THE DEPARTMENT OF BIOSTATISTICS

DOCTORAL STUDENT HANDBOOK

2024–2025

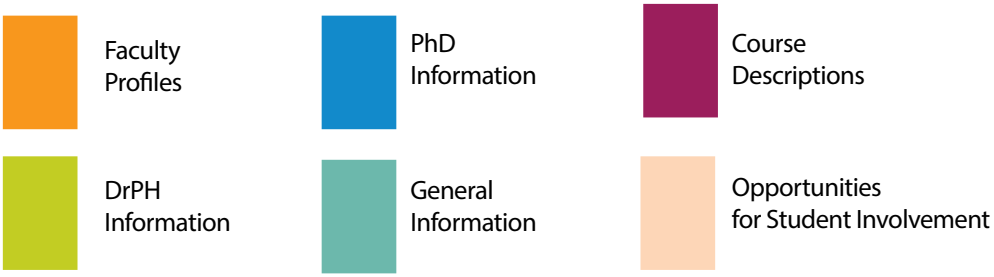
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Table of Contents

Information Key

Side Markers

Look for these colored markers on the central right hand side of each page as you scroll to quickly identify sections.



Colored Text

Within the text of each section there are several colors used to emphasize or call out information.

Green colored text is a clickable hyperlink
Blue colored text is a section header
Magenta colored text is the name of a course
Orange colored text emphasizes forms, requirements, or other improtant information

The Department of Biostatistics

Biostatistics is the science of developing and applying statistical methods for quantitative studies in biomedicine, health, and population sciences.

Biostatisticians play a crucial role in research design, collection and organization of data, analysis, presentation, and interpretation of results. Career opportunities are usually found in governmental agencies, private industry, and medical research institutions.

The Department of Biostatistics maintains collaborative relationships with other units of the University and with outside agencies and institutions. Among the many affiliated institutions and centers are: Columbia University Irving Medical Center, New York State Psychiatric Institute, the Department of Statistics at Columbia's Morningside Campus, the Gertrude H. Sergievsky Center (research in the field of neuroepidemiology), the Herbert Irving Comprehensive Cancer Center and Institute of Cancer Research, the HIV Center for Clinical and Behavioral Studies, and the Irving Center for Clinical Research.

Faculty in the Department of Biostatistics work at the frontier of public health, leading research teams that investigate some of today's most pressing health issues. Recruited from the top universities from around the world, the faculty bring to the school a wealth of experience that serves to inform their research and teaching.

HOWARD ANDREWS (PhD, Rutgers University)
Associate Professor at CUMC of Neuroscience (in Biostatistics)

Research interests: Clinical trials, data management systems, multi-level analysis, perinatal outcomes, environmental factors, Alzheimer's disease

SRIKESH ARUNAJADAI (PhD, University of California Berkeley)
Adjunct Assistant Professor of Biostatistics

Research interests: Statistical applications in solving problems in Biology, statistical research in time series, point processes, spatio-temporal processes, and modeling of various dynamic systems

KIROS BERHANE (PhD, University of Toronto)
Cynthia and Robert Citron-Roslyn and Leslie Goldstein Professor and Chair

Research interests: Longitudinal data modeling, multi-level growth curve models, nonparametric regression, multiple outcomes, quantile regression, mediation, applications to environmental data

MELISSA D. BEGG (ScD, Harvard University)
Professor of Clinical Biostatistics, and Dean of the Columbia School of Social Work

Research interests: Analysis of clustered data, oral health research, mental health statistics, clinical research training

XIAOYU CHE (PhD, Claremont Graduate University)
Assistant Professor of Biostatistics (in the Center for Infection and Immunity) at CUMC

Research interests: Design, development, and application of statistical methods for "multi-omics" analyses to bring new insights into the pathogeneses of chronic and neurodevelopmental diseases

QIXUAN CHEN (PhD, University of Michigan)
Associate Professor of Biostatistics

Research interests: Survey sampling, missing data, measurement error, Bayesian statistics, latent class modeling, integrative data analysis

BIN CHENG (PhD, University of Wisconsin-Madison)
Professor of Biostatistics at CUIMC

Research interests: Linear and generalized linear mixed models, statistical analysis of clinical trials, longitudinal non-normal data modeling, statistical computing, statistical inference on manifolds

YING-KUEN KENNETH CHEUNG (PhD, University of Wisconsin-Madison)
Professor of Biostatistics

Research interests: Design and analysis of clinical trials, methods in toxicology studies and bioassay, applications of Monte Carlo methods, nonparametric methods, bioethics

CODRUTA CHIUZAN (PhD, Medical University of South Carolina)
Adjunct Assistant Professor of Biostatistics

Research interests: Applications of statistical methods in psychiatric research, functional data analysis, statistical machine learning, treatment regime estimation and evaluation, integration of multi-source data

HANGA GALFALVY (PhD, University of Illinois)
Associate Professor of Biostatistics (in Psychiatry) at CUIMC

Research interests: Statistical methodology in psychiatric research, with a special focus on the prediction models for suicidal behavior from high-dimensional data, censored regression models, statistical genetics, and longitudinal data analysis in observational studies

JEFF GOLDSMITH (PhD, Johns Hopkins University)
Associate Professor of Biostatistics, and Associate Dean for Data Science

Research interests: Functional data analysis, high-dimensional regression, longitudinal data analysis, smoothing, Bayesian variable selection, neuroimaging, and accelerometry

PRAKASH GORROOCHURN (PhD, Monash University)
Associate Professor of Clinical Biostatistics at CUIMC

Research interests: Mathematical population genetics, genetic mapping of complex diseases

TIAN GU (PhD, University of Michigan)
Assistant Professor of Biostatistics

Research interests: Robust & efficient data integration in precision health research, methods for use in biobank data, COVID and disparity research

WENPIN HOU (PhD, University of Michigan)
Assistant Professor of Biostatistics

Research interests: Developing statistical machine learning methods, mathematical modeling of gene regulatory networks

JIANHUA HU (PhD, University of North Carolina-Chapel Hill)
Professor of Biostatistics (in Medicine and in the Herbert Irving Comprehensive Cancer Center)

Research interests: high-dimensional genomics/proteomics, imaging, and longitudinal data, modeling disease heterogeneity, and adaptive designs to achieve personalized treatments

IULIANA IONITA-LAZA (PhD, New York University)
Professor of Biostatistics (in Medicine and in the Center for Precision Medicine and Genomics)

Research interests: Statistical genetics and bioinformatics

HAOMIAO JIA (PhD, Case Western University)
Professor of Biostatistics (in Nuring) at CUMC

Research interests: Small area estimation, data smoothing, temporal-spatial analysis, survey sampling

ZHEZHEN JIN (PhD, Columbia University)
Professor of Biostatistics

Research interests: Survival analysis, resampling methods, ROC curves, smoothing methods, nonparametric regression, clinical trials

SEONJOO LEE (PhD, University of North Carolina-Chapel Hill)
Associate Professor of Clinical Biostatistics (in Psychiatry)

Research interests: Neuroimaging, cognitive neuroscience, machine learning, latent variable analysis, multivariate time series, stochastic process, and functional data analysis

SHING M. LEE (PhD, Columbia University)
Professor of Biostatistics at CUIMC

Research interests: Rapid dose finding techniques in Phase I trials, and the development of more sensitive endpoints (e.g. Toxicity Burden Scores) in Phase I Trials

CHENG-SHIUN LEU (PhD, Columbia University)
Professor of Biostatistics at CUIMC

Research interests: Adaptive design in Phase II clinical trials, clinical trials, design and analysis for clinical and behavioral studies.

MOLEI LIU (PhD, Harvard University)
Assistant Professor of Biostatistics

Research interests: High dimensional statistics, federated learning, distributed learning, transfer learning, semi-supervised learning, semi-parametric statistical methods and theory, electronic health record (EHR) data analysis

YING LIU (PhD, Columbia University)
Assistant Professor in Clinical Biostatistics, Mental Health Data Science, Department of Psychiatry

Research interests: Bridging Machine learning and Deep Learning to Biostatistical methodology research and land these methods to Psychiatry Research

ZHONGHUA LIU (PhD, Harvard University)
Assistant Professor of Biostatistics

Research interests: Statistical genetics/genomics, epigenetics, semiparametric efficiency theory, causal mediation analysis for integrative genomics and fairness, missing data problems, measurement error, and machine learning methods

SHAW-HWA LO (PhD, UC Berkeley)
Professor of Statistics and of Biostatistics

Research interests: Survival Analysis, Design of Clinical Trials, Resampling Methods, Coverage Problems, Analysis of Incomplete Data, Nonparametric Methods, Asymptotic Theory, Empirical Bayes Methodology. Statistical Genetics. Bioinformatics. Genetic network analysis.

XIN MA (PhD, Emory University)
Assistant Professor of Biostatistics at CUIMC

Research interests: High-dimensional statistics, variable selection, functional data analysis, measurement error, imaging applications, and deep learning methods

CHRISTINE MAURO (PhD, Columbia University)
Associate Professor of Biostatistics at CUIMC

Research interests: Analysis of clinical trials, longitudinal data analysis, statistical applications in mental health and substance use research, including evaluating state-level policies, statistics education

IAN MCKEAGUE (PhD, University of North Carolina at Chapel Hill)
Professor of Biostatistics

Research interests: Survival analysis, competing risks in HIV/AIDS studies, inference for stochastic processes, empirical likelihood, Markov chain Monte Carlo, functional data analysis, semiparametric efficiency, Bayesian statistics, and martingale and counting process methods

DANIEL MALINSKY (PhD, Carnegie Mellon University)
Assistant Professor of Biostatistics

Research interests: Causal inference, graphical models, missing data, stochastic processes, machine learning, algorithmic fairness, social & environmental determinants of health, health disparities

CALEB MILES (PhD, Harvard University)
Assistant Professor of Biostatistics

Research interests: Causal inference, HIV, Interference, Measurement error, Mediation analysis, Semiparametric inference

TODD OGDEN (PhD, Texas A&M University)
Professor of Biostatistics (in Psychiatry)

Research interests: Analysis of brain imaging data, functional data analysis, nonparametric regression, wavelet applications, statistical modeling

MARTINA PAVLICOVA (PhD, Ohio State University)
Associate Professor of Biostatistics at CUIMC

Research interests: Analysis of clinical trials, longitudinal data and analysis, multiple comparisons methods, hurdle and zero-inflated models, modern teaching methods

MIN QIAN (PhD, University of Michigan)
Associate Professor of Biostatistics

Research interests: Medical decision making, dynamic treatment regimes, variable selection/model selection for decision making, statistical machine learning, reinforcement learning, statistical inference, bootstrap, empirical processes, concentration inequalities, stochastic processes

YIFEI SUN (PhD, Johns Hopkins University)
Assistant Professor of Biostatistics

Research interests: General biostatistical methodology for survival, longitudinal and multivariate data, machine learning,electronic health record data, wearable device data

JOHN L.P. (SEAMUS) THOMPSON (PhD, University of California-Los Angeles)
Professor of Biostatistics and Neurology at CUMC

Research interests: Randomized clinical trials, trial design, neurology, data management systems

NAITEE TING (PhD, Colorado State University)
Adjunct Professor of Biostatistics

Research interests: Clinical development of new drugs, dose selection, Phase II

LINDA VALERI (PhD, Harvard University)
Assistant Professor of Biostatistics

Research interests: Causal inference, machine learning, statistical methods for measurement error, mediation analysis, missing data, multivariate survival, longitudinal and time series data, applications in perinatal epidemiology, mental health, environmental health, and health disparities

MELANIE WALL (PhD, Iowa State University)
Professor of Biostatistics (in Psychiatry)

Research interests: Latent variable modeling, spatial, and longitudinal data analysis

SHIKUN WANG (PhD, University of Texas, UTHealth Houston)
Assistant Professor of Biostatistics (in the Herbert Irving Comprehensive Cancer Center) at CUMC

Research interests: Developing novel statistical methods for health service research using joint modeling of longitudinal and survival data

SHUANG WANG (PhD, Yale University)
Professor of Biostatistics

Research interests: Statistical genetics, genetic epidemiology, quantitative trait loci analysis

YUANJIA WANG (PhD, Columbia University)
Professor of Biostatistics

Research interests: Machine learning, generative models, precision medicine, electronic health records, network analysis, mental health, and neurological disorders

YING WEI (PhD, University of Illinois-Urbana Champaign)
Professor of Biostatistics

Research interests: Quantile regression methods, growth charts estimation, longitudinal data analysis, semiparametric modeling, and robust statistics

XIAO WU (PhD, Harvard University)
Assistant Professor of Biostatistics

Research interests: Causal inference, environmental statistics, statistical learning, nonparametric statistics, Bayesian biostatistics, applications to climate science and health

Staff

PAUL MCCULLOUGH (pm2692@cumc.columbia.edu)
Director, Academic Programs

LUCIA LI (ll3414@cumc.columbia.edu)
Assistant Director, Academic Programs

ERIN ELLIOTT (ee2548@cumc.columbia.edu)
Program Coordinator (Communications)

KATY HARDY (ch336@cumc.columbia.edu)
Department Administrator

ANTHONY GUERRERO (gg292@cumc.columbia.edu)
Senior Grants Administrator

GEORGIA ANDRE (gg292@cumc.columbia.edu)
Administrative Coordinator

CAROLINA MELLADO (cm2556@cumc.columbia.edu)
Director, Personnel Operations

GESSIE MESAMOURS (gm3064@cumc.columbia.edu)
Executive Assistant to Chair, Dr. Kiros Berhane

LUMINITA HELLMANN (lh2023@cumc.columbia.edu)
Senior Financial Analyst

Doctoral Degree Progams Overview

The Department of Biostatistics offers two doctoral degree programs: the Doctor of Public Health (DrPH) and the Doctor of Philosophy (PhD). Both the DrPH and PhD programs train candidates to apply state-of-the art statistical methods to the analysis of important public health issues and potential solutions, but differ in their relative emphasis on application versus statistical theory.

Doctor of Public Health (DrPH)

DrPH training places relatively greater emphasis on the application of statistical methods to public health problems, although many DrPH students propose new methods and contribute to the advancement of statistical theory as part of their dissertation research.

Upon satisfactory completion of the DrPH degree in Biostatistics, graduates will be able to:

Data Analysis and Computing

- Identify and implement advanced statistical models for the purposes of estimation, comparison, prediction, and adjustment in non-standard settings

Public Health and Collaborative Research

- Describe the foundations of public health, including the biological, environmental, behavioral, and policy factors that affect the health of populations
- Develop and execute calculations for power and sample size when planning research studies with complex sampling schemes
- Formulate and prepare a written statistical plan for analysis of biomedical or public health research data that clearly reflects the research hypotheses of the proposal in a manner appropriate for scientists with varied backgrounds

Consulting

- Function as an effective consultant in biomedical and public health research projects
- Communicate and write effectively in order to describe complex topics in a consulting environment.

Data Management

- Identify the uses to which data management can be put in practical statistical analysis, including the establishment of standards for documentation, archiving, auditing, and confidentiality; guidelines for accessibility; security; structural issues; and data cleaning
- Differentiate between analytical and data management functions through knowledge of the role and functions of databases, different types of data storage, and the advantages and limitations of rigorous database systems in conjunction with statistical tool
- Assess database tools and the database functions of statistical software, with a view to explaining the impact of data management processes and procedures on their own research

Teaching

- Explain and illustrate principles of study design and data analytic techniques to public health students enrolled in first and second level graduate public health courses

Biostatistical Research

- Identify and integrate new developments in the statistical literature for challenging research problems in public health
- Generate efficient computer code to implement sophisticated statistical techniques

Doctor of Philosophy (PhD)

PhD training places relatively greater emphasis on the development of novel statistical theory and methods. A PhD dissertation must represent an original contribution to statistical theory or methods that has relevance to a real biomedical or public health application.

Upon satisfactory completion of the PhD degree in Biostatistics, graduates will be proficient in:

Biostatistical Research

- Identify and integrate new developments in the statistical literature for challenging research problems in biomedicine and public health
- Generate efficient computer code to implement sophisticated statistical techniques
- Recognize gaps in current inferential methods that limit further public health research and propose and develop solutions based on rigorous theoretical considerations

Data Analysis and Computing

- Identify and implement advanced statistical models for the purposes of estimation, comparison, prediction, and adjustment in non-standard settings.

Public Health and Collaborative Research

- Develop and execute calculations for power and sample size when planning research studies with complex sampling schemes;
- Formulate and prepare a written statistical plan for analysis of biomedical or public health research data that clearly reflects the research hypotheses of the proposal in a manner appropriate for scientists with varied backgrounds; and
- Evaluate research reports and proposals for research funding on the basis of their scientific integrity, validity, and the strength of the quantitative analysis.

Consulting

- Function as an effective consultant in biomedical and public health research projects; and
- Develop communication and writing skills in a consulting environment.

Data Management

- Identify the uses to which data management can be put in practical statistical analysis, including the establishment of standards for documentation, archiving, auditing, and confidentiality; guidelines for accessibility; security; structural issues; and data cleaning
- Differentiate between analytical and data management functions through knowledge of the role and functions of databases, different types of data storage, and the advantages and limitations of rigorous database systems in conjunction with statistical tools
- Describe the different types of database management systems, the ways these systems can provide data for analysis and interact with statistical software, and methods for evaluating technologies pertinent to both

Teaching

- Explain and illustrate selected principles of study design, probability theory, inference, and data analytic techniques to public health students enrolled in first and second level graduate public health courses
- Explain advanced concepts in the theory of statistical inference to graduate students in biostatistics and mathematical statistics

Doctor of Public Health

Director: Shing Lee, PhD

The Doctor of Public Health degree in Biostatistics (DrPH) prepares candidates to apply modern statistical methods to the solution of important public health problems as leaders of multidisciplinary research teams. The degree program is administered by the Standing Doctoral Committee of the Mailman School of Public Health, which carries out faculty policy on admission to the doctoral program and upholds the criteria for granting the degree.

Course Requirements

The Doctor of Public Health degree calls for completion of an approved program of study totaling no less than 36 doctoral credits. Upon completion of 36 credits of coursework, a student is permitted to take the written qualifying examination. In some instances, it may be determined by the Department that a student needs more than 36 post-MPH course credits before the qualifying examination.

No more than 10 credits may be tutorials, and no more than six may be earned at 6000-level courses at the Mailman School of Public Health or 4000-level courses at the Graduate School of Arts and Sciences; the Department may apply to the Doctoral Committee for a variance on the six-credit rule on a case-by-case basis.

A list of required courses is given below. A DrPH student who has not previously earned an MPH degree must complete courses in each of the core areas of public health: biostatistics, environmental health sciences, epidemiology, health policy, and social and behavioral sciences. Core courses will have to be completed through the Columbia MPH Core or completed by taking a comparable graduate level course at an accredited institution. These comparable courses must be approved by the School's Doctoral Committee. The credits accrued for completing the core requirements DO NOT count towards the 36 doctoral credits.

A grade of B or better is necessary in all required courses. Up to 2 elective courses may be taken pass/fail.

Training in Interdisciplinary Research

The curriculum is designed to enable students to integrate their training in statistical methods and theory with the role of biostatistical consultant/collaborator on interdisciplinary teams, which will comprise a major portion of their future professional practice. Statistical Practices and Research for Interdisciplinary Sciences I & II (P9185 & P9186) are courses in which students gain experience with design, data analysis, and both oral and written presentation communication through exposure to several consulting projects. DrPH students are required to enroll in P9185 prior to taking the Qualifying Exam, and to enroll in P9186 after taking the Qualifying Exam. Students with extensive work experience in the field may request to waive the P9186 requirement from the Director of the DrPH Program.

Criteria to Remain in Good Academic Standing in a DrPH Program

Consistent with CEPH guidelines for award of the DrPH, students are required to complete a minimum of 36 credits of coursework with a combined average grade point average of B+ (3.3), and pass all required and selective courses with a B or above. Students must meet the MSPH continuous registration policy, successfully complete a mentored Applied Practice project, and complete the competency-based portfolio. Only students in good standing may sit for the qualifying exam and proceed to undertaking an Integrative Learning Experience. Students not achieving these requirements are placed on academic notice and must develop a remediation plan in consultation with their advisors. The satisfactory progress of doctoral students is assessed annually through the Annual Progress Report on the basis of academic performance. A student who fails to maintain satisfactory progress will be alerted to their deficiencies, advised of the means to remedy them, and told the consequences of his or her failure to do so. This may include academic probation. Academic probation is set by the Office of Student Affairs in collaboration with the Biostatistics Department. The student is informed that they are on probation and notified of any conditions set. A student who fails to maintain satisfactory progress after such a probationary period will have their candidacy terminated. In cases of egregious failure to achieve progress, a breach of the honor code or other serious misconduct, a student may be dismissed from the degree program without a probationary period.

Departmental Colloquium

All doctoral students are required to attend the Departmental Colloquium and Research Talks held weekly each semester. Dates, times, and locations will be posted on the Department electronic board, as well as on the Department's website and social media.

Curriculum (TOTAL POINTS: 36 OR MORE)

Required Courses		Points
P8104*	Probability	3
P8105*	Data Science I	3
P8106*	Data Science II	3
P8108*	Survival Analysis	3
P8109*	Statistical Inference	3
P8120*	Analysis of Categorical Data	3
P8130*	Biostatistical Methods I	3
P8131*	Biostatistical Methods II	3
P8157*	Analysis of Longitudinal Data	3
P9185	Statistical Practices and Research for Interdisciplinary Sciences I	3
P9186	Statistical Practices and Research for Interdisciplinary Sciences II	1.5
P9070	DrPH Case Studies in Public Health Leadership I & II	3
P9050	DrPH Seminar in Strategic Management	1.5
P9040	DrPH Seminar in Management and Organizational Behavior	1.5
P9060	Essentials of Teaching and Communication for Doctoral	1.5

*requirements can be waived, consult with your faculty advisor

Sample Elective Courses		Points
P8142	Clinical Trial Methodology	3
P8116	Design of Medical Experiments	3
P8122	Statistical Methods for Causal Inference	3
P8123	Analysis of Health Surveys	3
P8140	Introduction to Randomized Clinical Trials	3
P8158	Latent Variable and Structural Equation Modeling for Health Sciences	3
P8160	Topics in Advanced Statistical Computing	3
P8180	Relational Databases and SQL Programming for Research and Data Science	3
P9120	Topics in Statistical Learning and Data Mining I	3
89260	Building Interdisciplinary Research Models	2

*elective courses not listed require approval from the program director

The Applied Practice Experience (APEx) for DrPH Students

Regardless of the amount or level of prior experience, all DrPH students are required to engage in an applied practice experience (APEx) in which students are responsible for completion of at least one project that is meaningful for an organization and to advance public health practice.

The work product may be a single project or a set of related projects that demonstrate a depth of competence. The deliverable must contain a reflective component that includes the student's expression of personal and/or professional reactions to the applied practice experience. This may take the form of a manuscript, journal article or other written product, a professional portfolio, or another deliverable that serves to assess the ability of the student to meet department and school competencies.

The applied practice experience (APEx) should take place within an organization external to the student's school or program so that it is not merely an academic exercise, but application of learning to a "real world" setting. Relevant organizations may include governmental, non-governmental, non-profit, industry and for-profit settings. The Office of Field Practice and individual departments identify sites in a manner that is sensitive to the needs of the agencies or organizations involved, and sites should benefit from Mailman students' experiences. The applied practice experience (APEx) may be completed within a student's own work setting, as long as the applied practice experience (APEx) differs substantially from a student's current job description and meets the required competencies described below.

The applied practice experience (APEx) must meet a minimum of five (5) foundational and/or concentration-specific competencies that are reinforced and/or assessed through application. One of these competencies must be a school-wide or a departmental-specific competency in leadership, management, and governance. Competencies for the applied practical experience (APEx) must be agreed upon by the student, advisor, and applied learning experience preceptor, as specified in the statement of work form.

While there is not a minimum number of hours for the applied practice experience, it does require substantive, quality opportunities that address the identified competencies.

Students must complete the Scope of Work (SOW) form prior to starting a practicum experience. The SOW, which is managed by the Mailman's Office of Careers and Practice, is an important tool for planning the practicum and meeting the School's requirements for engaging in a structured practicum process.

Qualifying Examination

There is a two-part qualifying examination for all DrPH candidates in Biostatistics that must be completed before going on to the oral comprehensive examination.

Basic Inference

The first part assesses basic familiarity with statistical inference as presented in the course Statistical Inference (P8109). Students who have taken this course (or a comparable graduate course) and have received a grade of B+ or above automatically satisfy the basic inference requirement. All others will be required to take a written examination testing their knowledge of the material in this course. In all cases, students must fulfill the requirement within two years of starting the doctoral program. Students must pass the basic inference requirement before they may sit for the Applications exam.

Applications

The Applications portion covers the practical analysis of data. This part focuses on addressing applied problems requiring statistical inference based on data analysis. The purpose of the Applications exam is to ensure that the student is able to determine the appropriate statistical and analytic approaches needed to solve real world public health / medical problems, correctly interpret the statistical results from these approaches, and translate and summarize those findings into language that public health and medical professionals would find useful.

Course Work and Progressing toward the Applications portion
Preparation should include additional coursework in skills classes, review and thorough understanding of the material in the suggested readings, group and individual study sessions, completion of timed practice tests, and enrollment in Statistical Practices and Research for Interdisciplinary Sciences I (P9185), a course in which students gain exposure to real world design, analysis, and report writing.

With approval and consent of his or her academic advisor, the student should inform the Director of Academic Programs two months in advance of sitting for the Applications portion of the exam.

Grading on the Applications exam
Grading is holistic, and can also take into account performance in coursework, and other factors deemed relevant. A score below 65% on the exam would generally be considered unsatisfactory. The student will be allowed no more than two attempts to pass either the Basic Inference or Applications parts of the examination. The Applications portion must be taken and passed by the end of the third year in the DrPH program.

Questions from prior years are available to the student to assist in preparing for the examination.

Reading List

The following list consists of textbooks that are used in the courses required for the DrPH degree, plus additional references which are generally at the appropriate level for the DrPH Qualifying Examinations. Those marked with an asterisk are highly recommended to students preparing for their examinations.

Breslow NE and Day NE, Statistical Methods in Cancer Research

Conover WJ, Practical Nonparametric Statistics

Cox DR and Oakes D, Analysis of Survival Data

Fleiss JL, The Design and Analysis of Clinical Experiments

* Fleiss JL, Levin B, and Paik MC, Statistical Methods for Rates and Proportions

Hogg RV and Craig AT, Introduction to Mathematical Statistics

* Hosmer D and Lemeshow S, Applied Logistic Regression

* Johnson RA and Wichern DW, Applied Multivariate Statistical Analysis

Kalbfleisch JD and Prentice RL, Statistical Analysis of Failure Time Data

Kleinbaum DG and Kupper LL, Applied Regression Analysis and other Multivariable Methods

* Lawless JF, Statistical Models and Methods for Lifetime Data

* Lee ET, Statistical Methods for Survival Data Analysis

Lehmann ER, Nonparametrics: Statistical Methods Based on Ranks

Mardia KV, Kent JT, and Bibby JM, Multivariate Analysis

* Mood AM, Graybill FA, and Boes D, Introduction to Statistical Inference

Morrison DF, Multivariate Statistical Methods

* Mosteller F and Tukey JW, Data Analysis and Regression

* Neter J, Wasserman W, and Kutner MH, Applied Linear Statistical Models

Rao CR, Linear Statistical Inference and Its Applications

Scheffe H, The Analysis of Variance

Searle SR, Linear Models

Snedecor GW and Cochran WG, Statistical Methods

Tukey JW, Exploratory Data Analysis

Oral Comprehensive Examination

After completing all course work and passing the qualifying examination described above, the DrPH candidate begins planning for the Integrative Learning Experience (ILE). The oral comprehensive examination for the DrPH in Biostatistics is intended to examine the student's mastery of the current state of knowledge about his or her project area, and thus to indicate whether the student is prepared to undertake such a project. The Oral Comprehensive Examination should be taken no later than six months after passing the qualifying exam.

Composition of the Examining Committee

The examining committee will consist of five members approved by the chair of the Doctoral Program Subcommittee on Biostatistics, and will include:

i) three members who are inside examiners (i.e. holding a formal appointment or approved as a dissertation sponsor);

ii) preferably two (but at least one) members who are outside examiners.

The latter faculty should represent disciplines closely related to the area of application of the student's proposed research. After the sponsor obtains consent from each member, the faculty sponsor submits the list of names to the Chair of the Department and to the Chair of the Departmental Subcommittee on Biostatistics (DPSOB) for approval, who then recommends the student's committee to the DrPH Committee of the Mailman School of Public Health.

Scheduling the Exam

The oral comprehensive examination should be taken within one year of passing the qualifying examination.

Nature of the Examination

After the committee selection and approval process has been completed, the student submits in writing a description of the current state of knowledge about the proposed area of research. This submission should be from 15 to 25 pages in length and contain between 15 and 20 references. This paper serves as the basis for the oral comprehensive examination. The student must give each member of the Examining Committee this written submission and discuss with each any additions or deletions that the committee member feels should be incorporated in the write-up. Since the final written submission and the references therein will constitute the basic material upon which the student will be examined, each member of the committee and the student must come to an agreement on the scope of the submission. It should be neither too narrow nor too broad in scope. After all members of the ad hoc committee approve the submission, the examination is scheduled within the next 60 days. The written submission may contain original results by the student, but this is not required.

ORAL COMPREHENSIVE EXAMINATION CONTINUED

Format of the Exam

The actual examination shall be an Oral Comprehensive Examination conducted by the Examining Committee as follows:

1. The chair of the Examining Committee will not be the ILE advisor but another member of the ad hoc committee.
2. The examination will run approximately two hours and will consist of an oral presentation of the content of the written submission by the student (a planned presentation of about 30 minutes is appropriate), which may be interrupted by members of the Examining Committee with appropriate questions on the material presented or relevant related material. The chair of the Examining Committee may challenge any question felt to be unrelated to the written submission and its background material.
3. After the presentation and questions, each member may ask additional questions of the examinee. Any such questions should be within the broad content of the written submission and its references. Again, the Examining Committee chair may challenge any question felt to be too far removed from the basic material upon which the examination is based, namely on the written submission and the references therein.
4. After all questions are completed, the examinee leaves the room and the committee then votes on whether or not the examinee passed the examination. All members must agree in order for the student to pass the examination. Instead of pass or fail, the committee may unanimously decide upon the option of retesting the student within a six-month period on the same written submission.

The committee's decision will be put into writing by the chair of the Examining Committee, as well as brief comments on the strengths and weaknesses of the student's performance as deemed necessary. Copies of this statement will be sent to the student and placed in the student's file.

Second Attempt at Passing

The student is entitled to no more than two attempts at passing the Oral Comprehensive Examination. The second attempt need not be based on the same written submission nor be examined by the same committee, but the same rules will govern the second attempt, including approval by the committee of the written submission. The second attempt must be made no more than 6 months after the first attempt.

Upon passing the Oral Comprehensive Examination, a student will typically ask his sponsor or another member of the faculty to agree to serve as the student's sponsor. No formal approval of an ILE topic is required; however, a suitable and mutually agreeable topic must be established by the student and advisor. As stated earlier, it is often the case that the Oral Comprehensive Examination is on a topic related to the student's ILE, although this is not a formal requirement.

Progressing toward the Integrated Learning Experience Defense

Between the Oral Exam and the ILE Defense, the DrPH student is required to present his/her project in two public settings. The first is the Doctoral Research Seminar, usually held in the spring, where doctoral students present their work to the faculty and their peers. The second setting is the preparation and presentation of a paper (or poster) at a conference of professional societies or at a statistics or biostatistics departmental talk for job interviews. A select, but not exhaustive, list of such societies is presented below. More information is available on the Doctoral Bulletin Board. Travel funds are often available.

Example of Professional Societies / Associations:

- American Statistical Association (ASA)
- American Public Health Association (APHA)
- International Biometric Society (ENAR/IBS)
- Joint Statistical Meetings (JSM)
- Society for Clinical Trials (SCT)

The Integrated Learning Experience (ILE)

Once a DrPH student has advanced to doctoral candidacy, they begin to develop a proposal for the ILE project. The topic must deal with an important problem or issue in public health which can be addressed by the sound and original application of existing statistical methods. It must demonstrate that the candidate has engaged in independent and original research that has advanced our understanding of or knowledge about the public health problem, though the methods themselves need not be original. After the project is successfully defended, the doctoral degree is awarded by the Mailman School of Public Health in the Faculty of Medicine.

In most cases, completion of DrPH course work and written qualifiers should take no more than two full-time academic years. On average, the ILE may take an additional two or three full-time academic years. An overall time limit of nine years is set from the date of first registration as a doctoral student.

In unusual instances a student may wish to change ILE sponsors, for instance, if the student's project requires different areas of expertise than originally anticipated. In such cases the student may seek approval from a new faculty sponsor. The candidate must inform the Department Chair and the previous sponsor that the new sponsor will assume the previous sponsor's duties. At this point the student may also decide to pursue a new project topic, with approval of the new sponsor, but in all cases the rules governing time limits and extensions remain in force.

DrPH candidates are required to submit an electronic copy of their final report to the department. Copies of past reports are available from the Director of Academic Programs.

Some Past DrPH ILE Titles

The titles below are provided to give students some idea of ILE topics which in past years have proved appropriate for the DrPH degree:

Symptom Cluster Analysis for Depression Treatment Outcomes and Growth Mixture Models for Analysis Association between Social Media Use Patterns and Anxiety Symptoms in Young Adults, Ying Chen (2023)

Analysis Approaches for Wearable Device Data, Patrick Hilden (2021)

Statistical Methods for Healthcare Cost Data: An Application to Administrative Claims Data for Pediatric Patients with Acute Lymphoblastic Leukemia, Elisabetta Malangone Monaco (2021)

Clustering Algorithm for Zero-Inflated Data, Anusorn Thanataveerat (2020)

Statistical Issues in Platform Trials with a Shared Control Group, Jessica Overbey (2019)

Bayesian Modeling of Latent Heterogeneity in Complex Survey Data and Electronic Health Records, Rebecca Anthopolos (2019)

Statistical Methods for the Study of Etiologic Heterogeneity, Emily Zabor (2019)

Statistical Methods for Integrated Cancer Genomic Data Using a Joint Latent Variable Model, Esther Drill (2018)

Bayesian Modeling for Mental Health Surveys, Sharifa Williams (2018)

Data-Driven Methods for Identifying and Validating Shorter Symptom Criteria Sets: The Case for DMS-5 Substance Use Disorders, Cheryl Raffo (2018)

Design and Analysis of Sequential Multiple Assignment Randomized Trial for Comparing Multiple Adaptive Interventions, Xiaobo Zhong (2018)

Prognostic Modeling in the Presence of Competing Risks: An Application to Cardiovascular and Cancer Mortality in Breast Cancer Survivors, Nicole Leoce (2016)

New Estimating Equation Approach for the Secondary Trait Analyses in Genetic Case-Control Studies, Xiaoyu Song (2015)

Identifying Patterns in Behavioral Public Health Data Using Mixture Modeling with an Informative Number of Repeated Measures, Gary Yu (2013)

A Life Expectancy-based Comprehensive Quantification of Structural-level Health Disparities, Emma Benn (2012)

An Index of Aging-Relatedness with Relevance to Genetic and Environmental Contributions to Mortality and Disease Incidence in a Population, Gilberto Levy (2011)

The Familial Aggregation of Epilepsy, Anna Peljto (2010)

Analysis of the MT CT-Plus Initiative: An Application of a Piecewise Multilevel Latent Variable Regression Model, Shean-Sheng Wang (2010)

Analyses on Double-Blinded Fetal Tissue Transplant Study on Patients with Severe Parkinson's Disease and Reliability Assessment on Outcome Measurements, Richard Kao (2010)

Stepwise Procedures for Dose Finding in an Adaptive Clinical Trial of Early Rehabilitation After Acute Stroke, Xi Wu (2010)

Registration

DrPH students matriculating in and after 2023 are expected to be working professionals in public health or related fields throughout their studies and may complete the DrPH program on a part-time or full-time academic schedule.

DrPH students must maintain continuous registration every semester from the start of the program until deposit of the doctoral dissertation. After completion of all coursework students register for Doctoral Registration (RSRHP0001) each term until they are ready to graduate.

Doctor of Philosophy

Director: Yuanjia Wang, PhD

The Doctor of Philosophy in Biostatistics (PhD) prepares candidates for leadership roles in the development and application of statistical methods to biomedical research for the advancement of public health. The PhD is awarded by the Graduate School of Arts and Sciences (GSAS) as governed by the Doctoral Program Subcommittee on Biostatistics. The program is administered by the faculty and staff of the Mailman School of Public Health.

Course Requirements

Students take courses in the department of biostatistics, and other academic units representing various fields of application and/or related background material. A student should plan his or her course work in consultation with his or her academic advisor and/or the PhD subcommittee chair. Students wishing to waive one or more required courses must request approval in writing from their faculty advisor and the Director of Academic Programs.

A grade of B or better is necessary in all required courses, except for P9111 which requires a B+ or better. Electives may be taken pass/fail, in order to encourage candidates to take courses outside his or her field of experience.

In advance of beginning the PhD program, any student who has not previously completed a bachelor's or master's degree in public health from a CEPH-accredited school or program, will need to complete the Introduction to Public Health non-credit, asynchronous requirement (administered by the Office of Educational Initiatives). Since this is not a course, you do not need to register. Throughout the requirement, you are expected to be actively engaged in the material and approach the work with the same effort you would a for-credit course.

Training in Interdisciplinary Research

The curriculum is designed to enable students to integrate training in statistical methods and theory with the role of biostatistical collaborator on interdisciplinary teams, which will comprise a major portion of their future professional practice. Statistical Practices and Research for Interdisciplinary Sciences I & II (P9185 & P9186) are courses in which students gain experience with design, data analysis, and both oral and written communication through exposure to several consulting projects. PhD students are required to enroll in P9185 during the spring semester prior to taking the Qualifying Exam, and to enroll in P9186 during the fall semester after taking the Qualifying Exam.

Statistical Inference Problem Seminar

To prepare for the written component of the Qualifying Exam, students are required to take the problem seminar in which students work on problems and discuss problem solving strategies useful for theoretical questions. The problem seminar is held in the months prior to the written portion of the Qualifying Exam.

GSAS Requirements

In addition to registering for individual courses, PhD students are required to register for the Residence Unit (RU) which provides the basis for tuition charges and provides full-time status. Six RU are required for the PhD degree. RUs may only be earned during fall and spring semesters, not during the summer. PhD students must register for 1 RU each semester up to the total required 6 RUs. After one year of study, students who enter with a Master's degree may apply for advanced standing of two residence units representing work completed in their Master's program. After the student has satisfied the residency requirement they must register for full-time Matriculation & Facilities (M&F) status until a successful dissertation defense.

Criteria to Remain in Good Academic Standing in a PhD Program

To remain in good standing, students must (1) receive grades of B or higher in all required courses and (2) achieve an overall grade point average (GPA) of B+ (3.3) or higher in required courses. Students whose grades do not qualify for good standing will be reviewed by the Doctoral Committee with input from the course director, the student, and their academic advisor. Based on this review, the Committee will determine whether the student should:

1) Continue in the program in good-standing without further conditions; or

2) Continue on a probationary basis until specified conditions (e.g., additional remedial coursework, tutoring) are met (at which point the student returns to good-standing); or

3) Be dismissed from the program. Dismissal from the program requires review by the Department Chair and review by the MSPH Office of Student Affairs.

Only students in good-standing may sit for the qualifying examinations.

Departmental Colloquium

All doctoral students are required to attend the Departmental Colloquium and Research Talks held weekly each semester. Dates, times, and locations will be emailed, as well as posted on the Department's website and social media.

Curriculum

Required Courses		Points
P6400*	Principles of Epidemiology	3
P8104*	Probability	3
P8105*	Data Science I	3
P8106*	Data Science II	3
P8109*	Statistical Inference	3
P8130*	Biostatistical Methods I	3
P8131*	Biostatistical Methods II	3
P8160*	Topics in Advanced Statistical Computing	3
P9104	Probability for Biostatisticians	3
P9109	Theory of Statistical Inference I	3
P9110	Theory of Statistical Inference II	3
P9111	Asymptotic Statistics	3
P9120	Topics in Statistical Learning and Data Mining I	3
P9130	Advanced Biostatistical Methods I	3
P9185	Statistical Practices and Research for Interdisciplinary Sciences I	3
P9186	Statistical Practices and Research for Interdisciplinary Sciences II	1.5

*requirements can be waived if taken a comparable course at the Master's level; consult with your faculty advisor

Sample Elective Courses		Points
P8108	Survival Analysis	3
P8116	Design of Medical Experiments	3
P8122	Statistical Methods for Causal Inference	3
P8123	Analysis of Health Surveys	3
P8133	Bayesian Analysis and Adaptive Designs in Clinical Trials	3
P8140	Introduction to Randomized Clinical Trials	3
P8142	Clinical Trial Methodology	3
P8144	Pharmaceutical Statistics	3
P8157	Analysis of Longitudinal Data	3

*elective courses not listed require approval from the program director

Qualifying Examination

There is a two-part qualifying examination for all PhD candidates in Biostatistics that must be completed prior to the oral comprehensive examination. The written and take-home portions of the exam are to be taken during the same summer semester.

Written Portion - Theory and Methods

The written or theory and method exam draws from material presented in the following MS and doctoral level courses: P8104, P8109, P8130, P8131, P9104, P9109, P9110, and P9130. The purpose of the written exam is to ensure that the PhD student is able to fully understand and use the mathematical and theoretical tools that form the basis of doctoral level biostatistical research. The exam requires solutions to five questions. Students entering with a Bachelor's are expected to take the exam after their second year; students entering with a relevant Master's are expected to take the exam after their first year.

Course Work and Progressing toward the written portion of the Qualifying Exam
Preparation should include coursework or mastery of content of the material in the required courses, review and thorough understanding of the material in the suggested readings, group and individual study sessions, and completion of timed practice tests. With approval and consent of the student's academic advisor, the student should inform the Director of Academic Programs two months in advance of sitting for the written portion of the Qualifying Exam.

Take Home Portion - Applications

The take-home exam covers the practical analysis of data. The examination focuses on applied problems requiring statistical inference based on data analysis, with particular emphasis on material from P8105, P8106, P8130, P8131, P9130 and P9185. The purpose of the take-home exam is to ensure that the student is able to determine the appropriate statistical and analytic approaches needed to solve real world public health / medical problems, correctly interpret the statistical results from these approaches, and translate and summarize those findings into language that public health and medical professionals would find useful. The take-home exam is administered over a two-day period. Students are encouraged to use personal laptops and any familiar software. Students entering with a Bachelor's are expected to take the exam after their second year; students entering with a relevant Master's are expected to take the exam the summer after their first year.

Course Work and Progressing toward the take home portion of the Qualifying Exam
Preparation should include additional coursework in skills classes, review and thorough understanding of the material in the suggested readings, group and individual study sessions, completion of timed practice tests, as well as enrollment in P9185. With approval and consent of his or her academic advisor, the student should inform the Director of Academic Programs two months in advance of sitting for the take-home exam.

Grading on the Qualifying Exam

Grading is holistic, taking into account performance in coursework, on both portions, and other factors deemed relevant. A score below 65% on either the written or take-home portion will generally be considered unsatisfactory. The student will be allowed no more than two attempts at passing either part of the exam. It is strongly recommended that the second attempt be made at the time of the next exam offering.

Exam questions from prior years are available to the student to assist in preparing for the examination.

Reading List

The following list consists of textbooks that are generally appropriate to use for preparing for the PhD qualifying examination.

Agresti A, Categorical Data Analysis

Bickel PJ and Doksum KA, Mathematical Statistics

Casella G and Berger RL, Statistical Inference

Cox D and Hinkley DV, Theoretical Statistics

Efron B and Tibshirani R, An Introduction to the Bootstrap

Hastie T, Tibshirani R, and Friedman J, The Elements of Statistical Learning

Hettmansperger TP and McKean JW, Robust Nonparametric Methods

Hollander M, Nonparametric Statistical Methods

Johnson RA and Wichern DW, Applied Multivariate Statistical Analysis

Klein JP and Moeschberger ML, Survival Analysis

Lehmann EL, Point Estimation

Lehmann EL, Testing Statistical Hypotheses Lehmann EL, Elements of Large-Sample Theory

McCullagh P and Nelder JA, Generalized Linear Models

Rao CR, Linear Statistical Inference and Its Applications

Robert CP and Casella, G, Monte Carlo Statistical Methods

Ruppert D, Wand MP, and Carroll R, Semiparametric Regression

Shao J, Mathematical Statistics

Wickham, H and Grolemund, G, R for Data Science

Oral Comprehensive Examination

After completing all course work and passing the two-part qualifying examination described in the previous sections, the PhD candidate begins planning for dissertation research and preparing for the Oral Exam. The Oral Comprehensive Examination is intended to demonstrate the student's mastery of the material in a defined statistical content area by verbally presenting a thorough description of the state of the art in that area, identifying limitations or areas of incomplete knowledge in that area, and proposing the development of new methods that would advance that area. This topic area may or may not end up being the student's dissertation topic. The Oral Comprehensive Examination should be taken approximately one year after passing the two-part qualifying exams. Fellows in the program, please note all tuition expenses incurred as a result of any delay in scheduling this exam shall be the responsibility of the student and not the Department of Biostatistics.

Composition of the Examining Committee

The examining committee will consist of at least four members approved by the chair of the Doctoral Program Subcommittee on Biostatistics, and will include:

i) three members who are inside examiners (i.e. holding a formal appointment or approved as a dissertation sponsor);

ii) preferably two (but at least one) members who are outside examiners.

The chair of the Examining Committee will be a member of the Doctoral Program Subcommittee on Biostatistics. One member of this committee should be the faculty member who acts as the student's sponsor and anticipated thesis advisor. With the consent of the members of the proposed committee, the faculty sponsor then submits their names for approval by the Chair of the Doctoral Program Subcommittee on Biostatistics.

Nature of the Examination

After the committee selection and approval process has been completed, the student submits in writing a description of the current state of knowledge about the proposed area of research. This submission should be from 15 to 25 pages in length and contain between 15 and 20 references. This paper serves as the basis for the oral comprehensive examination. The student must give each member of the committee this written submission and discuss with each any additions or deletions that the committee member feels should be incorporated in the write-up. Since the final written submission and the references therein will constitute the basic material upon which the student will be examined, each member of the committee and the student must come to an agreement on the scope of the submission and references. After such modifications to the written submission have been approved by all four members of the Examining Committee, the Comprehensive Exam is scheduled within the next 30 days. The written submission may contain original research by the student but need not be original in content. It should not be too narrow in scope and should reflect the necessary basic material relevant to the student's chosen area of research. Before and during the examination, the three faculty examiners other than the student's sponsor make suggestions for and may insist on changes in the student's perception of the topic. Part of the student's written submission is an enumeration of as yet unanswered questions. The examiners make their opinions plain as to how important and challenging they perceive these questions to be.

ORAL COMPREHENSIVE EXAMINATION CONTINUED

Format of the Exam

The actual examination shall be an Oral Comprehensive Examination conducted by the Examining Committee as follows:

1. The chair of the Examining Committee will not be the dissertation advisor but some other member of the ad hoc committee.
2. The examination will run approximately two hours and will consist of an oral presentation of the content of the written submission by the student (a planned presentation of about 30 minutes is appropriate), which may be interrupted by members of the Examining Committee with appropriate questions on the material presented or related material. The chair of the Examining Committee may overrule any question felt to be unfair or unrelated to the written submission and its background material.
3. After the presentation and questions, each member may ask additional questions of the examinee. Such questions should be within the broad scope of the written submission and references. Again, the Examining Committee chair may rule against any questions felt to be too far removed from the basic material upon which the examination is based, that is, the written submission and the references therein.
4. After all questions are completed, the examinee leaves the room, and the committee then votes on whether or not the examinee passed the exam. Three of the four members must vote to pass the student in order for the student to pass the exam.

The committee's decision will be put into writing by the chair of the Examining Committee, as well as brief comments on the strengths and weaknesses of the student's performance as deemed necessary. Copies of this statement will be sent to the student and placed in the student's file.

Second Attempt at Passing

The student is entitled to no more than two attempts at passing the Oral Comprehensive Examination. The second attempt need not be based on the same written submission nor be examined by the same committee, but the same rules will govern the second attempt, including approval by the committee of the written submission. The second attempt must be made no more than 6 months after the first attempt.

The examination and written submission are designed to focus the examination on basic material, which is important to the student's area of research, and allow the Examining Committee to judge that the student fully comprehends this material. Upon passing the Comprehensive Examination, a student will typically ask his sponsor or another member of the PhD subcommittee to serve as the student's dissertation advisor and sponsor. No formal approval of a dissertation topic is required; however, a suitable and mutually agreeable topic must be established by the student and advisor. While it is usually the case that the Oral Comprehensive Examination is on a topic that will become the student's dissertation topic, this is not a formal requirement.

Advancement of PhD Students to the Master of Philosophy Degree

Upon the student's passing the qualifying and oral comprehensive examinations and the successful completion of six residence units beyond the Master's degree (two residence units awarded for a completed Master's degree), he or she is awarded the Master of Philosophy degree. Failure on the Oral Comprehensive Examination implies that it is the Subcommittee's judgment the student is not yet prepared to carry out original research. The awarding of the Master of Philosophy to a student, on the other hand, certifies that the student has mastered the fundamental material necessary for him or her to conduct research in biostatistics. Students who apply for and receive two residence units of advanced standing are still required to complete four additional residence units before the Master of Philosophy may be awarded.

Progressing toward the Dissertation Defense

Between the Oral Exam and the Dissertation Defense, the PhD student is required to present his or her research in two public settings. Typically, one of these settings is at a Graduate Research Seminar, where doctoral students at various stages of their research present their work to their peers. A second setting is the preparation and presentation of a paper (or poster) at a conference of professional societies. A select, but not exhaustive, list of such societies is presented below. Students who are selected to present at a conference can apply for travel funds at the School and department levels. Information requests about available travel funds should be directed to the Director of Academic Programs.

Example of Professional Societies / Associations:

- American Statistical Association (ASA)
- American Public Health Association (APHA)
- International Biometric Society (ENAR/IBS)

Dissertation

The PhD dissertation is expected to contain original results in statistical theory and methods in the solution of a problem which has relevance to a biomedical application. As a rule, the content of the dissertation should be adequate for publication in peer-refereed journals in the topic area of the dissertation. Students begin work on their dissertation research with the approval of their thesis sponsor and comprehensive examination committee. The only time limitation is the Graduate School of Arts and Sciences maximum of seven years from the time of enrollment in the doctoral program (the maximum is six years for those receiving advanced standing). Candidates who are making satisfactory progress toward finishing the dissertation have, upon application, been granted extensions by the Dean of GSAS, with the approval of their sponsor. With proper advising, PhD students should be able to finish the degree within five years of entry into the PhD program.

PhD candidates are required to submit an electronic copy of their final dissertation to the department. Electronic copies of past dissertations are available from the Director of Academic Programs.

DISSERTATION CONTINUED

In some cases a student may wish to change dissertation sponsors – for instance, if the student's research leads to different areas of expertise than originally anticipated. In such cases, the student may seek approval from a new faculty sponsor. The candidate must inform the Doctoral Program Subcommittee Chair and the previous sponsor that the new sponsor will assume the previous sponsor's duties. At this point, the student may also decide to pursue a new dissertation topic, with approval of the new sponsor, but in all cases the rules governing time limits and extensions still apply. Upon completion of the dissertation, and with approval of the candidate's dissertation committee, the dissertation defense is scheduled.

PhD candidates are required to submit an electronic copy of their final dissertation to the department.

For more details regarding the PhD dissertation, the student is referred to the Dissertation Office website: www.gsas.columbia.edu/content/dissertation. The GSAS Dissertation Office is located on the Columbia Morningside Heights campus at 107 Low Memorial Library, 535 W. 116th Street, New York, NY 10027. Information is also available in the Department of Biostatistics and the Dean's Office of GSAS on Morningside Campus.

Some Past PhD Dissertation Titles

The titles below are provided to give students some idea of topics that in past years have proved appropriate for the PhD degree:

Multi-level Latent Variable Models for Integrating Multiple Phenotypes for Mental Disorders, Yinjun (Abby) Zhao (2024)

Innovations in Functional Data Analysis with Applications in Women's Health and Neuroscience, Madison Stoms (2024)

Causal Inference and Time Series Methods for N-of-1 Mobile Health Studies with Missing Data, Charlotte Fowler (2024)

Modeling and Causal Inference Methods for Estimating and Transporting an Environmental Mixture's Effect, Melanie Mayer (2024)

Causal Inference for Health Effects of Time-varying Correlated Environmental Mixtures, Zilan Chai (2024)

Machine Learning Methods for Intensive Longitudinal Data and Causal Inference in Multi-Study, Multi-Outcome Settings, Soohyun Kim (2024)

Dose Optimization Methods for Novel Cancer Therapies in the Presence of Patient Heterogeneity, Rebecca Silva (2024)

Phase II/III Transitional Seamless Trial Designs with Different Objectives and Endpoint Types, Robert A. Tumasian III (2023)

Correcting for Measurement Error and Misclassification using General Location Models, Muhire Kwizera (2023)

Computational Algorithms for Multi-omics and Electronic Health Records Data, Jia Guo (2023)

General Bayesian Calibration Framework for Model Contamination and Measurement Error, Siquan Wang (2023)

Statistical Methods for Learning Patients Heterogeneity and Treatment Effects to Achieve Precision Medicine, Tianchen Xu (2022)

The Joint Modeling of Longitudinal Covariates and Censored Quantile Regression for Health Applications, Bo Hu (2022)

Statistical Analysis of Large Scale Data with Perturbation Subsampling, Yujing Yao (2022)

Statistical methods for modeling progression and learning mechanisms of neuropsychiatric disorders, Qinxia Wang (2021)

Bayesian modeling in personalized medicine with applications to N-of-1 trials, Ziwei Liao (2021)

Dynamic graphical models and curve registration for high-dimensional time course data, Erin McDonnell (2021)

Statistical and machine learning methods for precision medicine, Yuan Chen (2021)

Topics in Bayesian design and analysis for sampling, Yutao Liu (2021)

GGQ-learning for indefinite horizon problem with L1 penalty, Xiaoqi Lu (2021)

Optimal Treatment Regimes for Personalized Medicine and Mobile Health, Eun Jeong Oh (2020)

Statistical Learning Methods for Depression Screening and Intervention, and Structured Missing Imputation, Huichen Zhu (2019)

Quantile regression for zero-inflated outcomes, Wodan Ling (2019)

Functional Data Analytics for Wearable Device and Neuroscience Data, Julia Wrobel (2019)

Statistical Methods for Constructing Heterogenous Biomarker Networks, Shanghong Xie (2019)

Machine Learning Methods in Personalized Medicine Using Electronic Health Records, Peng Wu (2019)

Statistical Methods for Genetic Studies with Family History of Diseases, Annie Lee (2019)

Varying-Coefficient Models and Functional Data Analysis for Dynamic Network and Wearable Device Data, Jihui Lee (2018)

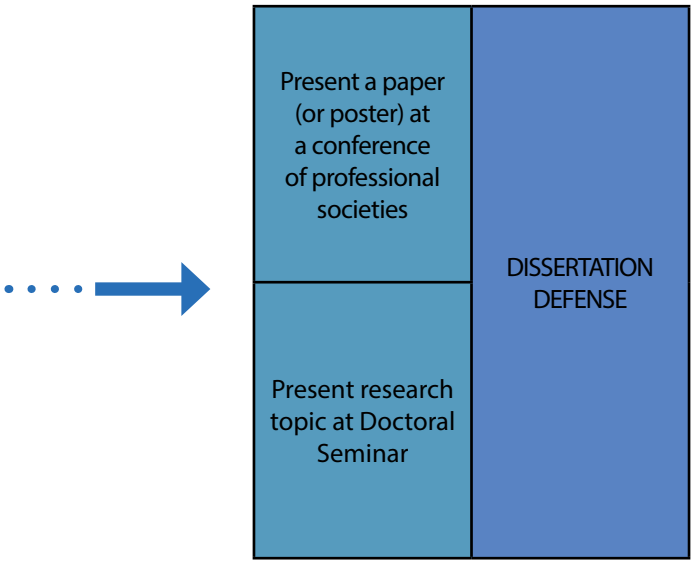
Statistical Methods for Modeling Biomarkers of Neuropsychiatric Disease, Ming Sun (2018)

Typical PhD Timeline (PhD student entering with a relevant Masters)

<table>
<tr><th>Fall I</th><th>Spring I</th><th>Summer I</th><th>Fall II</th><th>Spring II</th><th rowspan="2">Present a paper (or poster) at a conference of professional societies</th><th rowspan="8">DISSERTATION DEFENSE</th></tr>
<tr><td>P9104 Probability for Biostatisticians</td><td></td><td rowspan="7">QUALIFYING EXAMS Written/Take-home</td><td>P9111 Asymptotics</td><td rowspan="3">Formalize Research Topic</td></tr>
<tr><td>P9130 Advanced Methods I</td><td>P8160 Topics in Advanced Statistical Computing</td><td rowspan="2">P9120 Topics in Statistical Learning & Data Mining I</td><td rowspan="6">Present research topic at Doctoral Seminar</td></tr>
<tr><td>P9109 Theory of Statistical Inference I</td><td>P9110 Theory of Statistical Inference II</td></tr>
<tr><td rowspan="2">Electives</td><td>P9185 SPRIS I</td><td rowspan="2">P9186 SPRIS II</td><td rowspan="3">Prepare for Oral Exam</td></tr>
<tr><td>Electives</td></tr>
<tr><td colspan="3">Statistical Inference Problem Seminar</td><td>Electives</td></tr>
<tr><td colspan="6">Attend all Departmental Colloquium and Research Talks</td></tr>
</table>

Typical PhD Timeline (PhD student entering without a relevant Masters)

Fall I	Spring I	Summer I	Fall II	Spring II	Summer II	Fall III	Spring III
P6400 Principles of Epidemiology	P8109 Statistical Inference	Review material; begin research projects	P9104 Probability for Biostatisticians		QUALIFYING EXAMS (Written/Take-Home)	P9111 Asymptotics	Formalize Research Topic
P8104 Probability	P8131 Biostatistical Methods II			P8160 Topics in Advanced Statistical Computing		P9120 Topics in Statistical Learning & Data Mining I	
P8105 Data Science	P8106 Data Science II Elective		P9109 Theory of Statistical Inference I	P9110 Theory of Statistical Inference II		P9186 SPRIS II Electives	Prepare for Oral Exam
P8130 Biostatistical Methods I			P9130 Advanced Methods I	P9185 SPRIS I			
			Statistical Inference Problem Seminar				
Attend all Departmental Colloquium and Research Talks							



Attend all department colloquia and research talks

T32 Fellowship on Research Training in Mental Health Biostatistics and Data Science

Full-time doctoral students who are US citizens or permanent residents are eligible to receive the NIH-sponsored Ruth L. Kirschstein National Research Service Award (pre-doctoral T32 award). Mental health disorders are a major cause of disability that significantly reduces the quality of life. As new technologies emerge for assessing mental health disorders and as research design becomes increasingly complex, the Department of Biostatistics at Columbia University, in partnership with Columbia's Department of Psychiatry and New York State Psychiatric Institute, has developed an innovative training program.

This two-year program aims to train predoctoral scholars in biostatistical methods, machine learning, data science, and interdisciplinary research in order to meet the emerging challenges brought by cutting edge technologies and big data. Building on the Department's strengths and decades of success in rigorous training, the program prepares trainees to become next-generation leaders in mental health biostatistics and data science. Overall, this T32 program provides well-structured and in-depth training experience that will equip trainees with valuable skills and practice in biostatistics and mental health. This program will enable trainees to succeed in their future research careers in the intersection of biostatistics, data science and interdisciplinary areas such as mental health.

Interested doctoral students are encouraged to contact Program Directors, Drs. Yuanjia Wang and Todd Ogden, for more information.

Financial Support

The Columbia University Graduate School of Arts and Sciences (GSAS) and all Departments at the Mailman School of Public Health consider PhD training to be full-time effort for 12 months per year. The University's commitment to fully funding our trainees pursing a PhD is to allow them to focus entirely on all aspects of their training: coursework, research, teaching, dissertation-related activities, professional development, other training activities (e.g. seminars, colloquia, journal clubs, workshops), and service to the institution. PhD trainees, therefore, are discouraged from taking on additional employment.

Trainee appointments are set at 20 hours per week, irrespective of funding source (T32, F31, R36, TL1, NSF, GRA, or other support administered through the University). The educational components comprise the balance of full-time effort. PhD trainees are permitted to hold paid teaching positions of employment at Columbia during all terms in addition to their PhD trainee appointments (e.g. T32, F31, R36, TL1, GRA, or other apprenticeships), provided the MTE has been satisfied and cannot exceed 8 hours per week.

- Payment: MSPH PhD students will receive financial support at the stipulated guaranteed level for a given year. For the 2024-25 academic year, PhD students will be paid $48,080. The source of this financial support may be from one or more of the following sources: the Department, a training grant, a grant-funded Graduate Research Assistantship (GRA), another sponsored project, philanthropy, or other outside sources of funding. Stipends to training grant fellows are disbursed three times a year, and stipends to non-training grant students are paid twice a month. In both cases, the annual amount is the same: $48,080.

  Note: If a student in years 1-5 secures external funding, the externally-supporting stipend funding is applied towards the annual stipend, not awarded in addition to it.

- Tuition and Fees: MSPH will cover tuition & fees and health insurance for PhD students during the supported period, ensuring they can focus on their studies without financial burdens. Some sources of funding (e.g. training grant) help subsidize the tuition for our students who are part of a given training program.

Duration of Support

MSPH PhD students are guaranteed financial support for five (5) years, contingent upon satisfactory academic progress, adherence to program guidelines, School and University policies, and available funding.

Funding beyond five (5) years is not guaranteed. Students may not be funded on a training grant (e.g. T32) beyond year five (5). However, in year six (6), should a student secure an external award in their name from the National Institutes of Health (NIH) (e.g. F31, R36), or grants from foreign research entities or foundations that provide funding comparable to NIH funding levels for predoctoral students, departments will supplement the funding in year six (6) to bring students' financial support to at least union-mandated levels. Supplemental funding for year six (6) is subject to departmental and institutional review, including considerations of satisfactory academic progress. Neither School nor Departmental funding will be considered, nor will be available, beyond year six (6). Should a student's mentor wish to provide supplemental funding beyond year six (6) for the student who has secured external funding in their name (e.g. F31, R36), the School and the Department will permit the funding.

PhD students who are interested in pursuing external awards to support their dissertation research should plan to submit applications no later than the end of year three (3).

Administrative Policy

Parental Accomodation Policy

Parental accommodation policy (for Biostatistics PhD students) is governed by GSAS. Please refer to their set policy here on the GSAS website.

Time-Off Policy for Doctoral Students on Research and Teaching Appointments

In doctoral study, the periods between semesters (fall to spring, spring to fall) are times of active research, research training, and teaching preparation, rather than considered holidays.

However, in any given academic year, student officers on twelve-month research or teaching appointments are entitled to 10 days of vacation time per year without loss of compensation. Student officers of instruction who hold a teaching appointment for two consecutive terms in an academic year (fall to spring) are entitled to 8 days of vacation without loss of compensation.

Student officers observe the University academic holiday schedule. Those who are required to work on a University academic holiday shall receive an alternate day off approved in advance by their supervisor. Student officers of research who plan to take time off at other times may do so only after coordinating with a) their advisor or PI. Student officers of instruction may take vacation during academic breaks or as otherwise mutually agreed to by the faculty instructor or lead course coordinator. This will ensure that the requested time off does not conflict with the responsibilities attendant to the research enterprise or teaching the course.

Student officers of instruction and research are entitled to 1 paid personal day per semester. The personal day must be approved in advance in the same manner as vacation time described above. In addition, personal days must be used during the semester in which they are earned.

More information can be found here on the GSAS website.

Registration & Tuition

The PhD requires continuous registration and the completion of six Residence Units (RUs) prior to undertaking the qualifying exams. Tuition is calculated on a flat-fee basis, and not by individual course. Students entering with a master's or other terminal degree (e.g. MD) are eligible for "advanced standing" which reduces the number of RUs required to four. Advanced standing is determined after they have completed one semester of coursework in the PhD program. Students may register for a full or a half RU. The four RUs may be accumulated as four full RUs, eight half RUs, or a combination of both. Students registered for a full RU are considered full-time. While the RU is not itself a course, it is assigned a course ID number for registration purposes. PhD Students registered for a full RU can take as many courses as they want; students registered for a half RU may enroll in three courses per semester. Matriculation & Facilities (M&F) registration entitles the student to use university facilities but not to take courses.

After completion of all required courses, teaching experience ,students maintain continuous registration through Matriculation and Facilities (M&F), according to the following guidelines:

> Students who are defending a dissertation must register for M&F unless they have already distributed the dissertation.

After successfully pass the Qualifying Exam (which includes the Inference Exam and Application Exam), and complete the oral exam defense, students are awarded the MPhil degree.

Mentored Research Experience

Students are expected to engage Mentored Research Experiences (MRE), which vary by department, and which are necessary to meet the academic and training requirements of all MSPH PhD programs, and are supported by the funding package all PhD students receive, irrespective of funding source. MREs should be geared towards developing the knowledge base and research skills necessary to supporting the transition to independence as investigators. For students supported on a GRA with grant funding and specific responsibilities, the MRE constitutes the responsibilities outlined in GRA appointment letters. For students supported by training grants, the MRE should be negotiated with the program directors. For all MREs, the priority should be the development of essential research expertise and skills through mentorship that will ultimately enable students to complete successfully their PhD dissertations.

Mentored Teaching Experience

All PhD students at the Mailman School must engage in two semesters of a Mentored Teaching Experience (MTE). These two semesters constitute academic requirements. Beyond the two semester MTE, students may have to fulfill additional teaching requirements in their department, or may engage in elective additional teaching, both for additional compensation. Additional compensation is capped at 8 hours per week.

Professional Development

MSPH encourages and supports PhD students in presenting their research at national and international conferences to foster their professional development. PhD students may have access to funds, when available, to support their research endeavors, including for conference attendance, research materials, and other necessary resources, subject to Departmental and/or School approval.

Additional Work

The School and University permits PhD trainees to take on additional work throughout the year. This additional work cannot exceed 8 hours per week. Moreover, any additional work cannot interfere with a trainee's progress toward degree completion and is subject to a range of federal grant restrictions. Any ad comp work paid through Columbia must be pre-approved by the Department, the Dean's Office, and the Sponsored Projects Administration via submission to the Salary Review Committee, and must adhere to the stipulated pay rates for PhD students determined by the School, and meet the minimum compensation requirement stipulated by the UAW-SWC contract.

Work outside Columbia is not governed by the University, School, or Program, but must still conform with NIH rules, and should not exceed 10 hours per week, and should not hinder progress toward degree completion to degree (as evaluated by each Department).

Scholarships, foundation grants, and other awards (e.g. conference travel award) that do not require a student to be engaged in a traditional employee/employer relationship or a quid pro quo work effort are permissible but must be declared to the Department.

Students who wish to participate in externships with compensation in excess of the GRA minimum may waive their summer payments to accept the external payment. The School will continue to fund student health insurance and tuition, and fees. Externships must be approved by the Department, the Dean's Office, and Sponsored Projects Administration via submission to the Salary Review Committee.

General Information for Students

University Resouces Outside the Department

These are just a few of the many resources that exist within Columbia Unviersity that you may find helpful to know about or have on hand in your time as a student:

Office of Student Affairs (OSA)

The Office of Student Affairs (OSA) offers opportunities, social and alumni networking, personal advising, resources referrals, and assistance with navigating registration and student enrollment. From orientation to graduation, the office ensures students receive support and tools for a robust educational experience by advocating for our students' wellbeing, providing space for engagement and helping students build skills for success OSA also maintains the Student Conference and Leadership Experience (SCALE) program to help fund student travel for conferences or other educational experiences.

Office of Careers and Practice

The Office of Careers and Practice provides support for students throughout their Columbia Mailman education and beyond to their careers in public health. For students completing their Applied Practice Experience (APEx), the office can help in finding placements or submitting forms. Their Career Services staff helps students explore career goals, network, improve their resume, write cover letters, and negotiate salaries along with hosting regular events such as career fairs on campus.

The Office of Disability Services (ODS)

The Office of Disability Services (ODS) helps to facilitate equal access or students, including coordination of reasonable accommodations and support services for students with disabilities. ODS works with students with all types of disabilities, including physical, learning, sensory, psychological, AD/HD, and chronic medical conditions. ODS also helps students with temporary injuries and illnesses. The department is dedicated to accommodating student and maintaining equal access while preserving their confidentiality, in line with the provisions of the Americans with Disabilities Act. To register with the Office of Disability Services, students must complete a Graduate Application for Accommodations and Services, and submit documentation of their disability.

Ombudsman Office

The Ombudsman Office serves as a confidential, informal, and impartial resource for faculty, students, and staff who may have a University-related concern or problem. Their office is independent of University administration and adheres to the practices of the International Ombudsman Association. The identities and communications of those who contact the Ombudsman Office are held in strict confidentiality except as required by law, or where, in the judgment of the Ombuds Officers, there appears to be imminent risk of serious harm.

Well-Being and Health Promotion Programs

The Well-being and Health Promotion team leads, facilitates, and participates in collective, systemic, and innovative action for student health and well-being at CUIMC. They offer activities such as mindfulness, yoga, pet therapy, crafternoons, and much more. Subscribe to their weekly newsletter to stay up to date with their ongoing offerings.

Student Services for Gender-based and Sexual Misconduct

The Sexual Respect website offers a hub for services related to gender-based and sexual misconduct including confidential and non-confidential resources. You can find links to Student Health Services, Title IX Offices, Gender-Based Misconduct Offices, University Policy, and Spiritual Counseling among other resources.

Student Honor Code and Professional Guidelines

To clarify the school's expectation of professional and ethical conduct, the administration, faculty, and students have adopted the Honor Code and Professional Guidelines linked above to complement University policy.

Forms

These are a few forms that may be useful to you in your time as a student

Registration Form

Leave of Absence Request Form

Request for Incomplete Form

Dissertation Defense Form (DrPH Only)

Department of Biostatistics Courses

These are the complete list of all courses offered by the Department of Biostatistics. Some may not be applicable for doctoral students. Due to faculty commitments, the frequency of the courses changes from time to time. Students are advised to check the current schedule of courses listed on the MSPH web page: www.mailman.columbia.edu/academics/courses. Students may also review the course offerings of the Statistics Department at the Morningside Campus in the Graduate School of Arts and Sciences: www.stat.columbia.edu.

Students are encouraged to meet with their faculty advisors at least twice a year (in the fall and in the spring). Permission is not required for approved courses in a student's approved program of study. Students must first obtain permission from their faculty advisors to take courses outside the approved

P6103 Introduction to Biostatistics 3 points

Prerequisites: Permission of the instructor required for all non-Public Health students.

Biostatistics is essential to ensuring that findings and practices in public health and biomedicine are supported by reliable evidence. This course covers the basic tools for the collection, analysis, and presentation of data in all areas of public health. Central to these skills is assessing the impact of chance and variability on the interpretation of research findings and subsequent recommendations for public health practice and policy. Topics covered include: general principles of study design; hypothesis testing; review of methods for comparison of discrete and continuous data including ANOVA, t-test, correlation, and regression.

P6104 Introduction to Biostatistical Methods 3 points

Prerequisites: Instructor's permission for non-Biostatistics students

An enriched core course for students concentrating in biostatistics and others who expect to take additional courses in biostatistics beyond the two main second-level courses (P8100 and P8120). It covers in greater depth all of the topics in P6103 and is the best preparation for students anticipating a quantitative orientation in their degree programs. Topics covered include standard distributions, measures of central tendency and dispersion, hypothesis testing, point estimation, confidence intervals, and an introduction to correlation and regression.

P6110 Statistical Computing with SAS 3 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations core course

A logical follow-up course to an introductory biostatistics course. Covers uses of the computer in cleaning, summarizing, and cross-classifying data. Enhancement of the material covered in P6103/P6104— including regression, correlation, and contingency table analysis, and the analysis of variance–with data analysis carried out using SAS software.

P6170 New Drug Development: A Regulatory Overview 3 points

Prerequisite: P6104, P8130 or MPH Quantitative Foundations core course and P6400

Provides our CTSA fellows and scholars with insights into and understanding of the process of patient oriented/translational research and gives them an opportunity to meet active investigators from academia and industry, and learn about some career enhancing resources available at CUMC. Active researchers from various clinical disciplines and public health are invited to speak on research techniques, design, and laboratory methodology as applied to current studies. They present their experiences in conducting patient orientated research on the Health Sciences campus and elsewhere. Also features speakers from both the pharmaceutical and biotech industries who discuss drug development, and preclinical and clinical trials. Other lectures deal with FDA regulations, patent law, and the Institutional Review Board and ways to effectively build and succeed in a clinical/translational academic career.

P8100 Applied Regression I 3 points

Prerequisites: P6104 or MPH Quantitative Foundations core course. (Not open to MS/TM, PHDS, PS, or SG tracks)

This course will provide an introduction to the basics of regression analysis. The class will proceed systematically from the examination of the distributional qualities of the measures of interest, to assessing the appropriateness of the assumption of linearity, to issues related to variable inclusion, model fit, interpretation, and regression diagnostics.

P8101 Introduction to Health Data Science 3 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations

This course will introduce students to core data science skills and concepts through the exploration of applied biostatistics. The course will begin with an introduction to the R programming language and the RStudio IDE, focusing on contemporary tidyverse functions and reproducible programming methods. Then, the course will instruct students in contemporary data manipulation and visualization tools while systematically covering core applied biostatistics topics, including confidence intervals, hypothesis testing, permutation tests, and logistic and linear regression. Finally, the semester will end with an introduction to machine learning concepts, including terminology, best practices in test/training sets, cross-validation, and a survey of contemporary classification and regression algorithms.

P8103 Colloquium on Patient Oriented Research 2 points (0.5 points x 4 semesters)

Prerequisite: MS-POR students only

Provides our CTSA fellows and scholars with insights into and understanding of the process of patient oriented/translational research and gives them an opportunity to meet active investigators from academia and industry, and learn about some career enhancing resources available at CUMC. Active researchers from various clinical disciplines and public health are invited to speak on research techniques, design, and laboratory methodology as applied to current studies. They present their experiences in conducting patient orientated research on the Health Sciences campus and elsewhere. Also features speakers from both the pharmaceutical and biotech industries who discuss drug development, and preclinical and clinical trials. Other lectures deal with FDA regulations, patent law, and the Institutional Review Board and ways to effectively build and succeed in a clinical/translational academic career.

P8104 Probability 3 points

Prerequisites: P6104 or P8130 (may be corequisite), working knowledge of calculus

Topics include: Fundamentals, random variables, and distribution functions in one or more dimensions; moments, conditional probabilities, and densities; Laplace transforms and characteristic functions. Infinite sequences of random variables, weak and strong large numbers; central limit theorem.

P8105 Data Science I 3 points

Prerequisites: Experience in R programming (or programming in another language) and data analysis is recommended

Contemporary biostatistics and data analysis depends on the mastery of tools for computation, visualization, dissemination, and reproducibility in addition to proficiency in traditional statistical techniques. The goal of this course is to provide training in the elements of a complete pipeline for data analysis.

P8106 Data Science II 3 points

Prerequisites: P8105

With the explosion of "Big Data" problems, statistical learning has become a very hot field in many scientific areas. The goal of this course is to provide the training in practical statistical learning. It is targeted to MS students with some data analysis experience.

P8107 Introduction to Mathematical Statistics 3 points

Prerequisities: MPH Quantitative foundations or P6104 (Not open to MS/TM, PHDS, or SG tracks)

The first portion of this course provides an introductory-level mathematical treatment of the fundamental principles of probability theory, providing the foundations for statistical inference. Students will learn how to apply these principles to solve a range of applications. The second portion of this course provides a mathematical treatment of (a) point estimation, including evaluation of estimators and methods of estimation; (b) interval estimation; and (c) hypothesis testing, including power calculations and likelihood ratio testing.

P8108 Survival Analysis 3 points

Prerequisites: P8104, P8109, and P8130

This course focuses on methods for the analysis of survival data, or time-to-event data. Survival analysis is a method for survival data or failure (death) time data, that is time-to-event data, which arises in a number of applied fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography. A special course of difficulty in the analysis of survival data is the possibility that some individual may not be observed for the full time to failure. Instead of knowing the failure time t, all we know about these individuals is that their time-to-failure exceeds some value y where y is the follow-up time of these individuals in the study. Students in this class will learn how to make inference for the event times with censored.

P8109 Statistical Inference 3 points

Prerequisites: P8104, working knowledge of calculus and linear algebra

This course covers a review of mathematical statistics and probability theory at the Masters level. Students will be exposed to theory of estimation and hypothesis testing, confidence intervals and Bayesian inference. Topics include population parameters, sufficient statistics, basic distribution theory, point and interval estimation, introduction to the theory of hypothesis testing, and nonparametric procedures.

P8110 Applied Regression II 3 points

Prerequisites: P6104 or MPH Quantitative Foundations core course, and P8100 (Not open to MS/TM, PHDS, or SG tracks)

An introduction to the application of statistical methods in survival analysis, generalized linear models, and design of experiments. Topics to be covered include estimation and comparison of survival curves, regression models for survival data, log-linear models, logit models, analysis of repeated measurements, and the analysis of data from blocked and split-plot experiments. Examples are drawn from the health sciences.

P8112 Systematic Review and Meta-Analysis 1.5 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations core course and P6400

Research synthesis using systematic review and meta-analysis is one of the most valuable of research endeavors, and can be a particularly rewarding experience for junior investigators who want to develop expertise in a specific area of public health or medicine by producing a product with significant scientific impact. This course will combine lecture and workshop elements to introduce students to the principles and practices of systematic review and metaanalysis. It will be targeted to students who have previously been introduced to the concepts of basic biostatistics, epidemiology, and clinical trials.

P8116 Design of Medical Experiments 3 points

Prerequisites: P8104, P8109, and P8130

This course covers the fundamental principles and techniques of experimental designs in clinical studies. Topics include reliability of measurement, linear regression analysis, parallel groups design, analysis of variance (ANOVA), multiple comparison, blocking, stratification, analysis of covariance (ANCOVA), repeated measures studies; Latin squares design, crossover study, randomized incomplete block design, and factorial design.

P8119 Advanced Statistical and Computational Methods in Genetics and Genomics 3 points

Prerequisites: P6104 or P8130

This course introduces students to advanced computational and statistical methods used in the design and analysis of high-dimensional genetic data, an area of critical importance in the current era of Big Data. The course starts with a brief background in genetics, followed by in depth discussion of topics in genome-wide linkage and association studies, and next-generation sequencing studies. Additional topics such as network genetics will also be covered. Examples from recent and ongoing applications to complex traits will be used to illustrate methods and concepts.

P8120 Analysis of Categorical Data 3 points

Prerequisites: P6104 or P8130 or MPH Quantitative Foundations core course, and P6400 (Not open to MS/TM, PHDS, or SG tracks)

A comprehensive overview of methods of analysis for binary and other discrete response data, with applications to epidemiological and clinical studies. Topics discussed include the fourfold table, significance versus magnitude of association; estimation of relative risk; matching in design and analysis; interrater agreement; logistic regression analysis.

P8122 Statistical Methods for Causal Inference 3 points

Prerequisites: P8100 and P8110 or P8130 and P8131

This class will introduce students to both statistical theory and practice of causal inference. As theoretical frameworks, we will discuss potential outcomes, causal graphs, randomization and model-based inference, causal mediation, and suf ficient component causes. We will cover various methodological tools including randomized experiments, matching, inverse probability weighting, instrumental variable approaches, dynamic causal models, sensitivity analysis, statistical methods for mediation and interaction.

P8123 Analysis of Health Surveys 3 points

Pre-requisites: P8131 (or P8110) and P8104 (or P8107)

This is an applied statistical methods course. The course will introduce main techniques used in sampling practice,including simple random sampling, stratification, systematic sampling, cluster sampling, probability proportional to size sampling, and multistage sampling. Using national health surveys as examples, the course will introduce and demonstrate the application of statistical methods in analysing across-sectional surveys and repeated and longitudinal surveys, and conducting multiple imputation for missing data in large surveys. Other topics will include methods for variance estimation, weighting, post-stratification, and non-sampling errors. If time allows, new developments in small area estimation and in the era of data science will also be discussed.

P8124 Graphical Models for Complex Health Data 3 points

Pre-requisites: P8105 and P8109 or instructor's permission

This is a course at the intersection of statistics and machine learning, focusing on graphical models. In complex systems with many (perhaps hundreds or thousands) of variables, the formalism of graphical models can make representation more compact, inference more tractable, and intelligent data-driven decision-making more feasible. We will focus on representational schemes based on directed and undirected graphical models and discuss statistical inference, prediction, and structure learning. We will emphasize applications of graph-based methods in areas relevant to health: genetics, neuroscience, epidemiology, image analysis, clinical support systems, and more.

P8130 Biostatistical Methods I 3 points

Prerequisites: Students are required to have working knowledge of calculus and linear algebra

This course introduces basic applied descriptive and inferential statistics. The first part of the course includes elementary probability theory, an introduction to statistical distributions, principles of estimation and hypothesis testing, methods for comparison of discrete and continuous data including chi-squared test of independence, t-test, analysis of variance (ANOVA), and their non-parametric equivalents. The second part of the course focuses on linear models (regression) theory and their practical implementation.

P8131 Biostatistical Methods II 3 points

Prerequisites: P8130

Regression analysis is widely used in biomedical research. Non-continuous (e.g., binary or count-valued) responses, correlated observations, and censored data are frequently encountered in regression analysis. This course will introduce advanced statistical methods to address these practical problems. Topics include generalized linear models (GLM) for non-Gaussian response, mixed-effects models and generalized estimating equations (GEE) for correlated observations, and Cox proportional hazards models for survival data analysis. Examples are drawn from biomedical sciences.

P8133 Bayesian Analysis and Adaptive Designs for Clinical Trials 3 points

Prerequisites: P8104,P8109, and P8140

An introduction to sequential analysis as it applies to statistical problems in clinical trials, hypothesis testing, selection, and estimation. Emphasis is placed on a study of procedures, operating characteristics, and problems of implementation, rather than mathematical theory. Students obtain an overview of currently available sequential designs and the advantages and disadvantages they offer in comparison with classical designs.

P8134 Stochastic Approximation and Modern Dose-Finding 3 points

Prerequisites: P8104 and P8109 or their equivalents

Provides an in-depth study of statistical designs for dose-finding clinical trials of new drugs. This course is designed for advanced Master's, DrPH, and PhD students in biostatistics. The overall learning objective is to equip students with the techniques to construct, evaluate, and critique dose-finding designs. The course consists of two parts. The first is a review of modern dose-finding techniques with a focus on the continual reassessment method (CRM) and its clinical applications. The second part presents advanced topics on stochastic approximation and its related theory. Connections between the dose-finding methods (part 1) and the stochastic approximation (part 2) will be drawn. The practical implication of these connections is two-fold. First, the stochastic approximation will provide a versatile and mathematically rigorous framework for tailoring dose-finding designs to specific clinical situations. Second, the well-studied theory of stochastic approximation will be an effective analytical tool to approximate the theoretical properties of the CRM.

P8139 Statistical Genetics Modeling 3 points

Prerequisites: P6103 or P6104 or P8130, a working knowledge of calculus

Present to students statistical tools so that they can grasp the fundamentals of the design, conduct and analysis of genetic association studies. The course will thoroughly discuss current methods that are being used to map genes for common complex diseases. Great emphasis will be placed on candidate-gene and genome-wide association studies, but linkage methods will also be treated. Another key feature of this course will be a detailed treatment of the major findings of the Human Genome Project and HapMap Project.

P8140 Introduction to Randomized Clinical Trials 3 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations core course

Fundamental methods and concepts of the randomized clinical trial: protocol development, randomization, blindedness, patient recruitment, informed consent, compliance, sample size determination, crossovers, collaborative trials. Each student prepares and submits the protocol for a real or hypothetical clinical trial.

P8142 Clinical Trial Methodology 3 points

Prerequisites: P6104 or P8130

The main objective of this course is to provide students and investigators with a working knowledge of certain methodological issues that arise in designing a clinical trial in order to conduct complex study designs that yield valid and reliable results. With emphasis on several methodological and practical issues related to the design and analysis of clinical experiments, topics include: the design of small studies (Phase I and II studies), interim analyses and group sequential methods, survival studies, multiple outcome measures, surrogate outcomes, multicenter studies, issues in data analysis, and reporting and interpreting study results.

P8144 Pharmaceutical Statistics 3 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations core course. SAS knowledge recommended.

Drug development from compound discovery to marketing and commercialization registration is a lengthy and complex process in which statisticians play an important role from beginning to end. The main objective of this course is to provide students with working knowledge of the methodological and operational issues that arise in different stages of drug development that involve statistical contributions.

P8149 Human Population Genetics 3 points

Prerequisites: P8104 and the instructor's permission.

This course will cover all statistical aspects of population genetics. Upon completion of this course, the students will be able to model and do inference of underlying population genetic mechanisms and apply acquired knowledge about population genetics to the analyses of phenotypes.

P8157 Analysis of Longitudinal Data 3 points

Prerequisites: P8104, P8109, and P8130

The course will introduce students to statistical models and mthods for longitudinal data, i.e., repeatedly measured data over time or under different conditions. The topics will include design and sample size calculation, Hotelling's T2, multivariate analysis of variance, multivariate linear regression (generalized linear models), models for correlation, unbalanced repeated measurements, mixed effects models, EM algorithm, methods for non-normally distributed data, generalized estimating equations, generalized linear mixed models, and missing data.

P8158 Latent Variable and Structural Equation Modeling for Health Sciences 3 points

Prerequisites: P6104, P8130 or MPH Quantitative core course

This course is designed for those students (or any researchers) who want to gain a significant familiarity with a collection of statistical techniques that target the measurement of latent variables (i.e. variables that cannot be measured directly) as well as methods for estimating relationships among variables within causal systems. This course covers: both continuous and categorical latent variable measurement models (i.e. exploratory and confirmatory factor analysis, item response theory models, latent class and finite mixture models), as well as estimation of relationships in hypothesized causal systems using structural equation modeling. Data analysis examples will come from health science applications and practical implementation of all methods will be demonstrated using predominately the Mplus software, but also the R software.

P8160 Topics in Advanced Statistical Computing 3 points

Prerequisites: P8109, a basic understanding of Bayesian inference, working knowledge of a programming language

As statistical models become increasingly complex, it is often the case that exact or even asymptotic distributions of estimators and test statistics are intractable. With the continuing improvement of processor speed, computationally intensive methods have become invaluable tools for statisticians to use in practice. This course covers the basic modern statistical computing techniques and how they are applied in a variety of practical situations. Topics include numerical optimization, random number generation, simulation, Monte Carlo integration, permutation tests, jackknife and bootstrap procedures, Markov Chain Monte Carlo methods in Bayesian settings, and the EM algorithm.

P8170 Integrative Capstone Experience 2 points

Prerequisites: Biostatistics MPH students only

Required capstone course forall MPH students in Biostatistics. In this course, students will produce a written report that describes an analysis of relevant data using statistical techniques learned during the course of the MPH program.

P8180 Relational Databases and SQL Programming for Research and Data Science 3 points

Prerequisites: P6104, P8130 or MPH Quantitative Foundations core course, and the instructor's permission.

This class provides an overview of the specific techniques available to collect, store, retrieve, and control the quality of data in research projects. Students will be introduced to these concepts through a combination of lecture videos and a substantial hands-on component consisting of structured computer-based exercises. Spreadsheet and database technologies will be reviewed in detail to establish guidelines as to the appropriateness of their use to manage data in research.

P8182 Writing a Successful Grant Application 1 points

Prerequisites: Concurrent enrollment in the Columbia Summer Research Institute. Required for MS-POR students.

This seminar-style course will lead students through the process of writing an NIH-style grant application. By the end of the course, each student submits a research proposal outline following NIH guidelines for either an R01 or K (career development) award. The emphasis in this course is on the quality of the proposed research, taking into account feasibility, relevance, innovation, ethical foundation, and public health impact. As a culminating experience, students make oral presentations summarizing their research proposals to an invited panel of senior, experienced CUMC faculty, and receive feedback on their proposed research aims and approaches.

P8185 Capstone Consulting Seminar 1 point

Prerequisites: At least 15 points of required coursework in biostatistics. Biostatistics MS/MPH students only.

Required capstone course for the MS/PS, MS/SG, MS/TM and 2-year MPH students in Biostatistics. Provides experience in the art of consulting and in the proper application of statistical techniques to public health and medical research problems. Enables students to translate research objectives into statistical hypotheses, devise appropriate study designs, perform sample size calculations for studies employing simple random sampling, formulate and prepare written plans for statistical analysis for a research proposal, compose summaries of quantitative analyses, and communicate results clearly to public health colleagues. Based on seminars requiring active student participation.

P9104 Probability for Biostatisticians 3 points

Pre-requisites: P8109 and P8110, advanced calculus. Instructor's permission needed for MS students

The biostatistical field is changing with new directions emerging constantly. Doing research in these new directions, which often involve large data and complex designs, requires advanced probability and statistics tools. The purpose of this new course is to collect these important probability methods and present them in a way that is friendly to a biostatistics audience. This course is designed for PhD students in Biostatistics. Its primary objective is to help the students achieve a solid understanding of these probability methods and develop strong analytical skills that are necessary for conducting methodological research in modern biostatistics. At the completion of this course, the students will a) have a working knowledge in Law of Large Numbers, Central Limit Theorems, martingale theory, Brownian motions, weak convergence, empirical process, and Markov chain theory; b) be able to understand the biostatistical literature that involves such methods; c) be able to do proofs that call for such knowledge.

P9109 Theory of Statistical Inference I 3 points

Prerequisites: P8104, P8109. Instructor's permission needed for MS students

This course offers a general introduction to essential materials in advanced statistical theory for doctoral students in biostatistics. The course is designed to prepare doctoral students in biostatistics for their written theory qualifying exam. Students in this course will learn theory of estimation, confidence sets and hypothesis testing. Specific topics include a quick review of measure-theoretic probability theory, concepts of suf ficiency and completeness, unbiased estimation (UMVUE), least squares principle, likelihood estimation, a variety of estimators and their asymptotic properties, confidence sets, the Neyman-Pearson lemma and uniformly most powerful tests. If time permits, the likelihood ratio test, score test and Wald test, and sequential analysis will be covered.

P9110 Theory of Statistical Inference II 3 points

Prerequisites: P8104, P8109, and P9109. Instructor's permission needed for MS students

This course continues the introduction to mathematical statistics for doctoral students in biostatistics. Topics to be covered include: principles of decision theory, Bayesian estimation, Hypothesis testing, asymptotics, M-estimation, Wald tests, and score tests.

P9111 Asymptotic Statistics 3 points

Prerequisites: P8104, P8109, P9109, and P9110. Instructor's permission needed for MS students

The choice of topics will vary from year to year, but will typically include: empirical processes and M-estimation, bootstrap methods, empirical likelihood, contiguity, local asymptotic normality, counting process methods in survival analysis, semiparametric inference and efficiency.

P9120 Topics in Statistical Learning and Data Mining I 3 points

Prerequisites: Intended for Biostatistics PhD students and theoretically inclined MS students

Provide students a systematic training in key topics in modern supervised statistical learning and data mining. For the most part, the focus will remain on a theoretically sound understanding of the methods (learning algorithms) and their applications in complex data analysis, rather than proving technical theorems. Applications of the statistical learning and data mining tools in biomedical and health sciences will be highlighted.

P9130 Advanced Biostatistical Methods I 3 points

Prerequisites: Advanced calculus, linear algebra, basic probability, statistical inference and instructor's permission for MS students

The course will provide a solid foundation of the theory behind linear models and generalized linear models. More emphasis will be placed on concepts and theory with mathematical rigor. Topics covered including linear regression models, logistic regression models, generalized linear regression models and methods for the analysis contingency tables.

P9160 Master's Essay in Biostatistics: Clinical Research Methods 3 points

Prerequisites: At least 15 points of required coursework. MS/CRM students only.

Students produce a Master's essay in the form of a research article of publishable quality, supervised by faculty members from Biostatistics and from the student's own clinical field.

P9165 Master's Essay in Biostatistics: Patient Oriented Research 0 points

Prerequisites: At least 15 points of required coursework. MS/POR students only

Students produce a Master's essay in the form of an NIH-style grant application, supervised by a project sponsor from Biostatistics and a mentor from the student's own clinical field. A formal presentation to the POR advisory board is required for successful completion of the course.

P9185 Statistical Practices and Research for Interdisciplinary Sciences (SPRIS) 3 points

Prerequisites: DrPH and PhD Biostatistics only

Required course for the DrPH and PhD students in biostatistics. Provides experience in the art of consulting and in the proper application of statistical techniques to public health and medical research problems. Enables students to translate research objectives into statistical hypotheses, devise appropriate study designs, perform sample size calculations for studies employing simple random sampling, formulate and prepare written plans for statistical analysis for a research proposal, compose summaries of quantitative analyses, and communicate results clearly to public health colleagues. Based on seminars requiring active student participation.

P9186 Statistical Practices and Research for Interdisciplinary Science (SPRIS) II 1.5 points

Prerequisites: P9185

Students will apply the concepts and methods introduced in Statistical Practices and Research for Interdisciplinary Science (SPRIS) I to a real research setting. Each student will be paired with a Biostatistics faculty member. The student will participate in one of the mentor's collaborative projects to learn how to be an effective member of an interdisciplinary team. Student experience will vary depending on the assigned faculty member, but all students will gain exposure to preparing collaborative grant applications, designing research studies, analyzing real data, interpreting and presenting results, and writing manuscripts. Mentors will help to develop the student's data intuition skills, ability to ask good research questions, and leadership qualities. Where necessary, students may replicate projects already completed by the faculty mentor to gain experience.

P8190/P9190 Tutorials in Biostatistics 1 to 6 points

For appropriately qualified students wishing to enrich their programs by undertaking literature reviews, special studies, or small group instruction in topics not covered in formal courses. Hours to be arranged.

89260 Building Interdisciplinary Research Models 2 points

Interdisciplinary research is an approach to advancing scientific knowledge requiring mastery of specific competencies. This seminar will introduce the students to competencies in interdisciplinary research through a combination of readings and lectures in each necessary aspect, chosen from fields essential to successful interdisciplinary research. This course will assist learners to understand why and how different professional disciplines, each representing a body of scientific knowledge, must work together to generate and disseminate knowledge. Learners will develop a set of skills specific to be an effective member and leader of an interdisciplinary research team, and will become familiar with the advantages of team science.

P8449 Optimization for Interventions 1.5 points

Prerequisites: Quantitative portion of the MPH Core curriculum or the equivalent.

Students will learn how to apply engineering-inspired concept of optimization to the study of behavioral, biobehavioral and biomedical interventions across public health fields. The course will be grounded in the multiphase optimization strategy (MOST) framework. Under the optimization phase of the MOST framework, the course will introduce experimental designs with an emphasis on sequential, multiple assignment randomized trial (SMART) which is a way to develop high-quality adaptive interventions. Micro-randomized Trails which are referred to as MRTs, a way to develop mHealth Just-in-Time Adaptive Interventions (JITAIs) also will be covered.

Opportunities for Student Involvement

Colloquia

During the Fall and Spring semesters, the Department of Biostatistics holds seminars on a wide variety of topics which are of interest to both students and faculty. The speakers are occasionally departmental faculty members themselves but very often are invited guests who spend the day of their seminar discussing their research with Biostatistics faculty and students. While all students are strongly urged to attend, doctoral student attendance is mandatory.

Social Events

The Department regularly hosts various social events students are encouraged to attend to get to know their faculty, staff, and fellow students. Some include:

- T-Time: A regular social event hosted in the building on Wednesdays with snacks and drinks
- Doctoral Social: Happy hours at local businesses where doctoral students can mingle in a casual setting with their faculty and fellow students
- Holiday Party: A party to celebrate the successful Fall semester with drinks and food
- Bake-Off: a yearly event where everyone in the department is invited to enter their baked goods to compete for prizes or come to judge the various tasty treats
- Holiday Themed Events: We host a variety of events around holidays such as Easter, Lunar New Year, Halloween and more.

You can also suggest potential events or ideas for socials to the programs team!

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

BIOSTATISTICS

2024-2025 STUDENT HANDBOOK

revised August 2024