Biostatistics & Health Data Science

DEPARTMENT CHAIR:

John Hughes, PhD (https://health.lehigh.edu/faculty/hughes-john/) - *Associate Professor*

Associate Chair:

Eric Delmelle, PhD (https://health.lehigh.edu/faculty/delmelle-eric/) - *Associate Professor*

DEPARTMENT FACULTY:

Ahmed Najeeb Albatineh, PhD, MS, MSOR (https://health.lehigh.edu/faculty/albatineh-ahmed-najeeb/) - *Teaching Professor*
Gideon Gogovi, PhD, MS, Mphil (https://health.lehigh.edu/faculty/gogovi-gideon/) - *Assistant Professor*
Bilal Khan, PhD (https://health.lehigh.edu/faculty/khan-bilal/) - *Professor*
Hsuan-Wei "Wayne" Lee, PhD (https://health.lehigh.edu/faculty/lee-hsuan-wei-wayne/) - *Assistant Professor*
Thomas McAndrew, PhD, MS (https://health.lehigh.edu/faculty/mcandrew-thomas/) - *Associate Professor*
Vinod Namboodiri, PhD (https://health.lehigh.edu/faculty/namboodiri-vinod/) - *Professor and Forlenza '75 Endowed Chair in Health Innovation and Technology*

Contact information:

Health | Science | Technology Building
College of Health Administrative Suite #155
124 East Morton Street
610.758.1800 | cohadvising@lehigh.edu
website: health.lehigh.edu (https://health.lehigh.edu/)
social: @lehighcoh

The Department of Biostatistics and Health Data Science at Lehigh University advances the development and application of statistical, computational, and technological methods to address complex problems in health. The department conducts cutting-edge methodological research, collaborates on interdisciplinary studies, and educates the next generation of health data scientists. Areas of expertise include bioinformatics, computational biology, artificial intelligence, network science, Bayesian methods, spatiotemporal methods, visualization, evolutionary game theory, infectious disease modeling, and assistive and accessibility technologies, with a shared commitment to improving health outcomes through data-driven discovery.

Major & Minor Programs

Biostatistics & Health Data Science	BS Degree, Major
Biostatistics	Minor

B.S. BIOSTATISTICS & HEALTH DATA SCIENCE

The Biostatistics & Health Data Science major draws on knowledge from many disciplines including mathematics, statistics, computing, and epidemiology, but frames these to the singular applied objective of advancing public health. It spans hypothesis generation, study design, data collection, data storage, data processing, analytic methods development, application and interpretation of analyses, dissemination, and translation. It emphasizes rigor, reproducibility, effective communication, and ethical practices. The major is intended for students who are interested in health, healthcare, and health policy from a data focused perspective, or students who seek to acquire analytic, computational, and data skills within the context of human health. The BS degree requires a minimum of 120 credits.

CORE REQUIREMENTS		**36**
Programming Core		
BSTA 030	Data Exploration in R	
BSTA 040	Data Exploration in Python	
Statistics Core		
BSTA 132	Health Data Science I: Inference	
BSTA 133	Health Data Science II: Regression	
AI Core		
BSTA 141	Health Data Science III: Supervised Machine Learning in Health	
BSTA 142	Health Data Science IV: Unsupervised Machine Learning in Health	
Health Core		
POPH 001	Introduction to Population and Public Health	
POPH 002	Population Health Research Methods & Application	
EPI 104	Fundamentals of Epidemiology	
EPI 305	Intermediate Epidemiology	
ELECTIVES		**24**
(24 credits from 3 clusters, at least one course from each cluster and a minimum of 6 credits of which are from Data or Methods).		
Elective courses may count towards college distribution requirements.		
Society Cluster		
POPH 003	Justice, Equity, and Ethics in Population Health	
BSTA 007	Frontiers of AI in Health	
CGH 103	Biological & Environmental Determinants of Health	
CGH 104	Sociocultural & Political Determinants of Health	
CGH 105	Commercial Determinants of Health	
CGH 313	Health Policy and Politics	
CGH 332	Aging, Health, and Social Policy	
Data Cluster-all courses have a prerequisite completion of AI Core		
BSTA 372	Analyzing Electronic Health Record Data	
BSTA 373	Analyzing Clinical Natural Language Data	
BSTA 374	Analyzing Health GIS Data	
BSTA 375	Analyzing Health Sensor Data	
BSTA 376	Deep Learning for Healthcare	
Methods Cluster-all courses have a prerequisite completion of Statistics Core		
BSTA 381	Analysis of Dependent Data	
BSTA 383	Survival Analysis	
BSTA 384	Network Analysis	
BSTA 309	Outbreak Science & Public Health Forecasting I	
BSTA 386	Bayesian Analysis	
BSTA 387	Analyzing Data in SAS	
Portfolio Project-Concurrent with Data/Methods electives		**1**
BSTA 399	Portfolio Project	
MATH & COMPUTER SCIENCE DISTRIBUTION		**14**
MATH 051	Survey of Calculus I	
MATH 052	Survey of Calculus II	
MATH 043	Survey of Linear Algebra	
CSE 012	Introduction to Programming with Python	
Total Credits		**75**

MINOR PROGRAMS

Minor programs in the College of Health are open to students from across the university. Students who have completed courses in their major that are also required for a minor may only count one course for both. For more information, contact the College of Health at cohadvising@lehigh.edu. To declare

any minor offered by the College of Health, complete this form (https://powerforms.docusign.net/329016c8-371a-40cf-9ed9-f08c3197cc71/?env=na3&acct=4522e8bc-42ec-46ec-af83-a167d8a26e3f&accountId=4522e8bc-42ec-46ec-af83-a167d8a26e3f&recipientLang=en).

Minor in Biostatistics

The Biostatistics minor provides quantitatively oriented students with conceptual knowledge and hands-on skills in applied statistics and data science techniques commonly employed in the field of biostatistics. The curriculum seeks to prepare students to interpret and contribute to quantitative research in health-related fields, including community and population health. The minor serves to broaden student employment possibilities post-Lehigh while making them more competitive as applicants to health-related graduate programs that favor prior training in applied statistics.

BSTA 101 & BSTA 102	Population Health Data Science I and Population Health Data Science I Algorithms Lab	4
BSTA 103 & BSTA 104	Population Health Data Science II and Population Health Data Science II Algorithms Lab	4
Electives (choose 3 from the list below, or in consultation with your adviser)		9
BSTA 308	Advanced R Programming	
BSTA 309	Outbreak Science & Public Health Forecasting I	
BSTA 310	Assistive Technologies	
BSTA 320	Independent Study or Research in Biostatistics	
Total Credits		**17**

Courses

BSTA 003 Computational Thinking 3 Credits
This course introduces computational thinking as a problem-solving methodology in health and biological sciences. You will explore the approach of developing theoretical models for natural events and converting them into computer simulations using tools like R, Python, MATLAB, or SAS. The course emphasizes fundamental programming concepts, making it suitable for beginners, while also highlighting computational thinking in health. Additionally, the course explores ethics in computational science, covering responsible algorithmic decision-making, data management, privacy, bias, and transparency in computing.

BSTA 005 Statistical Literacy in Health 3 Credits
This course is designed to introduce students with a fear of all things mathematical to the importance of statistics in health research. Students will learn how to read and understand basic statistical concepts and methods used in health research, such as probability, sampling, hypothesis testing, and correlation. Students will also learn to interpret tables and statistical findings in the health literature.

BSTA 007 (POPH 007) Frontiers of AI in Health 3 Credits
This course presents a broad contemporary survey of the actual and potential contributions of Artificial Intelligence and Health Data Science in addressing public health challenges. By reading recent articles that describe case studies of AI in health and healthcare and by engaging in discussions both in class and online, students will come to appreciate the many unsolved problems in public health and how one may evaluate the potential benefits and risks of exciting new data-centric solutions made possible by AI.

BSTA 008 The Art of AI Conversation: Prompting GPT and Its Peers 3 Credits
This introductory course explores Large Language Models (LLMs) like ChatGPT and Claude, emphasizing effective prompt engineering and critical evaluation of AI-generated content. Students will learn how to formulate queries, assess outputs, and refine prompts while addressing ethical and domain-specific challenges. Using health-related examples, the focus is on general, cross-disciplinary interactive AI methods, not computer science or software development. Students will learn to use AI dialogue systems responsibly and creatively, with an understanding of the tradeoffs of various prompting techniques.

BSTA 030 Foundations of Health Data Science Using R 3 Credits
This course introduces students to the mathematical and computing principles that underlie health data science. Topics include R programming fundamentals, exploratory data analysis, introductory probability theory, and stochastic simulation. Students will use R to conduct exploratory analyses of real health-related datasets, to do computing pertaining to theoretical probability distributions, and to simulate data from probability models that arise frequently in health data science. Knowledge of differential and integral calculus would be helpful but is not required.
Prerequisites: CSE 012

BSTA 040 Data Exploration in Python 3 Credits
This course provides an introduction to the fundamentals of programming in Python. Students will gain experience designing, implementing, and testing their Python code, as well as in using Jupyter Notebooks, and IPython for statistics and data analysis. Multiple programming paradigms will be explored. The course covers Python data types, input, and output, and control flow in the context of preparing, cleaning, transforming, and manipulating data. In addition, students will use Python to conduct exploratory data analyses, including computing descriptive statistics.
Prerequisites: CSE 012

BSTA 101 Population Health Data Science I 3 Credits
This course provides an introduction to the use of statistics in health. Topics include data presentation, descriptive statistics, probability and probability distributions, parameter estimation, hypothesis testing, analysis of contingency tables, analysis of variance, linear and logistic regression models, and sample size and power considerations. Students develop the skills necessary to perform, present, and interpret basic statistical analyses. Must be taken in conjunction with BSTA 102.
Corequisites: BSTA 102

BSTA 102 Population Health Data Science I Algorithms Lab 1 Credit
Students will use a statistical computing platform to apply concepts learned in BSTA 101 and attain autonomy in handling real-world data. Lab must be taken concurrently with lecture (BSTA 101 Population Health Data Science I).
Corequisites: BSTA 101

BSTA 103 Population Health Data Science II 3 Credits
This course is a continuation of BSTA 101. Topics include an overview of generalized linear models, simple and multiple linear regression, regression models for binary data, regression models for count data, quasi-likelihood methods, and extensions of generalized linear models. Must be taken with BSTA 104.
Prerequisites: BSTA 101 and BSTA 102
Corequisites: BSTA 104

BSTA 104 Population Health Data Science II Algorithms Lab 1 Credit
Students will use a statistical computing platform to apply regression techniques learned in BSTA103 Population Health Data Science II to health datasets. Lab must be taken concurrently with lecture (BSTA103 Population Health Data Science II).
Prerequisites: BSTA 101
Corequisites: BSTA 103

BSTA 120 (CGH 120, EPI 120, POPH 120) Independent Study or Research 1-4 Credits
This course can be directed readings or research in Biostatistics or an experiential learning experience that puts student's understanding of Biostatistics into practice. Department permission required.
Repeat Status: Course may be repeated.

BSTA 130 Internship 1-4 Credits
In this introductory course, students will engage in supervised work in Biostatistics. Placements will be arranged to suit individual interests and career goals. Potential internship sites include government agencies, non-profit organizations, and the private sector. A written report is required, and a preceptor evaluation will be required. Department permission is required.
Repeat Status: Course may be repeated.

BSTA 132 Health Data Science I: Inference 4 Credits
This course provides an introduction to methods of statistical inference as applied to health data. Topics covered include hypothesis testing, confidence intervals, analysis of variance, correlation, and non-parametric methods. The course will illustrate these concepts using data from the health context. In addition to traditional methods of learning, computing will be a significant component of the course, ensuring students acquire the skills to both formulate and answer pressing questions in population health.
Prerequisites: MATH 052 and MATH 043 and BSTA 030

BSTA 133 Health Data Science II: Regression 4 Credits
This course provides an introduction to generalized linear models as applied to health data. Topics covered include models for binary data, models for nominal and ordinal data, models for count data, quasi-likelihood methods, and Bayesian generalized linear models. The course will illustrate these concepts using data from the health context. In addition to traditional methods of learning, computing will be a significant component of the course, ensuring students acquire the skills to both formulate and answer pressing questions in population.
Prerequisites: BSTA 132

BSTA 141 Health Data Science III: Supervised Machine Learning in Health 4 Credits
Supervised machine learning is used to create automated systems that sift through labeled/continuous data at high speed to make predictions with minimal human intervention. This course provides students with skills in applying supervised machine learning in contexts of population health. We will cover regression, classification, cross-validation, hyperparameter selection, feature selection, feature engineering, ensemble methods, regularization, and reinforcement learning. Students will learn concepts through hands-on engagement with health data sets, preparing them to contribute effectively to data-driven precision population health.
Prerequisites: MATH 052 and BSTA 040

BSTA 142 Health Data Science IV: Unsupervised Machine Learning in Health 4 Credits
Unsupervised machine learning is used to discover hidden patterns and structures in high-dimensional unlabeled health data. This course will survey leading techniques for clustering and dimensionality reduction. The course will cover hierarchical and density-based clustering techniques, along with modeling using Gaussian mixtures, factor analysis, and principal component analysis. Applications considered will include patient clustering for personalized treatment, anomaly detection for early disease identification, and dimensionality reduction for efficient analysis of diverse and complex medical datasets.
Prerequisites: BSTA 141 and MATH 052 and MATH 043 and BSTA 040

BSTA 150 Special Topics in Biostatistics 3-4 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an intermediate level.
Repeat Status: Course may be repeated.

BSTA 160 Biostatistics Study Abroad 1-4 Credits
Biostatistics focused course taken during an abroad experience.
Repeat Status: Course may be repeated.

BSTA 300 Apprentice Teaching 1-4 Credits
Repeat Status: Course may be repeated.

BSTA 308 Advanced R Programming 3 Credits
R language syntax and structure. R programming techniques. Emphasis on structured design for medium to large programs. R package development fundamentals. Capstone development project.
Prerequisites: (BSTA 101 and BSTA 102) or (BSTA 103 and BSTA 104)

BSTA 309 Outbreak Science & Public Health Forecasting I 3 Credits
This course aims to introduce students to models that describe the spread of a pathogen through a population, and how models can support public health decisions. The course will be split into four parts: (i) the factors that motivate public health actions, (ii) epidemic models such as the Reed-Frost and SIR, (iii) statistical time series and forecasts, (ii) a focus on ensemble building. Students will be expected to complete mathematical/statistical exercises and write code that simulates infectious processes.
Prerequisites: BSTA 101 and BSTA 102 and BSTA 103 and BSTA 104

BSTA 310 (CSE 310) Assistive Technologies 3 Credits
This class will introduce typical challenges faced by persons with disabilities and the role of assistive technologies (ATs) in solving such challenges. The class will examine opportunities presented by recent advances in mobile and AI technologies. Working in groups, each student will be expected to acquire and apply relevant skills in designing AT solutions. The class can be taken by students with diverse backgrounds including the following: community and population health, social and behavioral sciences, business, engineering and computer science.
Prerequisites: CSE 017 or (BSTA 101 and BSTA 102)
Attribute/Distribution: Q

BSTA 320 (CGH 320, EPI 320, POPH 320) Independent Study or Research in Biostatistics 1-4 Credits
This course can be directed readings or research in Biostatistics or an experiential learning experience that puts student's understanding of Biostatistics into practice. Department permission required.
Repeat Status: Course may be repeated.

BSTA 330 Internship 1-4 Credits
In this advanced course, students will engage in supervised work in Biostatistics. Placements will be arranged to suit individual interests and career goals. Potential internship sites include government agencies, non-profit organizations, and the private sector. A written report is required, and a preceptor evaluation will be required. Department permission is required.
Repeat Status: Course may be repeated.

BSTA 350 Special Topics in Biostatistics 3-4 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an advanced level.
Repeat Status: Course may be repeated.

BSTA 360 Biostatistics Study Abroad 1-4 Credits
Upper-level biostatistics focused course taken during an abroad experience.
Repeat Status: Course may be repeated.

BSTA 372 Analyzing Electronic Health Record Data 3 Credits
This course will explain the structure and provide computing skills to analyze Electronic Health Record (EHR) data. Through a series of health-related case studies, students will have the opportunity to experience EHR as a comprehensive platform to support best-in-class evidence-based care and as the core component for big data analytics to help care organizations adapt and transform into learning organizations. The course will present a number of EHR data architectures, data standards, quality assessment, and workflow methods.
Prerequisites: BSTA 142

BSTA 373 Analyzing Clinical Natural Language Data 3 Credits
This course will convey specialized clinical natural language processing (NLP) principles and methods, as well as how to write regular expressions and parse and collate information from text-rich health documents such as electronic health records, clinical notes, and peer-reviewed medical literature. The course will engage real-world data sets for students to develop text-processing strategies. Computing will be a significant component of the course, ensuring students acquire the skills necessary to work with clinical natural language data.
Prerequisites: BSTA 142

BSTA 374 Analyzing Health GIS Data 3 Credits
This course will convey specialized methodologies of data collection and the statistical analysis of spatial data. Through a series of health-related case studies, students will have the opportunity to explore spatial statistical analysis at a variety of spatial resolutions. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to apply these techniques to health-related GIS data.
Prerequisites: BSTA 142

BSTA 375 Analyzing Health Sensor Data 3 Credits
This course will convey specialized methodologies of data collection and the statistical analysis of health-related time-series data collected from sensors. Of particular interest are data generated by environmental sensors, wearable devices, and medical instrumentation. Through a series of health-related case studies, students will have the opportunity to explore signal processing, filtering, modeling, and forecasting techniques. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to apply these techniques to health-related sensor data.
Prerequisites: BSTA 142

BSTA 376 Deep Learning for Healthcare 3 Credits
This course will convey the specialized methods of deep learning in the context of health data. Through health-related case studies, students will learn to engage deep learning models and healthcare applications such as clinical predictive models, computational phenotyping, patient risk stratification, treatment recommendation, and medical imaging analysis. The course will engage with real-world data sets via computing using Jupyter and PyTorch, ensuring that students acquire the skills necessary to apply deep learning techniques to health data.
Prerequisites: BSTA 142

BSTA 381 Analysis of Dependent Data 3 Credits
This course will convey specialized methodologies needed to analyze and model dependent data. By considering dependent data from a series of health-related case studies, students will have the opportunity to explore different types of statistical association, random effects models, generalized estimating equations, copula models, and nonparametric methods for dependent data. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to carry out a wide range of analyses of health-related dependent data.
Prerequisites: BSTA 133

BSTA 383 Survival Analysis 3 Credits
This course will present methodologies needed to model time-to-event data. By considering censored (i.e., incomplete) health data from a series of case studies, students will explore nonparametric estimation (e.g., life table methods, Kaplan–Meier estimator), nonparametric methods for comparing the survival experience of populations, and semiparametric and parametric methods of regression for censored outcome data. Computing will be a significant component of the course, ensuring students acquire the skills necessary to conduct time-to-event analyses of health-related data.
Prerequisites: BSTA 133

BSTA 384 Network Analysis 3 Credits
This course will convey specialized methodologies needed to analyze and model network data. By considering relational data from a series of health-related case studies, students will have the opportunity to explore mathematical description of networks, social network measures, exponential random graph models of networks, network sampling, and visualization. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to carry out a wide range of network-based analyses of health-related data.
Prerequisites: BSTA 133

BSTA 386 Bayesian Analysis 3 Credits
This course will provide a basic introduction to Bayesian concepts and methods with an emphasis on the data analysis in the context of health. We will discuss model choice, including the assessment of prior distributions. We will discuss how to conduct inference in a Bayesian setting, through posterior means, credible intervals and hypothesis testing. The Analyses will be performed using the freely available software Jags as implemented in the R packages rjags and R2jags.
Prerequisites: BSTA 133

BSTA 387 Analyzing Data in SAS 3 Credits
This course will introduce the student to the SAS programming language in a lab-based format. The objective is for the student to develop programming and statistical computing skills to address data management and analysis issues using SAS. The course will also provide a survey of some of the most common data analysis tools in use today and provide decision-making strategies in selecting the appropriate methods for extracting information from data.
Prerequisites: BSTA 133

BSTA 396 1-4 Credits
Repeat Status: Course may be repeated.

BSTA 399 Portfolio Project 1 Credit
This course will must be taken concurrently with an elective in either the Data or Methods clusters of the program. Students must inform the instructor for the associated elective about their registration in the Portfolio Project course. Portfolio Project students may be assigned additional material/assignments, and will be required to complete a significant report in the associated elective course.

BSTA 402 Biostatistics in Health 3 Credits
This course provides an introduction to the use of statistics in health. Topics include descriptive statistics, probability distributions, parameter estimation, hypothesis testing, analysis of contingency tables, analysis of variance, regression models, and sample size and power considerations. Students develop the skills necessary to perform, present, and interpret statistical analyses; and attain autonomy in handling real-world data using a statistical computing environment.

BSTA 403 Health Applications in Statistical Learning 3 Credits
This course will explore common statistical models used to analyze both continuous, discrete, and time to event data: simple and multivariate linear regression, logistic regression, poisson and negative binomial regression, and survival models. An emphasis will be placed on supervised learning. Throughout the semester, students will apply the theoretical background they learn in class to population health data sets, generating their own hypotheses and testing them with rigorous statistical methods.
Prerequisites: BSTA 402

BSTA 404 Data Architecture, Mining, and Linkage 3 Credits
This course will focus on collecting, storing, and formatting data for use in population health data analysis. Students will learn fundamental concepts and best practices for working with data, how to use Python to scrape the internet for data related to population health and learn how to link a diverse set of data together to test novel hypotheses students themselves pose during class.

BSTA 405 Survey Sampling Methods 3 Credits
In this course, students are introduced to key concepts such as sampling theory, questionnaire design, survey planning, questions ordering, sources of errors, types of bias in surveys , and sampling from finite vs. infinite populations. Furthermore, students will explore sampling designs including simple random sampling, stratified and systematic sampling, and cluster sampling. Students will explore concepts like design effects and implement methods to conduct power and sample size calculations for different population parameters in different sampling designs using standard/free software.
Prerequisites: BSTA 402

BSTA 409 Outbreak Science & Public Health Forecasting I 3 Credits
This course aims to introduce students to models that describe the spread of a pathogen through a population, and how models can support public health decisions. The course will be split into four parts: (i) the factors that motivate public health actions, (ii) epidemic models such as the Reed-Frost and SIR, (iii) statistical time series and forecasts, (ii) a focus on ensemble building. Students will be expected to complete mathematical/statistical exercises and write code that simulates infectious processes.

BSTA 410 (CSE 410) Assistive Technologies 3 Credits
This class will introduce typical challenges faced by persons with disabilities and the role of assistive technologies (ATs) in solving such challenges. The class will examine opportunities presented by recent advances in mobile and AI technologies. Working in groups, each student will be expected to acquire and apply relevant skills in designing AT solutions. The class can be taken by students with diverse backgrounds including the following: community and population health, social and behavioral sciences, business, engineering and computer science.

BSTA 420 (CGH 420, POPH 420, PUBH 420) Independent Study or Research in Biostatistics 1-4 Credits
This course can be directed readings or research in Biostatistics or an experiential learning experience that puts student's understanding of Biostatistics into practice. Department permission required.
Repeat Status: Course may be repeated.

BSTA 450 Special Topics in Biostatistics 3 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an advanced level.
Repeat Status: Course may be repeated.