



INDIANA UNIVERSITY

University Graduate School
2008-2009
Academic Bulletin

Biostatistics

School of Medicine
Indianapolis

and

Purdue School of Science
Indianapolis

Program Directors

Professor Benzion Boukai*

Assistant Professor Jaroslaw Harezlak*

Program E-mail

grad-program@math.iupui.edu

Program URL

<http://www.math.iupui.edu/program/phd/biostat/>

Graduate Faculty

(An asterisk [*] denotes membership in the University Graduate School faculty with the endorsement to direct doctoral dissertations.)

Professors

Benzion Boukai*, Sujuan Gao*, Siu L. Hui*, Barry P. Katz*, Constantin Yiannoutsos*

Associate Professors

Lang Li*, Xiaochun Li, Patrick O. Monahan, Hanxiang Peng, Susan M. Perkins, Chandan K. Saha, Jyotirmoy Sarkar*, Wanzhu Tu

Assistant Professors

Xueya Cai, Samiran Ghosh, Jaroslaw Harezlak*, Fang Li, Yunlong Liu, Ryan Martin, Changyu Shen, Menggang Yu, Zhangsheng Yu

Degree Offered

Doctor of Philosophy

Special Departmental Requirements

(See also general University Graduate School requirements.)

Graduate Office
Union Building 207
Indiana University-Purdue
University Indianapolis
Indianapolis, IN 46202
(317) 274-1577
Contact: gradoff@iupui.edu

Admission Requirements

The application deadline for the program is January 15 of each year for the following fall. The program is designed for individuals with strong quantitative and analytical skills and a strong interest in biological, medical and/or health related sciences. A bachelor's degree or equivalent is required. Students seeking admission must apply online. Applications must include a complete application form, three letters of recommendation, official undergraduate transcripts, personal statement, resume or CV, and the Graduate Record Examination (GRE) general test. The TOEFL is also required if the applicant's native language is not English.

Course Requirements

A minimum of 90 credit hours are required for the degree. The 90 credit hours will consist of the following:

Core Courses (33 hours): A common core of 33 credit hours of course work will be required of all students who begin the program after the completion of a bachelor's degree.

STAT 51200	Applied Regression Analysis
BIOS S515	Biostatistics Practicum
STAT 51900	Introduction to Probability +
STAT 52500	Intermediate Statistical Methods +
BIOS S527	Introduction to Clinical Trials
STAT 52800	Mathematical Statistics I +
STAT 53600	Introduction to Survival Analysis +
BIOS S546	Applied Longitudinal Data Analysis +

Any three of the following:

STAT 61900	Probability Theory ^
BIOS S621	Advanced Statistical Computing ^
STAT 62800	Advanced Statistical Inference ^
BIOS S636	Advanced Survival Analysis ^
BIOS S646	Advanced Longitudinal Data Analysis ^

(+ indicates the program's core courses; ^ indicates course is still subject to approval by The University Graduate School)

Elective Courses (12 hours): All students must take 12 credit hours of elective statistics/biostatistics courses. At most six credit hours of the electives may be taken from 500-level courses.

Minor Area (9 to 15 hours): All students must complete a minor in an area related to any of health and life sciences. The minor may be obtained in areas such as pharmacology and toxicology, epidemiology, genetics, biology, physiology bioinformatics, public health and health economics, among many others and it must be approved by the student's advisor or

graduate committee. The minor must contain a minimum of three graduate level courses (9 cr.) in the chosen area and it must comply with the minor requirements of the respective department/unit.

Dissertation (27 to 33 hours): The remaining hours to total 90 will be guided research dissertation hours.

Qualifying Examination

Students must pass an initial qualifying examination on the five core courses. The qualifying examination is a written examination offered once a year during a two-day Qualifier Exam Session the week before classes start in August and is administered in two sections – Theoretical Biostatistics and Applied Biostatistics. The preparation and the administration of the qualifying examination are overseen by the Graduate Examination Committee. Students are expected to have completed and passed both sections of the qualifying examination on or before their qualifier deadline.

Deadline for full-time students:

The deadline for passing the qualifying examination for full-time students who enter the program with a master's degree or equivalent is August at the end of their second year; the deadline for full-time students who enter the program without a master's degree is August at the end of their third year.

Deadline for part-time students:

The deadline for passing the qualifying examination for part-time students who enter the program with a master's degree or equivalent is August at the end of their third year; the deadline for part-time students who enter the program without a master's degree is August at the end of their fourth year. If students do not pass both sections of the examination by their qualifier deadline, they will have their privilege to continue in the program terminated.

A student will have at most two attempts to pass the examination. The first attempt must include the entire examination, i.e. both the Theoretical and Applied sections. If one or both sections are not passed on the first attempt, then a second attempt on or before the deadline is allowed. During the final attempt, the student may only sit for the section(s) not passed in the first attempt.

A student's first attempt at the qualifying examination will result in one of the following three outcomes:

1. **Pass Both Sections:** The student has demonstrated fundamental understanding of the core material and the examination committee believes he/she will be successful in completing the Ph.D. program.
2. **Pass One Section:** The student has demonstrated fundamental understanding of one section, but lacks adequate understanding of the other section. The student must sit for the section not passed at a future examination session.
3. **Fail:** The student has failed to demonstrate an adequate understanding of the material from the core courses and thus fails the examination. The student must sit for both sections at a future examination session.

A student's second and final attempt at the qualifying examination will result in one of the following two outcomes:

1. **Pass:** The student has demonstrated fundamental understanding of the core material and the examination committee believes he/she will be successful in completing the Ph.D. program.
2. **Fail:** The student has failed to demonstrate an adequate understanding of the material from the core courses and thus fails the examination, with privilege to continue in the program terminated.

Minor Area

In addition to the 46 credit hours of formal statistics/biostatistics coursework, the student must complete a minor (9 – 15 credit hours) in an area related to any of the health and life sciences disciplines. The minor may be obtained in areas such as pharmacology and toxicology, epidemiology, genetics, biology, physiology, bioinformatics, public health and health economics, among many others, and it must be approved by the student's advisory committee. The minor must contain a minimum of three graduate level courses (nine credits) in the chosen area and it must comply with the minor requirements of the respective department/unit.

Preliminary Examination

A student becomes eligible to take the preliminary examination after passing the qualifying examination. The student must prepare and pass a preliminary examination, which consists of an oral presentation on an advanced research topic suggested by the student to an appointed committee of at least four faculty members, including the student's advisor and at least one member from the student's minor area. The committee may consist of the same members as the student's original advisory committee, but is not required to. This committee will serve as the research committee for the student, and must be approved by the dean. Prior to the examination, the student must provide the committee with a paper (10 – 15 pages) outlining the topic to be covered, clearly indicating the scope and depth of the planned research along with relevant references. In the examination, the student is expected to display an in-depth understanding of the chosen subject matter. The committee may ask the student questions which normally will be directed to the subject matter of the research but may, by natural extension, also cover any other relevant topic.

Admission to Candidacy

Following the passing of the preliminary examination and the completion of all required coursework, the student's advisory committee will nominate the student to candidacy. Upon approval of the dean of the The University Graduate School, the student will be admitted to candidacy.

Final Examination

Oral, primarily a defense of the dissertation.

Normal Progress and Termination

Once students begin research, they must maintain normal progress toward their degree objective to ensure continued

financial support and/or active status. If, in the opinion of the research committee, satisfactory research progress is not being made or if the GPA continues to be below 3.0, a meeting of the student's research committee may be convened. This meeting will include a brief presentation by the student on the work accomplished up to that point, and/or a discussion concerning the problems which have hampered progress. If the consensus of the committee is that the student needs to show improvement, he/she will have 60 days to demonstrate a change in performance. At the end of this time, financial support may be discontinued, if applicable.

If a student finds it necessary to withdraw from the graduate program, then he/she should provide as much notice as possible. In the case of teaching or research assistants, students are expected to complete the semester once it has begun. Similarly, the program will provide a student with as much advance notice as possible if the student is dropped from the program for reasons of poor performance.

Courses Offered

(^ indicates course is still subject to approval by The University Graduate School)

500 Level

STAT 51200 Applied Regression (3 cr.) P: STAT 51100. Inference in simple and multiple linear regression, residual analysis, transformations, polynomial regression, model building with real data, nonlinear regression. One-way and two-way analysis of variance. Use of existing statistical computing package.

STAT 51300 Statistical Quality Control (3 cr.) P: STAT 51100. Control charts and acceptance sampling, standard acceptance plans, continuous sampling plans, sequential analysis, statistics of combinations, and some nonparametric methods. Use of existing statistical computing packages.

BIOS S515 Biostatistical Practicum (1-3 cr.) P: STAT 52100; BIOS S527, S546; or consent of instructor. Real-world projects in biostatistics involving participation in consulting sessions, directed reading in the literature, research ethics, design of experiments, collection of data and applications of biostatistical methods. Detailed written and oral reports required. May be repeated, up to 6 credits.

STAT 51400 Design of Experiments (3 cr.) P: STAT 51200. Fundamentals, completely randomized design, randomized complete blocks. Latin squares, multi-classification, factorial, nested factorial, incomplete blocks, fractional replications, confounding, general mixed factorial, split-plot and optimum design. Use of existing statistical computing packages.

STAT 51900 Introduction to Probability (3 cr.) P: MATH 26100. Algebra of sets, sample spaces, combinatorial problems, conditional probability, independence, random variables, distribution functions, characteristic functions, special discrete and continuous distributions, distributions of function of random variables,

limit theorems.

STAT 52000 Time Series and Applications (3 cr.) P: STAT 51900. A first course in stationary time series with applications in engineering, economics, and physical sciences. Stationary, auto-covariance function and spectrum; integral representation of a stationary time series and interpretation; linear filtering; transfer function models; estimation of spectrum; multivariate time series; Kalman filtering, Burg's algorithm.

STAT 52100 Statistical Computing ^ (3 cr.) P: STAT 51200. This course demonstrates how computing can be used to understand the performance of core statistical methods and introduces modern statistical methods that require computing in their application. Covers relevant programming fundamentals in at least two programming environments (e.g. SAS and R/ Splus).

STAT 52200 Sampling and Survey Techniques (3 cr.) P: STAT 51200 or STAT 51100. Survey designs, simple random, stratified, cluster and systematic sampling; systems of sampling; methods of estimation, ratio and regression estimates, costs; non-response analysis; spatial sampling.

STAT 52300 Categorical Data Analysis Models (3 cr.) P: STAT 52800 or equivalent, or consent of instructor. Generating binary and categorical response data, two-way classification tables, measures of association and agreement, goodness-of-fit tests, testing independence, large sample properties. General linear models, logistic regression, probit and extreme value models. Log-linear models in two and higher dimensions; maximum likelihood estimation, testing Goodness-of-fit, partitioning Chi-square, models for ordinal data. Model-building, selection and diagnostics. Other related topics as time permits. Computer applications using SAS.

STAT 52400 Applied Multivariate Analysis (3 cr.) P: STAT 52800 or equivalent, or consent of instructor. Extension of univariate tests in normal populations to the multivariate case, equality of covariance matrices, multivariate analysis of variance, discriminate analysis and misclassification errors, canonical correlation, principal components, factor analysis.

STAT 52500 Intermediate Statistical Methodology (3 cr.) C: STAT 52800 or equivalent or consent of instructor. Generalized linear models, likelihood methods for data analysis, diagnostic methods for assessing model assumptions. Methods covered include multiple regression, analysis of variance for completely randomized designs, binary and categorical response models, and hierarchical log-linear models for contingency tables.

BIOS S527 Introduction to Clinical Trials (3 cr.) P: STAT 51200, exposure to survival analysis; or consent of instructor. Prepares biostatisticians for support of clinical trial projects. Topics: fundamental aspects of the appropriate design and conduct of medical experiments involving human subjects including ethics, design, sample size calculation, randomization, monitoring, data collection analysis and reporting of the results.

STAT 52800 Mathematical Statistics I (3 cr.) P: STAT 51900. Sufficiency and completeness, the exponential family of distributions, theory of point estimation, Cramer-Rao inequality, Rao-Blackwell Theorem with applications, maximum likelihood estimation, asymptotic distributions of ML estimators, hypothesis testing, Neyman-Pearson Lemma, UMP tests, generalized likelihood ratio test, asymptotic distribution of the GLR test, sequential probability ratio test.

STAT 52900 Bayesian Statistics and Applied Decision Theory (3 cr.) C: STAT 52800 or equivalent. Bayesian and decision theoretic formulation of problems; construction of utility functions and quantification of prior information; choice of prior; methods of Bayesian decision and inference; Bayesian computations; MCMC methods; empirical Bayes; hierarchical models, Bayes factors; combination of evidence; game theory and minimax rules, Bayesian design and sequential analysis.

BIOS S530 Statistics Methods in Bioinformatics ^ (3 cr.) P: STAT 51200, 51900; or consent of instructor. Covers statistical methods used in many areas of bioinformatics research, including sequence alignment, genome sequencing and gene finding, gene expression microarray analysis, transcriptional regulation and sequence motif finding, comparative genomics, and proteomics.

BIOS S531 Sequence Analysis ^ (3 cr.) P: BIOS S530. This course is designed to develop students' skills in sequence analyses and communications through multiple real life projects covering pairwise alignment; multiple alignment, evolution and phylogeny, and cis-regulatory analysis.

STAT 53200 Elements of Stochastic Processes (3 cr.) P: STAT 51900 or equivalent. A basic course in stochastic models including discrete and continuous time processes, Markov chains and Brownian motion. Introduction to topics such as Gaussian processes, queues and renewal processes and Poisson processes. Applications to economics, epidemic models, birth and death processes, point processes, and reliability problems.

STAT 53300 Nonparametric Statistics (3 cr.) P: STAT 51900 or equivalent. Binomial test for dichotomous data, confidence intervals for proportions, order statistics, one-sample signed Wilcoxon rank test, two-sample Wilcoxon test, two-sample rank tests for dispersion, Kruskal-Wallis test for one-way layout. Runs test and Kendall test for independence, one and two sample Kolmogorov-Smirnov tests, nonparametric regression.

STAT 53600 Introduction to Survival Analysis (3 cr.) P: STAT 51700. Deals with the modern statistical methods for analyzing time-to-event data. Background theory is provided, but the emphasis is on the applications and the interpretations of results. Provides coverage of survivorship functions and censoring patterns; parametric models and likelihood methods, special life-time distributions; nonparametric inference, life-tables, estimation of cumulative hazard functions, the Kaplan-Meier estimator; one and two-sample nonparametric tests for censored data; semiparametric proportional hazards regression (Cox Regression), parameters' estimation, stratification, model fitting strategies and model interpretations. Heavy use of statis-

tical software such as Splus and SAS.

BIOS S546 Applied Longitudinal Data Analysis (3 cr.) P: STAT 51200, 52500; or permission of instructor. Covers modern methods for the analysis of repeated measures, correlated outcomes and longitudinal data. Topics: repeated measures ANOVA, random effects and growth curve models, generalized estimating equations (GEE) and generalized linear mixed models (GLMMs). Extensive use of statistical software, e.g. SAS, R.

BIOS S587 Nonlinear Mixed Models ^ (3 cr.) P: Undergraduate statistics course and familiarity with statistical inference. This course will develop the student's ability to understand the pharmacokinetic/pharmacodynamic model, fit the nonlinear mixed model through the required software package, conduct the diagnosis of model fitting, perform the hypothesis tests, and provide the interpretation of the data.

BIOS S598 Topics in Biostatistical Methods (0-3 cr.) P: Consent of instructor. Directed study and reports for students who wish to undertake individual reading and study on approved topics.

600 Level

STAT 61900 Probability ^ (3 cr.) P: STAT 51900, 52800. Theory Measure theory based course in probability. Topics include Lebesgue measure, measurable functions and integration. Radon-Nikodym Theorem, product measures and Fubini's Theorem, measures on infinite product spaces, basic concepts of probability theory, conditional probability and expectation, regular conditional probability, strong law of large numbers, martingale theory, martingale convergence theorems, uniform integrability, optional sampling theorems, Kolmogorov's Three series Theorem, weak convergence of distribution functions, method of characteristic functions, the fundamental weak compactness theorems, convergence to a normal distribution, Lindeberg's Theorem, infinitely divisible distributions and their subclasses.

BIOS S621 Advanced Statistical Computing ^ (3 cr.) P: STAT 52100, experience with R/Splus programming. This course covers selected computational techniques useful in advanced statistical applications and statistical research, such as methods for solving linear equations, numerical optimization, numerical integration, Bayesian methods, bootstrap methods, and stochastic search algorithms).

BIOS S627 Statistics in Pharmaceutical Research (3 cr.) P: STAT 51200; BIOS S527, S546. An overview of the drug development process, including the various phases of development from pre-clinical to post-marketing. Topics: statistical issues in design, study monitoring, analysis and reporting. Additional topics may include regulatory and statistical aspects of population pharmacokinetics and real world applications.

STAT 62800 Advanced Statistical Inference ^ (3 cr.) P: STAT 51900, 52800, C: STAT 61900. Real analysis for inference, statistics and subfields, conditional expectations and probability distributions, UMP tests with applications to normal distributions and confidence sets, invariance, asymptotic theory of esti-

mation and likelihood based inference, U-statistics, Edgeworth expansions, saddle point method.

BIOS S634 Stochastic Modeling in Biomedical and Health Sciences ^ (3 cr.) P: STAT 52800. The aim of this course is to develop those aspects of stochastic processes that are relevant for modeling important problems in health sciences. Among the topics to be covered are: Poisson processes, birth and death processes, Markov chains and processes, semi-Markov processes, modeling by stochastic diffusions. Applications will be made to models of prevalence and incidence of disease, therapeutic clinical trials, clinical trials for prevention of disease, length biased sampling, models for early detection of disease, cell kinetics and family history problems.

BIOS S636 Advanced Survival Analysis ^ (3 cr.) P: STAT 53600, 62800. Addresses the counting process approach to the analysis of censored failure time data. Standard statistical methods in survival analysis will be examined, such as the Nelson-Aalen estimator of the cumulative hazard function, the Kaplan-Meier estimator of the survivor function, the weighted logrank statistics, the Cox proportional hazards regression model, and the accelerated failure time model.

STAT 63800 Stochastic Processes I ^ (3 cr.) P: STAT 61900. Advanced topics in probability theory which may include stationary processes, independent increment processes, Gaussian processes; martingales, Markov processes, ergodic theory.

STAT 63900 Stochastic Processes II ^ (3 cr.) P: STAT 63800. This is the continuation of STAT 63800. We will concentrate on specific chapters from the textbook, including Ch VI-IX (Local Times, Generators, Girsanov's theorem, Stochastic Differential Equations). Some material from another textbook (Karatzas and Shreve, Brownian Motion and Stochastic Calculus), and the instructor's own work, may also be used, especially to cover Feynman-Kac formulas and the connection to PDEs and Stochastic PDEs. New topics not treatable using martingales will also be investigated, include stochastic integration with respect to Fractional Brownian Motion and other, more irregular Gaussian processes; anticipative stochastic calculus; Gaussian and non-Gaussian regularity theory.

BIOS S646 Advanced Generalized Linear Models ^ (3 cr.) P: BIOS S546. The theory of classical and modern approaches to the analysis of clustered data, repeated measures, and longitudinal data: random effects and growth curve models, generalized estimating equations, statistical analysis of multivariate categorical outcomes, estimation with missing data. Discussion of computational issues: EM algorithm, quasi-likelihood methods, Bayesian methods for both traditional and new methodologies.

STAT 69500 Seminar in Mathematical Statistics ^ (1-3 cr.) P: Consent of advisor. Individual Study that meets 3 times per week for 50 minutes per meeting for 16 weeks.

STAT 69900 Research Ph.D. Thesis ^ (1-18 cr.)