



## DEPARTMENT OF BIostatISTICS AND COMPUTATIONAL BIOLOGY

### About the Department

The Department of Biostatistics and Computational Biology at the University of Rochester has a long record of excellence in methodologic and collaborative research and in the education of professionals in and users of statistics. The Department currently has 17 faculty who are eligible to advise students with research interests spanning traditional biostatistics, stochastic modeling, bioinformatics, and computational biology.

The Department offers masters and doctoral degrees and has an active postdoctoral program supported by research and training grants. We have strong collaborative relationships with Anesthesiology, Biomedical Genetics, the Cancer Center, Cardiology, Environmental Medicine, Microbiology and Immunology, Neurology, Orthopedics, Pediatrics, Psychiatry, Public Health Sciences and many other departments, centers, and units throughout the School. We are also an important contributor to the Clinical and Translational Science Institute of the University of Rochester.

### About Our Faculty

Faculty of the Department have played major roles in important breakthroughs in medical research at Rochester. Examples include the better understanding of the role of calcium channel blockers in treating patients who have had a heart attack, demonstration of both the clinical effectiveness and the cost effectiveness of implantable defibrillators in reducing mortality among certain heart disease patients, demonstration of the effectiveness of deprenyl in slowing onset of disability in early Parkinson's disease, and of surfactant therapy for respiratory distress syndrome in premature infants, the benefits and risks of early surgical therapy for drug-resistant temporal lobe epilepsy, and an ongoing epidemiologic study of the assessment of the effects of low levels of dietary mercury intake on childhood development.

Methodologic and collaborative interests of our graduate research advisors are shown below.

#### Professors

- Robert Strawderman, ScD, Harvard University (Department Chair).* Survival analysis; statistical inference for point process data (e.g., recurrent events); statistical methods for risk and outcome prediction in medicine, epidemiology and public health; statistical and computational methods for high dimensional data; statistical methods for evaluating the cost and quality of health care; asymptotics (theory and approximation); statistical computing.
- Michael McDermott, PhD, University of Rochester (Director, Statistics PhD Program).* Order-restricted inference; receiver operating characteristic (ROC) curves and surfaces; methods for combining p-values; meta-analysis; missing data problems; clinical trials methodology; applications in neurological disease.
- David Oakes, PhD, London University.* Survival analysis, including multivariate survival data and frailty models; semiparametric inference; clinical trials; applications in environmental medicine and neurological disease.
- Derick Peterson, PhD, University of California, Berkeley.* Construction of prognostic gene expression profiles from high-dimensional microarray data; model selection; nonparametric estimation and inference; data-driven smoothing methods; analysis of censored survival data.

#### Associate Professors

- Anthony Almudevar, PhD, University of Toronto.* Genetics and bioinformatics, especially in the area of graphical modelling, with applications to cellular networks and population biology; optimization and control theory, particularly in the area of Markov decision processes; analysis of technological motion data, particularly data collected from home monitoring systems.
- Christopher Beck, PhD, University of Rochester.* Inference concerning receiver operating characteristic (ROC) curves and surfaces; change-point problems; clinical trials; applications in orthopedics and neurological disease.
- Changyong Feng, PhD, University of Rochester.* Multivariate survival analysis; empirical processes theory; longitudinal data analysis; statistical methods in epidemiology and clinical trials.

*Brent Johnson, PhD, North Carolina State University.* Semi-parametric methods for missing data problems with specific applications to causal inference, survival and longitudinal data; dynamic treatment regimes in therapeutic HIV and AIDS studies, infusion trials, neurological and behavioral disorders; pattern analysis in HIV prevention; statistical methods for epidemiology.

*Tanzy Love, PhD, Iowa State University.* Clustering and latent variable models; mixed membership models and model choice; normalization and preprocessing issues relating to gene expression and proteomics data; Bayesian models for QTLs and growth curves; hierarchical Bayesian models for gene expression data; scalable parallel model-based clustering.

*Xing Qiu, PhD, University of Rochester.* Computational biology, particularly microarray data analysis, intergene correlation structure, normalization procedures, and stability of various gene selection procedures; nonparametric inference; theory and applications of stochastic differential equations, especially in modeling infectious disease; information geometry and its application to hypothesis testing procedures based on correlation/covariance.

*Sally Thurston, PhD, Harvard University.* Modeling multiple outcomes; methods of correcting for measurement error bias; exposure assessment; non-parametric smoothing; Bayesian inference; informative prior specification; latent variable models; applications in environmental health.

*Tongtong Wu, PhD, University of California, Los Angeles (Director, Master's Programs).* High-dimensional data analysis; survival analysis; machine learning; computational statistics; computational biology and statistical genetics; longitudinal data analysis.

### **Assistant Professors**

*Ashkan Ertefaie, PhD, McGill University.* Causal inference; dynamic treatment regimes; sequential multiple assignment randomized trials; comparative effectiveness studies using electronic health records; instrumental variable analyses; high-dimensional data analysis; post selection inference; survival analysis.

*Matthew McCall, PhD, Johns Hopkins University.* Statistical genomics; systems biology; bioinformatics; methods to estimate gene regulatory networks involved in cancer; within-subject genomic heterogeneity; methods for preprocessing and analysis of genomic data; effect of cellular composition on tissue-level gene expression.

*Andrew McDavid, PhD, University of Washington.* Single cell gene expression; zero-inflated models; high dimensional estimation; convex optimization; applied Bayesian modeling; immunology; bioinformatics and computing.

*Michael Sohn, PhD, University of Arizona.* Mediation analysis; causal inference; high-dimensional data analysis; statistical machine learning; compositional data analysis; multivariate data analysis; statistical genomics and metagenomics.

*Zhengwu Zhang, PhD, Florida State University.* Statistical image analysis; statistical shape analysis; functional data analysis; Bayesian statistics; network analysis; computational neuroscience.

## **Degree Programs**

### **PHD DEGREE**

#### **Program for the Degree of Doctor of Philosophy in Statistics (Traditional)**

The department administers the doctoral program in statistics. The department interprets the term “statistics” very broadly. The program permits specialization in probability, statistical theory and analysis, biostatistics, and interdisciplinary areas of application. Students have opportunities for supervised teaching and supervised consulting experience, requiring approximately 12 to 15 hours of effort per week.

A candidate for admission to the PhD program should have a strong background in mathematics, including advanced calculus or mathematical analysis, a course in linear and/or matrix algebra, and a year of probability and mathematical statistics. A course in statistical methods is also recommended; however, promising students may make up deficiencies after matriculation. While some background in biology may be helpful for pursuing certain avenues of research, it is not required for admission to the program.

All MA/PhD students take a comprehensive (basic) examination at the beginning of the second year. PhD students take another written (advanced) examination at the beginning of the third year. Both examinations cover material in the areas of probability, inference, and data analysis.

After beginning research on a dissertation topic, PhD students take an oral qualifying examination, consisting largely of a presentation of a thesis proposal to a faculty committee, the student's Thesis Committee. Upon completion of the dissertation, doctoral candidates present their work at a public lecture followed by an oral defense of the dissertation before the Thesis Committee.

Prior to completing degrees, most students have some publications underway, including some work related to their dissertation research, possibly other methodological work done in collaboration with other members of the faculty, and often some applied papers with scientific researchers in other fields. In general, the PhD program requires a minimum of four years of study, with five years of study being more common.

### Course Requirements

Course work in statistics is concentrated in three areas – probability, inference, and data analysis. Beginning students should expect to spend all of their first year, most of their second year, and some of their third year taking formal courses. The balance of time is spent on reading and research. Students entering with advanced training in statistics may transfer credits at the discretion of their advisors and in accordance with University policy.

A typical program for an entering student without previous advanced training is as follows:

#### *Year 1 Fall*

Probability Theory  
Statistical Inference  
Introduction to Statistical Computing  
Seminar in Statistical Literature  
Ethics in Research  
Supervised Teaching

#### *Year 1 Spring*

Bayesian Inference  
Linear Models  
Categorical Data Analysis  
Seminar in Statistical Literature  
Supervised Teaching

#### *Year 2 Fall*

Stochastic Processes  
Generalized Linear Models  
Elective  
Seminar in Statistical Literature  
Supervised Teaching

#### *Year 2 Spring*

Large Sample Theory  
Analysis of Longitudinal and Dependent Data  
Elective  
Seminar in Statistical Literature  
Reading Course at the PhD Level

#### *Year 3+*

Mostly reading and research, with some 400- and 500-level courses

Recent examples of advanced elective courses include:

Functional Data Analysis	Frailty Models in Survival Analysis
Causal Inference and its Applications	Missing Data
Advanced Bayesian Inference with an Emphasis on Computation	The Bootstrap, the Jackknife, and Resampling Methods
Smoothing Methods	Time Series
ROC Curve Analysis	Model Selection and Validation

### Program for the Degree of Doctor of Philosophy in Statistics with Concentration in Bioinformatics and Computational Biology

The Bioinformatics and Computational Biology (BCB) concentration is designed to educate the next generation of biostatisticians with the knowledge required to address critical scientific and public health questions, and in particular, equip them with the skills necessary to both develop and use quantitative and computational methodologies and tools to manage, analyze, and integrate massive amounts of complex biomedical data. Students learn core statistical methods and obtain training in data analysis methodologies and computational skills and techniques necessary for handling “Big Data” in the biomedical and public health sciences. In addition to this training in core methods, the program also places great emphasis on cross-training to prepare students to work as part of interdisciplinary teams that require expertise in statistical data science: 1) training students with quantitative/computational science backgrounds to enhance their understanding of biological questions and biological interpretation; and 2) training students with biomedical science backgrounds to proficiently use bioinformatics and computational methods and tools to address scientific questions.

Entering PhD students need undergraduate preparation in mathematics, including advanced calculus or mathematical analysis, a course in linear and/or matrix algebra, and a year of probability and mathematical statistics. Basic courses in computer science and/or biology are also required. A course in statistical methods is also recommended; however, promising students may make up deficiencies after matriculation.

## Course Requirements

Formal course and examination requirements for students in the BCB concentration are essentially the same as those for students in the traditional program, with the main differences being in some required and elective courses related to bioinformatics and computational biology. Beginning students should expect to spend all of their first year, most of their second year, and some of their third year taking formal courses. The balance of time is spent on reading and research. Students entering with advanced training in statistics, bioinformatics, or computational biology may transfer credits at the discretion of their advisors and in accordance with University policy.

A typical program for an entering student without previous advanced training is as follows:

### *Year 1 Fall*

Probability Theory  
Statistical Inference  
Introduction to Statistical Computing  
Seminar in Statistical Literature  
Ethics in Research  
Supervised Teaching

### *Year 1 Spring*

Bayesian Inference  
Linear Models  
Categorical Data Analysis  
Seminar in Statistical Literature  
Supervised Teaching

### *Year 2 Fall*

Stochastic Processes  
High Dimensional Data Analysis  
Generalized Linear Models  
Seminar in Statistical Literature  
Supervised Teaching

### *Year 2 Spring*

Large Sample Theory  
Genomic Data Analysis  
Analysis of Longitudinal and Dependent Data  
Seminar in Statistical Literature  
Reading Course at the PhD Level

### *Year 3+*

Mostly reading and research, with some 400- and 500-level courses

## MASTER'S DEGREES

### Master of Arts in Statistics

The requirements for entry into the MA program in statistics are the same as those for entry into the PhD program. The MA degree requires satisfactory completion of at least 32 credits and a final examination (the basic examination requirement above or an oral examination); no thesis is required. Of the 32 credits, at least 24 must be in departmental courses primarily at the 400-level or above. Appropriate substitutions may be made as long as the spirit (distribution and level) of the requirements is met. A balanced program is worked out with the student's advisor. A typical program for the MA is the same as that in the first three semesters of the PhD program.

Students in the PhD program receive an MA degree upon satisfactory completion of the requirements for this degree (typically during the second year of graduate study).

### Master of Science in Medical Statistics

The MS program in medical statistics is primarily intended for students who wish to follow careers in health-related professions such as those in the pharmaceutical industry and biomedical or clinical research organizations. For entry into the program, three semesters of calculus, a course in linear and/or matrix algebra, a course in probability, a course in mathematical statistics, and a course in applied statistics are required.

The MS program in medical statistics consists of one core year (two semesters) of coursework as well as an internship/applied project, which is normally taken in the summer after the core program. The degree requires 32 credit hours consisting of the courses listed below; substitutions may be made with approval of the faculty program advisor. A comprehensive oral examination to determine the student's qualifications for the MS degree will be administered upon completion of coursework and a written report of the internship/applied project.

A typical program for an entering student without previous advanced training is as follows:

### *Fall*

Statistical Inference  
Statistical Computing  
Applied Linear Regression

### *Spring*

Design of Clinical Trials  
Categorical Data Analysis  
Elective (e.g., Bayesian Inference, Linear Models)

### *Summer*

Internship/Applied Project

## About Our Students

Students have opportunities for supervised teaching and statistical consulting experience. The department gives individual attention to each student through intensive advising, extensive small seminars, and research collaboration. Prior to completing their degrees, most doctoral students have several publications underway based on research done in collaboration with faculty members in biostatistics/statistics and in various medical departments.

PhD program graduates have found employment at Georgetown University, State University of New York at Buffalo, Carnegie Mellon University, Case Western Reserve University, Harvard University, Emory University, University of Florida, University of Illinois – Chicago, Johns Hopkins University, Lehigh University, Medical University of South Carolina, University of Rochester, Rochester Institute of Technology, University of Pittsburgh, Southern Methodist University, other state universities, numerous companies such as Google, Merck, Novartis, AbbVie, DuPont, and Bell Laboratories, and governmental agencies. MA and MS graduates are in various academic programs and in industrial, government, research, and consulting positions.

Recent PhD graduates with their advisors and thesis titles:

2018	Majumder, Madhurima	Dr. M. P. McDermott	Conditional Tests for Multivariate One-sided Hypotheses with Missing Data
2018	Zhang, Yun	Dr. X. Qiu and Dr. J. Thakar	Novel Statistical Methods for Gene Set Enrichment Analysis with Empirical Memberships for Overlapping Genes
2017	Chen, Chongshu	Drs. S.W. Thurston and O. Hyrien	Finite Mixtures of Nonlinear Mixed-Effects Models for Longitudinal Data
2017	Ciminelli, Joseph	Dr. T. Love	Mixed-Membership and Spatial Models for Social Network Data
2017	LaLonde, Amy	Dr. T. Love	Bayesian Model-Based Clustering Methods: Procedures for Data with Unknown Numbers of Clusters
2017	Grzesik, Katherine	Dr. D. Peterson	Local Cross-Validated Smoothing Parameter Estimation for Linear Smoothers
2017	Liu, Chang	Dr. S.W. Thurston	Bayesian Semiparametric Measurement Error Models: Estimation, Variable Selection and Fast Algorithms
2016	Hebert, Donald	Dr. M. P. McDermott	Global Tests for Multiple Outcomes in Randomized Trials
2016	Singh, Kyra	Dr. T. Love	Variable Selection Methods for Model-Based Clustering: Procedures for Functional Data and Bayesian Inference
2016	Yu, Ziji	Dr. G. S. Mudholkar	Theory and Application of the Mode Centric M-Gaussian Distribution
2016	Chowdhry, Amit	Dr. M. P. McDermott	Missing Data in Meta-Analysis
2015	Chen, Tian	Dr. X. M. Tu	A New Class of Functional Response Models for Robust Regression Analysis
2015	Xia, Changming	Drs. H. Liang and S. W. Thurston	Generalized Semiparametric Linear Mixed-Effects Models
2015	Tran, Thanh Van	Dr. A. Almudevar	Threshold Boolean Network Inference and Experimental Design
2015	Ma, Fei	Drs. O. Hyrien and X. M. Tu	Composite Likelihood Inference for Multivariate Finite Mixture Models and Application to Flow Cytometry
2014	Han, Yu	Drs. C. Feng and X. M. Tu	New Semiparametric Methods for Clustered Time-to-Event Data
2014	Morrisette, Jason	Dr. M. P. McDermott	Order Restricted Analysis of Covariance with Interactions
2014	Zhang, Xiao	Drs. M. P. McDermott and G. S. Mudholkar	Hypothesis Testing Problems Involving Order Restricted Parameters
2014	Yang, Hui	Dr. H. Liang	Model Selection and Model Averaging for Longitudinal Data with Applications in Personalized Medicine
2014	Evans, Katie	Drs. T. Love and S. W. Thurston	Extensions to Model-Based Clustering for Mixed-Type Data: A New Model Framework, Variable Selection, and Outlier Identification
2013	Yu, Yao	Dr. H. Liang	Single-index Model with Application in Estimation of ODE and Gene Regulatory Network
2013	Wu, Pan	Dr. X. M. Tu	A New Class of Structural Functional Response Models for Causal Inference and Mediation Analysis
2013	Chen, Juofan	Dr. H. Wu	State Space Models and Differential Equations for Dynamic Gene Regulatory Network Identification
2013	Chen, Zhen	Drs. D. Oakes and O. Hyrien	A Flexible Copula Model for Bivariate Survival Data
2012	Awadalla, Saria	Drs. G. S. Mudholkar and M. P. McDermott	Some Contributions to the Theory and Applications of R-Symmetry
2012	Liu, Xiang	Dr. H. Liang	Penalized Variable Selection for Semiparametric Regression Models

2011	Gunzler, Douglas	Dr. X. M. Tu	A Class of Distribution-free Models for Longitudinal Mediation Analysis
2011	Lu, Tao	Dr. H. Wu	ODE Based Statistical Models for Dynamic Gene Regulatory Network Identification
2011	LaCombe, Jason	Dr. A. Almudevar	Non-Informative Priors for Structural Inference in Bayesian Networks
2011	Lynch, Miranda	Dr. S. W. Thurston	Estimation, Model Checking and Diagnostics, and Identifiability in Finite Mixture Models for Point Mass Data: Methods in a Bayesian Framework
2010	Zhang, Hui	Dr. X. M. Tu	Distribution-free Models for Latent Population Mixtures
2009	Stokes-Riner, Abbie	Dr. S. W. Thurston	Residual Diagnostic Methods for Bayesian Structural Equation Models
2009	Yu, Qin	Dr. X. M. Tu	Distribution-free Models for Longitudinal Count Data
2009	Su, Haiyan	Dr. H. Liang	Empirical Likelihood-Based Inference for Multiple Regression and Treatment Comparison
2008	Pearson, Alexander	Dr. D. Peterson	Subset Selection for High-Dimensional Data, with Applications to Gene Array Data
2008	Chen, Linlin	Drs. A. Yakovlev and A. Almudevar	The Correlation Structure of Microarray Data and Its Statistical Implications
2008	Ma, Yan	Dr. X. M. Tu	Inference for Instrument Reliability and Rater Agreement within a Multi-rater and Longitudinal Data Setting
2008	He, Hua	Drs. M. P. McDermott and L.-S. Huang	Correcting Verification Bias in the Assessment of the Accuracy of Diagnostic Tests
2008	Georger, Lesley	Dr. G. S. Mudholkar	Some Skew Models for Quantal Response Analysis

Recent MS graduates with their advisors and project titles:

2018	Catalfamo, Kayla	Dr. M. P. McDermott	Evaluation of Baseline Predictors of One-Year Change in Outcomes in Myotonic Dystrophy Type 1 in Preparation for Clinical Trials
2017	Maletz-Novick, Kristina	Dr. D. Peterson	Identification of Patients at High Risk for Sub-Optimal Bladder Filling During Radiation Therapy for Prostate Cancer
2017	Lee, Charles	Dr. T. Love	Hospital Charges: The Elephant in the Doctor's Office
2016	Tran, Lam	Dr. C. Feng	Structured Deep Learning on Medical Image Segmentation
2016	Kang, Hongyi	Dr. X. Qiu	The Change in Neuronal Dendritic Volume from Wakefulness to Sleep
2016	Chapman, Benjamin	Dr. A. Almudevar	A False Discovery Rate Upper Bound For Two-Stage Testing
2014	Bandyopadhyay, Sanjukta	Drs. H. Yang and H. Miao	Identifying Factors Affecting Treatment Success in Adult Drug Treatment Court Clients
2013	Boselli, Danielle	Dr. D. Peterson	Assessment of Left and Right Ventricular Volume Reduction to Identify Clinical Responders to CRT-D Among Non-LBBB Patients: A MADIT-CRT Sub-study
2011	Xia, Yinglin	Dr. X. M. Tu	Modeling HIV Risk Reduction Intervention: The Zero-inflated Negative Binomial and Zero-inflated Poisson Regression Models
2010	Miao, Hongyu	Dr. H. Liang	Understanding B Cell Kinetics in Humans via Heavy Water Labeling Using Nonlinear Mixed Effects Models and Stochastic Approximation EM Algorithms
2009	Rollins, Nicholas	Dr. X. M. Tu	Predictors of Depression in Pregnant Women and New Mothers Six Weeks Post-Partum
2009	Berry, Andrea	Dr. C. Beck	Evaluation of Influenza Vaccine Effectiveness Among the Rochester Population: 2008-2009 Season
2008	Osuch, Neil	Dr. M. P. McDermott	Cognitive Outcomes in a Randomized Trial of Early Surgical Intervention versus Continued Optimal Pharmacotherapy in Mesial Temporal Lobe Epilepsy
2008	Miller, Wesley	Dr. L.-S. Huang	Factors Influencing Primary and Secondary Educational Attainment of Children in the Seychelles Republic
2008	Eldar-Lissai, Adi	Dr. O. Hyrien	Identifying Factors Influencing Oncologists' Decision to Use Primary Prophylaxis G-CSF

## Applying to Our Programs

Prospective students apply for admission online at

[www.urmc.rochester.edu/education/graduate/prospective-students/](http://www.urmc.rochester.edu/education/graduate/prospective-students/)

### Application Requirements

- Online application
- Statement of purpose
- 3 letters of recommendation
- Transcripts
- Official test scores (GRE, TOEFL)
- Payment of the \$60 application fee

### Application Timeline for PhD Program

November 15: Deadline to submit application to receive automatic waiver of the \$60 application fee  
December 1: Complete application due (\$60 application fee applies)  
February: Interviews  
February - March: Admission offers sent  
April 15: Deadline to reply to admission offer

### Application Timeline for Master's Programs

May 1: Complete application due  
May-June: Offer of admission notices mailed  
July 1: Responses due for offers of admission

### For more information

Karin Gasaway  
Graduate Program Administrator  
585-275-6684  
URBiostat@urmc.rochester.edu

Department of Biostatistics and Computational Biology  
University of Rochester School of Medicine and Dentistry  
265 Crittenden Boulevard, CU 420630  
Rochester, New York 14642-0630

Fellowships and scholarships are awarded for full-time graduate study at the PhD level. All full-time PhD students receive a full tuition scholarship, single health insurance, and a 10-month stipend (\$24,639 in academic year 2019-20), with summer support possible depending on availability. PhD students are expected to assist with teaching duties and consulting projects. A training grant in Environmental Health Biostatistics funded by the NIEHS supports three PhD students each year; trainees must be US citizens or permanent residents.

Masters students generally rely on need-based financial aid. The School's Financial Aid Office web site ([www.urmc.rochester.edu/education/financial-aid/](http://www.urmc.rochester.edu/education/financial-aid/)) provides information about the aid programs that students use most frequently.

[www.urmc.rochester.edu/biostat](http://www.urmc.rochester.edu/biostat)

<https://www.urmc.rochester.edu/education/graduate/phd/statistics.aspx>