD
Assistant Professor in Biostatistics

Daphne Lew, PhD, MPH
Instructor in Biostatistics

Ruijin Lu, PhD
Assistant Professor of Biostatistics

J.Phillip Miller, PhD
Professor of Biostatistics

Tara Payne, MA, FAMIA
Lecturer

DC Rao, PhD
Professor of Biostatistics

Ken Schechtman, PhD
Professor of Biostatistics

Yun Ju Sung, PhD
Associate Professor of Biostatistics

RJ Waken, PhD
Instructor of Biostatistics

Peter Wang, PhD
Assistant Professor of Biostatistics

Chengjie Xiong, PhD
Professor of Biostatistics

Linying Zhang, PhD
Assistant Professor of Biostatistics

Visit our website for more information about our faculty and their appointments.

Courses

Visit online course listings to view offerings for M21 MSB.

M21 MSB 503 Statistical Computing with SAS
Intensive hands-on summer training in SAS (Statistical Analysis System) during seven full weekdays. Students will learn how to use SAS for handling, managing, and analyzing data. Instruction is provided in the use of SAS programming language, procedures, macros, and SAS SQL. The course will include exercises using existing programs written by SAS experts.
Credit 2 units.

M21 MSB 506 Introduction to R For Data Science
This is an introduction to the R Statistical Environment for new users. R is "a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, etc." The goal is to give students a set of tools to perform statistical analysis in medicine, biology, or epidemiology. At the conclusion of this primer, students will: be able to manipulate and analyze data, write basic models, understand the R environment for using packages, and create standard or customized graphics. This primer assumes some knowledge of basic statistics as taught in a first semester undergraduate or graduate sequence. Topics should include: probability, cross-tabulation, basic statistical summaries, and linear regression in either scalar or matrix form.
Credit 2 units.

M21 MSB 507 R and Python For Biomedical Sciences
Students will explore data manipulation techniques using Pandas in Python, focusing on leveraging its native functionalities to efficiently handle and analyze biomedical data. Emphasis will be placed on transitioning from traditional looping methods to utilizing Pandas'

powerful features for data manipulation tasks. In Python, students will delve into essential libraries including Pandas and NumPy using Jupyter Notebooks, gaining proficiency in data structuring, exploration, and analysis. Through hands-on exercises and projects, students will learn to leverage Python's ecosystem for efficient data handling and visualization, preparing them for real-world applications inbiomedical research and analysis. The course also introduces R programming within the tidyverse framework which includes a number of R packages designed to make data manipulation, visualization, and analysis easier and more intuitive. We will focus on the use of packages including dplyr and ggplot2 for data manipulation and visualization, respectively. Students will learn to harness R's capabilities for basic statistical analysis and visualization, complementing their Python skills and expanding their toolkit for biomedical informatics and biostatistics.
Credit 2 units.

M21 MSB 512 Ethics in Biostatistics and Data Science
This course prepares biostatisticians to analyze and address ethical and professional issues in the practice of biostatistics across the range of professional roles and responsibilities of a biostatistician. The primary goals are for biostatisticians to recognize complex situational dynamics and ethical issues in their work and to develop professional and ethical problem-solving skills. The course specifically examines ethical challenges related to research design, data collection, data management, ownership, security, and sharing, data analysis and interpretation, and data reporting and provides practical guidance on these issues. The course also examines fundamentals of the broader research environment in which biostatisticians work, including principles of ethics in human subjects and animal research, regulatory and compliance issues in biomedical research, publication and authorship, and collaboration in science. By the conclusion of the course, participants will understand the ethical and regulatory context of biomedical research; identify ethical issues, including situational dynamics that serve to foster or hinder research integrity, in the design and conduct of research and the management, analysis, and reporting of data; and utilize strategies that facilitate ethical problem-solving and professionalism.
Credit 2 units.

M21 MSB 515 Fundamentals of Genetic Epidemiology
Lectures cover causes of phenotypic variation, familial resemblance and heritability, Hardy-Weinberg Equilibrium, ascertainment, study designs and basic concepts in genetic segregation, linkage and association.The computer laboratory portion is designed as hands-on practice of fundamental concepts. Students will gain practical experience with various genetics computer programs (e.g. SOLAR, MERLIN, QTDT, and PLINK). Auditors will not have access to the computer lab sessions. Prerequisite: R for Data Science (M21-506).
Credit 3 units.

M21 MSB 5483 Human Genetic Analysis
Basic Genetic concepts: meiosis, inheritance, Hardy-Weinberg Equilibrium, Linkage, segregation analysis; Linkage analysis: definition, crossing over, map functions, phase, LOD scores, penetrance, phenocopies, liability classes, multi-point analysis, non-parametric analysis (sibpairs and pedigrees), quantitative trait analysis, determination of power for mendelian and complex trait analysis; Linkage Disequilibrium analyses: allelic association (case control designs and family bases studies), QQ and Manhattan plots, whole genome association analysis; population stratification; Quantitative Trait Analysis: measured genotypes and variance components. Hands-on computer lab experience doing parametric linkage analysis with the program LINKAGE, model free linkage analyses with Genehunter and Merlin, power computations with SLINK, quantitative trait anaylses with SOLAR, LD computations with Haploview and WGAViewer, and family-based and case-control association anaylses with PLINK and SAS. The methods and exercises are coordinated with the lectures and students are expected to understand underlying assumptions and limitations and the basic calculations performed by these computer programs. Auditors will not have access to the computer lab sessions. Prerequisite: M21-515 Fundamentals of Genetic Epidemiology. For details, to register and to receive the required permission of the Coursemaster contact the MSIBS Program Manager (biostat-msibs@email.wustl.edu or telephone 362-1384).
Same as L41 Biol 5483
Credit 3 units.

M21 MSB 550 Introduction to Bioinformatics
Provide a broad exposure to the basic concepts, methodology and application of bioinformatics to solve biological problems. Specifically, the students will learn the basics of online genomic/protein databases and database mining tools, and acquire understanding of mathematical algorithms in genome sequence analysis (alignment analysis, gene finding/predicting), gene expression microarray (genechip) analysis, and of the impact of recent developments in the protein microarray technology. Prerequisite: R for Data Science (M21-506).
Credit 3 units.

M21 MSB 560 Biostatistics I
This course is designed for students who want to develop a working knowledge of basic methods in biostatistics. The course is focused on biostatistical and epidemiological concepts and on practical hints and hands-on approaches to data analysis rather than on details of the theoretical methods. We will cover basic concepts in hypothesis testing, will introduce students to several of the most widely used probability distributions, and will discuss classical statistical methods that include t-tests, chi-square tests, regression analysis, and analysis of variance. Both in-class examples and homework assignments will involve extensive use of SAS. Prerequisite: M21-503, Statistical Computing with SAS®, or student must have good practical experience with SAS®. Students are required to participate in the "Computing/Unix" workshops offered free of charge prior to this course.
Credit 3 units.

M21 MSB 570 Biostatistics II
This course is designed for students who have taken Biostatistics I or the equivalent and who want to extend their knowledge of biostatistical applications to more modern and more advanced methods. Biostatistical methods to be discussed include logistic and Poisson regression, survival analysis, Cox regression analysis, and several methods for analyzing longitudinal data. Students will be introduced to modern topics that include statistical genetics and bioinformatics. The course will also discuss clinical trial design, the practicalities of sample size and power computation and meta analysis, and will ask students to read journal articles with a view towards encouraging a critical reading of the medical literature. Both in-class examples and homework assignments will involve extensive use of SAS. Prerequisite: M21-560, Biostatistics I or its equivalent as judged by the course masters.
Credit 3 units.

M21 MSB 600 Mentored Research
Student undertakes supervised research in a mentor's lab. The goal is to acquire important research skills as well as good writing and presentation skills. The student finds a mentor and they together identify a research topic. A written thesis based on the research, prepared in the format of an actual scientific publication, must be submitted and presented to a select audience. The course instructor will organize a few meetings throughout to facilitate the whole process. The course instructor will determine the grade (pass/fail) in consultation with the mentors. Permission of the Course Instructor is required.

Credit variable, maximum 6 units.

M21 MSB 617 Study Design and Clinical Trials

The course will focus on statistical and epidemiological concepts of study design and clinical trials. Topics include: different phases of clinical trials, various types of medical studies (observational studies, retrospective studies, adaptive designs, and comparative effectiveness research), and power analysis. Study management and ethical issues are also addressed. Students will be expected to do homework and practice power analysis during lab sessions. Prerequisites: M21-560 Biostatistics I and M21-570 Biostatistics II. Permission of the Course Instructor is required.

Credit 3 units.

M21 MSB 618 Survival Analysis

This course will cover the basic applied and theoretical aspects of models to analyze time-to-event data. Basic concepts will be introduced including the hazard function, survival function, right censoring, and the Cox-proportional hazards (PH) model with fixed and time dependent covariates. Additional topics will include regression diagnostics for survival models, the stratified PH model, additive hazards regression models and multivariate survival models. Permission of the Course Instructor required. Prerequisites: M21-560 Biostatistics I and M21-570 Biostatistics II.

Credit 3 units.

M21 MSB 621 Computational Statistical Genetics

This course is designed to give the students computational experience with the latest statistical genetics methods and concepts, so that they will be able to computationally implement the method(s)/model(s) developed as part of their thesis. Concentrating on the applications of genomics and computing, it deals with creating efficient new bioinfomatic tools to interface with some of the latest, most important genetic epidemiological analysis software, as well as how to derive, design and implement new statistical genetics models. The course also includes didactic instruction on haplotype estimation and modeling of relationship to phenotype, LD mapping, DNA pooling analysis methods, analysis approaches in pharmacogenomics (with an emphasis on possible genomic role in drug response heterogeneity), and epistasis (GxG) and GxE interactions; data mining methods, including clustering, recursive partitioning, boosting, and random forests; and fundamentals of meta-analysis, importance sampling, permutation tests and empirical p-values, as well as the design of monte-carlo simulation experiments. Prerequisite: Biostatistics I and II, permission of the instructor.

Credit 3 units.

M21 MSB 630 Internship

The primary goal of the Internship program is for students to acquire critical professional experience so that they will be well prepared to enter the job market upon graduation. This provides an opportunity for students to develop contacts, build marketable skills and perceive likes and dislikes in the chosen field. Students will have an opportunity to work with experienced mentors (PIs) on a range of projects that may include data management, data analysis, study design, and protocol development among other things. Students may have opportunities to contribute to and participate in the preparation of publishable quality manuscripts. As part of the Internship requirements, each student will submit a one-page Abstract of the work performed as part of the internship and will give a presentation of the Internship experience. The grade (pass/fail) for each student will be determined in consultation with the mentor.

Credit variable, maximum 6 units.

M21 MSB 660 Biomedical Data Mining

This course introduces methods and applications of biomedical data mining. Various computational and statistical methods will be presented, such as model selection and regularization, resampling methods, tree-based methods, and artificial intelligence. In addition to the common applications of the covered methods in biomedical sciences, this course will prepare students for future challenges and opportunities in data science. Prerequisites: M21 506, M21 560, M21 570, and M21 550. Matrix algebra is also highly recommended.

Credit 3 units.