



The Master's in Applied Biostatistics Program at Boston University

Program Handbook
2026-2027



Boston University School of Public Health

Biostatistics Program Contacts

Biostatistics Department

Sph.bu.edu/biostatistics

biostat@bu.edu

(p) 617-358-2222

(f) 617-378-2752

Sr. Academic Program Administrator

Lauren Hall

lmhall21@bu.edu

(p) (617)358-2706

MSAB Program Co-Directors

Chunyu Liu, PhD

liuc@bu.edu

617-358-5104

Helen Jenkins, PhD

helenje@bu.edu

617-358-2640

Other Contacts

Biostatistics Department Chair

Scarlett Bellamy

bellamys@bu.edu

Mission Statements

The mission of the Boston University School of Public Health is to improve the health of local, national and international populations, particularly the disadvantaged, underserved and vulnerable, through excellence and innovation in education, research and service.

The mission of the Department of Biostatistics:

- To teach students the proper conduct of research studies through rigorous study design and appropriate descriptive and analytic methods that enable valid, interpretable conclusions to be drawn
- To collaborate in research projects to ensure that studies are properly designed, appropriately analyzed, and suitably interpreted
- To develop and evaluate new methods of biostatistical analysis and strategies for study design

Table of Contents

Section 1: Introduction	5
Section 2: MS in Applied Biostatistics	6
Section 3: Admission Requirements	7
Section 4: MS Degree Requirements	9
Curriculum	10
Part-Time Option	11
Section 5: Research Experience Policies and Procedures	12
Section 6: Course Descriptions	25
Section 7: Biostatistics Faculty	29
MS Degree Audit Sheet	36

Section

1

Introduction

OVERVIEW

Biostatistics is the branch of statistics used in biomedical research, biology, and public health and is at the root of all major change in public health and discovery in biomedical sciences. Biostatisticians work with multidisciplinary teams to design, analyze, and summarize data from experimental science, genetics, clinical trials, and observational studies. Biostatisticians' work is compelling because it helps to distinguish real patterns from randomness and ultimately leads to sound decisions that promote better health and higher quality of life around the world. From discussions with investigators about fine-tuning research questions to developing appropriate study designs, planning and implementing proper statistical analyses, and writing up the results, biostatisticians are involved in all aspects of research investigations.

Job prospects for biostatisticians are robust and track the projected growth of health care and related areas with a predicted 21% growth in employment opportunities over the next 6 years, according to [Fortune in 2016](#). The Bureau of Labor Statistics predicts a 27% growth in employment opportunities for statisticians between 2017 and 2024, much faster than the average growth of most other occupations

There is a growing demand for biostatisticians with expertise in design and analysis of clinical trials and observational studies, statistical modeling skills for big data including genetic, genomic and health outcomes data, facility with multiple statistical computing programs, strong oral and written communication skills, and the ability to work effectively both independently and in interdisciplinary teams. Boston University's innovative 1-year MS program in Applied Biostatistics will provide you with comprehensive training in biostatistical methods, real-world experience through rotation/internship, and the chance to start building your professional network in Boston.

Section

2

MS in Applied Biostatistics

THE MASTER OF SCIENCE

The MS in Applied Biostatistics prepares students for the challenges of an ever-changing, technology-driven world, giving them the skills to adapt quickly to new data collection and analysis methods. The MS program is an accelerated training focusing on the practical skills needed for future work in fields such as biomedical research, pharmaceuticals, local and national government agencies, nonprofits, and contract research organizations. For a full-time student, the MS program will only take a year to complete, starting with an accelerated Advanced Statistical Training in August, which provides the probability, statistical inference, and computing skills necessary for a solid background in biostatistics. Through the fall and spring semesters, students take all of the necessary classes to acquire the essential skills and knowledge of a biostatistician and perform a research rotation, which provides practical experience working in the field. The research rotation ideally will develop into practical training in the form of a paid or unpaid internship during mid-May, June and July. These rotations allow students to make contacts within the field of biostatistics, building valuable relationships, as well as gaining professional experience. The MS in Applied Biostatistics is meant to be a very practical training program giving students the background and experience they need to build a fulfilling career in the growing field of biostatistics.

Section

3

Admission Requirements

APPLICATION FOR ADMISSION

All applications for admission are to be completed online through [SOPHAS](#). Review all of the application instructions carefully, including the [Frequently Asked Questions](#). The priority application deadline for all applicants to the MS in Applied Biostatistics is February 11.

APPLICATION REQUIREMENTS

- SOPHAS online application
- Official transcripts in English or WES ICAP Evaluation (international applicants) from all colleges and universities attended
- Three letters of recommendation
- Statement of Purpose and Objectives
- Proof of English Proficiency (international applicants)
- CV or Resume

PROGRAM PRE-REQUISITES

- At least the equivalent of a Bachelor's degree; no specific undergraduate major is required
- One year of calculus, including multivariate calculus
- One formal course in linear algebra (with a minimum of two credits)*

***Applicants who lack the linear algebra prerequisite can be accepted on the condition that they complete the requirement before the beginning of the Advanced Statistical Training.**

International students

Students from abroad must submit a [WES ICAP](#) evaluation to SOPHAS instead of their original transcripts. Directions can be found in the link above.

Students must also submit results of the Test of English as a Foreign Language (TOEFL), or the International English Language Testing System (IELTS) with minimum score

TOEFL ibt 100 or higher
TOEFL bpt 600 or higher
IELTS (academic exam only) 7.0 total band score

Applicants who do not meet the minimum requirements of English proficiency are not eligible for admission. This requirement is waived *only* if the student has received, or expects to receive, an undergraduate or graduate degree from a college or university in any of the following countries prior to enrollment in the MS in Applied Biostatistics: The United States, Canada, The United Kingdom, Australia, New Zealand, or Ireland. Studying in the medium of English in a country other than the ones previously listed or studying as an exchange student in an English-speaking country without completing a degree program is not sufficient to waive the English proficiency requirement.

Department-specific admissions questions should be directed to biostms@bu.edu. Questions about the general admissions process should be directed to the School of Public Health at asksph@bu.edu.

Section

4

MS Degree Requirements

OVERVIEW OF THE MS DEGREE

The Master of Science in Applied Biostatistics program is aimed primarily at students with the equivalent of a Bachelor's degree who wish to pursue advanced study in the practical application of biostatistics. It is an intensive one-year degree that will provide training in biostatistical methods, real-world experience through one rotation and one internship, and the opportunity to begin building a professional network. The program prepares students for successful careers in biomedical research enterprises, pharmaceutical companies, local and national government agencies, and research.

MS LEARNING OUTCOMES

By the end of the program, students will have the knowledge, skills and professional confidence to:

- Apply the statistical methods commonly used in biomedical research, including
 - Analysis of variance
 - Linear regression, logistic regression and log-linear models
 - Survival Analysis
 - Mixed models and analysis of correlated data
 - Bayesian Analysis
 - Statistical computing
 - Analysis of observational studies
- Apply basic principles and methods to design, plan, conduct, and interpret biomedical studies in
 - Clinical trials
 - Observational studies
 - Big genomic and genetics data
- Provide effective biostatistical advice as a member of a team with strong
 - Consultancy skills
 - Oral and written communication skills

Students will gain these competencies through a blend of active learning opportunities including in-class and field-based experiences.

MS DEGREE REQUIREMENTS

The Boston University School of Public Health requires students pursuing a Master of Science in Applied Biostatistics to complete:

- the 34-credit curriculum
- PH 746 Career P.R.E.P. Professional Development Course (0 credits)
- PH 700 Foundations of Public Health (0 credits)
- 100 hours of Research Rotation
- 400 hours of Practical Training

All students are required to complete the degree within five years.

MS Curriculum

Full-time students in the MS program must complete a total of 34 credits as follows:

August	Fall (18 cr)		Spring (16 cr)		May-July
BS800 (2cr). Accelerated statistical training PH 700 (0 cr) Foundations of Public Health (OL, completed before arrival)	PH746(0 cr) Professional development course		Research Rotation in Biostatistics (100 hours)		400 hours practical training
	BS806 (4cr). Statistical Learning with Applications in R		BS853 (4cr). Generalized Linear Models		
	BS852 (4cr). Biostatistical Methods for Observational Studies		BS857 (4cr). Analysis of Correlated Data		
	BS858 (4cr). Statistical Genetics I		BS851 (4cr). Applied Statistics in Clinical Trials I		
	BS803 (2cr). Statistical Programming for Biostatisticians	BS728 (2cr). Public Health Surveillance	BS849 (2cr). Bayesian Modeling	BS831 (2cr). Analysis of Big Genomic Data	

*BS800 needs to be taken before Fall I, unless applicants can show equivalent material already completed

Grade Requirements: Students must earn a grade of B- or better in all courses applied to the degree.

Policies

All students must adhere to all Boston University School of Public Health [academic policies](#) and the University's [Administrative Policies](#). Note that this information [may change](#) at any time.

MS Part-time Option

Students in the MS program may opt to complete the required 34 credits in two years. A possible part-time schedule could look like this:

	Year 1				
August	Fall I (8 cr)**		Spring I (8 cr)***		May-July
BS800* (2cr). Accelerated statistical training PH 700 (0 cr) Foundations of Public Health (OL, completed before arrival)	PH746 Professional development course				
	BS806 (4cr). Statistical Learning with Applications in R		BS853 (4cr). Generalized Linear Models		
	BS852 (4cr). Biostatistical Methods for Observational Studies		BS851 (4cr). Applied Statistics in Clinical Trials I***		
	Year 2				
	Fall II (8 cr)**		Spring II (8 cr)***		May-July
			Research Rotation in Biostatistics (100 hours)#		400 hours of practical training
	BS858 (4cr). Statistical Genetics I		BS857 (4cr). Analysis of Correlated Data		
	BS803 (2cr). Statistical Programming for Biostatisticians	BS728 (2cr). Public Health Surveillance	BS849 (2cr). Bayesian Modeling	BS831 (2cr). Analysis of Big Genomic Data	

*BS800 needs to be taken before Fall I, unless applicants can show equivalent material already completed

** BS806 and BS852 must be taken in Fall I and BS858 must be taken in Fall II.

*** BS851 could be swapped with block (BS849 & BS831) between Spring I and Spring II

#The 100 hour research rotation can be spread out throughout the two year program

Note For Students Conducting Practical Training Outside of MA (within the U.S.)

The Massachusetts Department of Higher Education, in its capacity as the State Authorization Reciprocity Agreement (SARA) portal entity for Massachusetts, reviews and evaluates student complaints regarding distance education activities offered by Boston University in accordance with 610 CMR 12.07. This includes practicums taking place outside of Massachusetts. The Boston University state authorization approval and related complaint

resolution process can be found on the State Authorization and Distance Education website at www.bu.edu/state-authorization. Additional information can be found on the [BU Compliance website](#) and the [Academic Bulletin](#) regarding ways that students can report concerns and utilize University complaint procedures.

Section

5

Research Experience Policies and Procedures

1. Background Information

Every MS student admitted to the Boston University MS in Applied Biostatistics Program is required to fulfill a 100-hour biostatistics or bioinformatics research experience during the Fall, Spring or Summer semester supervised by a Boston University faculty. Full-time students complete the research experience **by May**, while part-time students may continue working over the Summer semester. The work does not have to be remunerated as it is a necessary degree requirement of the MS program. This document describes expectations and guidelines related to the Research Experience Module of the MS in Applied Biostatistics.

2. Identification of a mentor

In the fall semester, the APA will work with the Biostatistics department to hold a 'Meet a Mentor Meeting **during October** (date and details to be announced in Sept.). Boston University faculty members who are willing to mentor MSAB students are invited to present their research interests. Full-time MSAB students need to attend this meeting, identify potential mentor(s), and contact the potential mentor(s) to set up meeting(s) to discuss mutual interests, and to form mentor-student teams **by December 1**. Part-time MSAB students have a more flexible timeline to identify a mentor. Trainee and mentor must agree to a work schedule **by December 15** and need to submit MS Research Expectation Form **by December 31**.

3. Student Workload

Students are expected to devote no more than 20 hours per week to their research experience during any academic period (fall semester, spring semester, summer semester). The usual workload is 10-hours per week over a 10-week period. During official University break periods, such as intersession and spring break, or during any academic period in which students are not classified as full-time, they may work up to 40 hours per week.

4. Student Responsibilities

4.1 Set expectations. After a student and a mentor agree to work together, the research mentor and the student must discuss and document expectations. This may include short and long-term project goals; specific duties and responsibilities of the student; criteria/policies on collaboration & publication, and on confidentiality and research data protection; contingencies for sick and vacation absences; and the specific criteria that will be used for performance evaluation.

4.2 Student performance. Students are required to perform the work agreed to and specified by their research mentors for up to 20 hours a week. The usual workload is 10-hours per week over a 10-week period. Students are expected to work hard, respond to feedback, and be responsible in performing the duties of their research experience.

4.3 Feedback. The department will conduct an annual evaluation of Research experience that will be sent to students and research mentors separately. Students are encouraged to provide candid feedback at that time. If problems arise at any point during the year, students should work with their research mentors and academic advisors to resolve the problem; students and research mentors are welcome to consult with the program Co-Directors for guidance as needed.

5. Research Mentor Responsibilities

5.1 Set expectations. At the beginning of the research experience, the mentor and the student must discuss and document expectations. This may include short and long-term project goals; specific duties and responsibilities of the student, criteria/policies on collaboration & publication, and on confidentiality and research data protection; contingencies for sick and vacation absences; *and the specific criteria that will be used for performance evaluation.*

5.2 Performance evaluations. There must be an ongoing communication between the student and the research mentor with feedback on a regular basis about the student's performance. Any concerns related to the student's performance must be discussed with the student as soon as possible, and the research mentor should report those concerns to the Co-Directors. Additionally, every spring the department will send out student evaluation forms to both the research mentor and student, allowing each to provide feedback on their experience.

6. Required documents and timelines

6.1. MS Research Expectation Form (**December 31 for full-time students**). This form should include the following information: 1) student and mentor names and contact information, 2) main aims of the research experience, 3) brief descriptions of the planned research (e.g., background information, aims, methods, and expected outcomes), 4) the start and end dates of the research experience, and 5) hours per week.

We provide an example of MS Research Expectation Form in Appendix.

6.2. MS Research Mid-term Evaluation Report (**March 15 or at the completion of 50 hours for full-time students**). This report should include the following information: 1) student and mentor names and contact information, 2) main aims of the research experience, 3) aims that have been completed, 4) experience summary (good experience or challenges encountered). Codes and results generated from the research are optional to be submitted.

We provide an example of MS Research Mid-Term Evaluation report in Appendix

6.2. MS Research Final Report (**May 15 or at the completion of 100 hours for full-time students**). This report should include the following information: 1) student and mentor names and contact information, 2) main aims of the research experience, 3) a concise report of your project methods, results and conclusion with figures and tables included as necessary and an additional appendix of your code.

We provide an example of MS Research Final report in Appendix

Research Experience Protocol

Research Experience Procedures and Requirements

All students in the MS in Applied Biostatistics need to complete a 100-hour research experience. Each research experience must last at least one semester with the expectation that the student devotes 10 hours per week for 10 weeks performing hands-on data-analysis.

The following details the process and timing for securing research experiences.

Research Experience Appendix 1

Table 1. Summary of Research Experience Procedures, Requirements and Timelines

Process and Timing	Fall	Spring
Step 1: Students are expected to attend the “meet the mentor meeting”	October	
Step 2: Select Research Experience Choose a research experience and contact a research mentor	October/November	
Step 3: Delineate Roles and Responsibilities. Students meet with the research mentor to discuss roles and responsibilities and will draft a document of expectations for the research experience and a brief description of the planned activities. The research experience form must be approved and signed by the research mentor and by one of the co-Directors.	By December 1st	
Step 4: Finalize Schedule Trainee and mentor must agree to a work schedule.	Dec 15	

<i>Process and Timing, continued</i>	<i>Fall</i>	<i>Spring</i>
Step 5: MS Research Expectation Form The research experience form must be signed by the student, the mentor, and one of the Co-Directors. The form should then be submitted to the Co-Directors and the APA. Students need to submit the expectation form that contains the following information: 1) student and mentor names and contact information, 2) main aims of the research experience, 3) brief descriptions of the planned research (e.g., background information, aims, methods, and expected outcomes), 4) the start and end dates of the research experience, and 5) hours per week.	Dec 31	
Step 6: Mid-term evaluation Students need to submit a signed (by the student, mentor, and one Co-Director) interim report to the co- Directors and the APA with the following information: 1) student and mentor names and contact information, 2) main aims of the research experience, 3) aims that have been completed, 4) experience summary (good experience or challenges encountered). Codes and results generated from the research are optional to be submitted.		March 15 Or at the completion of 50 hours

Step 7: Final Report At the completion of the research experience, the student produces a report detailing skills acquired and lessons learned from the research experience. Students should address each of the objectives for the research experience and how each objective was met. The report should not exceed 5 pages with supporting tables/figures as needed. A copy of the document must be submitted to the co-Directors and APA for approval. A completed signature page with student signature, mentor signature, and at least one Co-Director's signature should be submitted for approval with the report.		May 15th Or at the completion of 100 hours
--	--	--

Note that some research experience may require training in human subject research. Please consult the instructions about CITI training <https://www.bumc.bu.edu/ohra/required-training/human-subjects-protection-training/initial-certification/>.

Appendix 2a. MS Research Expectation Form

Template for MS Research Expectation Document

Student name:

Mentor name:

Mentor contact information:

Goals of the research experience:

Brief description of work activity planned for the research experience:

Start date:

End date:

Hours planned per week for research experience work

MSAB Research Rotation Signature Form

Student Name:

Mentor Name:

Mentor Title & Contact Information:

Date:

Research Experience Start Date:

Research Experience:

☐ Genetics

☐ Bioinformatics

☐ Clinical Trials/Clinical Research

☐ Observational Studies

☐ Other (please specify):

X

Student Signature

Date:

X

Mentor Signature

Date:

X

Director's Signature

Date:

Appendix 2b. MS Research Expectation Form (An example)

Student Name: xxxx

Mentor Name: xxx

Mentor Contact Information: xxx@bu.edu

Goals of the Research Experience:

Apply the knowledge and skill obtained through my program in biostatistics to assist and advise the research team on statistical matters related to their projects. In addition, learn about the roles of a biostatistician throughout the process of medical research.

Description of work activity planned for the research experience:

For the beginning portion of the research experience, it is expected that the student will perform descriptive statistics and regressions on a current project evaluating the association between patients with substance use disorder and [DISEASE X]. Once that project is completed, the student will work on other projects available and perform appropriate statistical analysis. There will be biweekly meetings with the mentor to plan tasks and objectives for each two week time period.

Start Date: December 28th, 2020

End Date: March 8th, 2020

Hours planned per week: 10

Appendix 3a. Template MS Research Mid-term Report

Student Name:

Mentor Name:

Mentor Contact Information:

Goals:

Aims that have been completed:

Experience summary (good experience or challenges encountered):

MSAB Research Rotation Signature Form

Student Name:

Mentor Name:

Mentor Title & Contact Information:

Date:

Research Experience Start Date:

Research Experience:

☐ Genetics

☐ Bioinformatics

☐ Clinical Trials/Clinical Research

☐ Observational Studies

☐ Other (please specify):

X

Student Signature

Date:

X

Mentor Signature

Date:

X

Director's Signature

Date:

Appendix 3b. MS Research Mid-term Report

Student Name:

Mentor Name:

Mentor Contact Information:

For my 100-hour research rotation I have been working [mentor name] from the [institute] to analyze the impact of substance use disorder on [DISEASE X]. [Mentor's Name] has been kindly mentoring me regarding statistical methods throughout this process as well.

This experience has been very beneficial so far in testing and expanding my skills and knowledge. I started my rotation becoming IRB approved and gaining access to data on demographics, comorbidities, social history, symptoms, and clinical course from the electronic medical records of 409 patients with [DISEASE X] who were admitted to Boston Medical Center during the first month after the pandemic spread to Boston. The dataset includes 427 variables. The main goal at the beginning of my research was to develop a table one summarizing characteristics and outcomes of patients with [DISEASE X] stratified by the substance use disorder group vs controls. I also read literature surrounding the area of research to understand what other studies have done and what variables might be of interest for our study as well. I utilized SAS to look at the unadjusted associations between the substance use disorder group vs controls for variables of interest. The procedures I used in SAS include proc contents, proc sql, proc univariate, proc freq, proc sort, proc means, and proc t-test. I then utilized R to combine datasets, manipulate the data, and create a presentable table 1 for publication using the package "CreateTableOne" and "KableExtra".

This experience has not only helped my coding skills in R, but has challenged me to deal with a large number of variables and being able to communicate statistical terminology and findings to people with little to no statistical background. It has also brought together many different concepts from the courses I have taken so far in this program. Instead of specific instructions to use methods specific for a course project, I have the responsibility to use my best judgement on what statistical analysis should be applied to this dataset and what would be best for the researcher's goals. I am currently working on models for a few main outcomes of interest for those with [DISEASE X] to understand the impact of substance use disorder and some of the directed therapies. This has challenged me to determine which variables are statistically and clinically relevant, whether missing data is at random or not, checking assumptions of models, and investigating new approaches to modeling such as the different generalized linear models. I am also helping to write and edit a paper to potentially be published that summarizes the work myself and the researchers have accomplished since December. Starting the 21st I will begin another project on outcomes for patients with obstructive sleep apnea and COVID.

Appendix 4a. Template MS Research Final Report

Student Name:

Mentor Name:

Mentor Contact Information

Introduction

Methods

Results

Discussion

Conclusion

Appendix (Code)

MSAB Research Rotation Signature Form

Student Name:

Mentor Name:

Mentor Title & Contact Information:

Date:

Research Experience Start Date:

Research Experience:

☐ Genetics

☐ Bioinformatics

☐ Clinical Trials/Clinical Research

☐ Observational Studies

☐ Other (please specify):

X

Student Signature

Date:

X

Mentor Signature

Date:

X

Director's Signature

Date:

Appendix 4b. MS Research Final Report (an example)

Student Name:

Mentor Name:

Mentor Contact Information:

For my 100-hour research rotation I have been working with [Mentor name] from the Boston Medical Center to analyze the impact of substance use disorder on [Disease X].

The goals originally set for this research experience were to apply the knowledge and skills obtained through my program in biostatistics to assist and advise the research team on statistical matters related to their projects. Additionally, another goal was to learn about the roles of a biostatistician throughout the process of medical research. I believe that these goals have been met and even exceeded. This experience has been very beneficial in testing and expanding my skills and knowledge. I started my rotation becoming IRB approved and gaining access to data on demographics, comorbidities, social history, symptoms, and clinical course from the electronic medical records of 409 patients with [DISEASE X] who were admitted to [Hospital Y] during the [period]. This dataset includes 427 variables. The main objective at the beginning of my research was to develop a table one and two summarizing characteristics and outcomes of patients with [DISEASE X] stratified by the substance use disorder group vs controls. I also read literature surrounding the area of research to understand what other studies have done and what variables might be of interest for our study as well.

To carry this out, I utilized SAS to look at the unadjusted associations between the substance use disorder group vs controls for variables of interest as specified by the researchers. The procedures I used in SAS include proc contents, proc sql, proc univariate, proc freq, proc sort, proc means, and proc t-test. I then utilized R to combine datasets, manipulate the data, and create a presentable table one and two for publication using the package "CreateTableOne" and "KableExtra". Once descriptive tables were created, I built three different models in R to assess the impact of substance use disorder on death, length of stay in the hospital, and receiving biologics. For the first model with the outcome of death (coded 1=yes and 0=no) I used a logistic regression. For the outcome of length of stay in the hospital, since it is a count variable and not normally distributed, I first fit the data with a poisson regression. However, I found that there was overdispersion so negative binomial was used as the distribution for length of stay in hospital. Finally, for the outcome of biologics (coded 1=yes and 0=no) I again used a logistic regression. There was a difficulty at first choosing out of the 427 variables available which to include in these models. I dealt with this problem by speaking with the researchers to better understand what variables may be clinically relevant for each of the outcomes. I also ran univariable analyses with the outcome of interest and all variables that were included in tables one and two. Associations that had a p-value of 0.2 or less in univariable analyses were considered in the models. Once all potential variables were included in the models, manual backwards elimination was utilized. The lowest AIC assisted in choosing the best model. For the outcome of death, since there were only 36 events in total and 9 of those in the substance use disorder group, there was a limitation to the number of predictor variables included in the model

in order to not overfit the model. I also found that there were some variables with missing values so I analyzed whether the missingness created any bias. Variables that had a lot of missing observations ended up not being used in the models.

Once I completed the majority of the analysis for the impact of substance use disorder on [DISEASE X] outcomes, I had the opportunity to contribute to writing the statistical methods in the paper that will potentially be published. I also began to look at the impact of obstructive sleep apnea on [DISEASE X] outcomes. The data for this was taken from the same dataset as substance use disorder and many of the variables of interest were the same, so similar code was used. I am still in the process of tweaking this table and translating findings to the researchers. All tables that I have created for the substance use disorder project have been included in the appendix.

This experience has not only helped my coding skills in R, but has challenged me to deal with a large number of variables and being able to communicate statistical terminology and findings to people with little to no statistical background. It has also brought together many different concepts from the courses I have taken in the applied biostatistics program. Instead of specific instructions to use methods for a course project, I have the responsibility to use my best judgement on what statistical analysis should be applied to this dataset and what would be best for the researcher's goals. I have been challenged to determine which variables are statistically and clinically relevant, whether missing data is at random or not, checking assumptions of models, and investigating different generalized linear models. I have had many great conversations with [Mentor's Name] about modeling such confounding variables, collinearity, interactions, and remembering to keep in mind the clinical reasons for decisions we make as statisticians, not just completely relying on the numbers. Not only have I applied my knowledge and skills from my coursework, but I have been able to better understand the role of biostatistician in medical research. The job of a statistician is not just about manipulating data and plugging in variables into models. I learned that it is important to be able to communicate why you are performing certain methods and how to translate your findings. I found that it is also crucial to be able to work in a collaborative team setting across disciplines and never hesitate to ask questions.

I am grateful for this opportunity that has taught me many lessons, solidified my understanding of statistical methods, connected me to mentors, and has given me a better understanding as to what a biostatistician role in research may look like.

Appendix

Table 1. Characteristics of Patients by Substance Use Disorder Status (content deleted)

Table 2. Patient Disease summary (content deleted)

Table 3. Patient Outcomes by Substance Use Disorder Status (content deleted)

Table 4. Multivariable logistic regression model for the outcome of death (content deleted)

Table 5. Multivariable negative binomial model for the outcome of length of stay in hospital (content deleted)

Table 6. Multivariable logistic regression model for outcome of receiving biologics (content deleted)

Code and output

Section

6

Course Descriptions

MS IN APPLIED BIOSTATISTICS COURSE DESCRIPTIONS

Comprehensive and up to date course descriptions can be found on the [BU Bulletin](#)

BS800 - Accelerated Statistics Training

Graduate Prerequisites: One year of college-level calculus, including multivariable calculus, and linear algebra to cover matrix operations, matrix functions, and singular value decomposition. - This course is designed for the MS in Applied Biostatistics program and will cover concepts of descriptive statistics and exploratory data analysis, measures of association in epidemiological studies, probability, statistical inference and computing in R and SAS. It is intended to equip students enrolling in the MS in Applied Biostatistics program with sufficient probability, statistics and computing background to enter 800 levels courses and finish the MS program within a year. The course will be offered during 2 weeks preceding the Fall semester and will involve 10 day-long modules. Modules will generally run from 10am to 5pm, combining a traditional lecture (10am to 12pm), a practice session in which students will practice the notions learned in class through exercises (1pm to 2:30pm), and a computer lab (3pm to 5pm) in which the students will learn basic computing in R and SAS and also apply the notions learned in class to real data. Allowing a student to waive this course is at the discretion of the MS in Applied Biostatistics program directors. [2 cr.]

SPH PH 700 – Foundations of Public Health

Addressing current and emerging public health challenges requires a diverse skill set and effective collaboration across multiple disciplines. This online, asynchronous course is designed to provide students with foundational knowledge of the field, from core scientific principles to the factors influencing public health. This required academic component (0 credits) meets the foundational knowledge criteria set by CEPH for all MPH and DrPH students. [0 cr.]

SPH PH746 - Professional Development Course (Career P.R.E.P)

This career development course is made up of 14 weeks, each 90 minutes long, designed to give you the tools and techniques to effectively market yourself during the job search process and advance in your career. It will also enable you to research potential career options and to manage your job searches and careers as proactively and effectively as possible. [0 cr.]

SPH BS 728 - Public Health Surveillance, a Methods Based Approach

Thacker wrote, "Surveillance is the cornerstone of public health practice." This course will provide an introduction to surveillance and explore its connections to biostatistics and public health practice. Topics will include complex survey design, weighted sampling, capture-recapture methods, time series analyses and basic spatial analyses. Students will learn about available surveillance data, how to analyze these data, and how to write about their findings. This class carries Epidemiology concentration credit. [2 cr.]

SPH BS 803 – Statistical Programming for Biostatisticians

This course will focus on skills required for advanced computing applications in biostatistics. Students will learn statistical programming and methods such as loops, functions, macros as well as data visualization techniques in SAS and R. Furthermore, the course will provide an introduction to Linux and basic statistical programming in Python. Lab sessions S will also provide students with basic computing skills to enroll to more advanced statistical classes such as BS830 and BS857. [2 cr.]

SPH BS 806 – Statistical Learning with Applications in R

This course will focus on skills required for effective conduct of data analysis with statistical packages, primary with R. This course will focus on the multiple regression modeling and multivariate analysis to cover multi-way ANOVA, multiple linear regression, classification and regression trees, automated model search, model fit and diagnostic, an multivariate analysis (PCA and cluster analysis) with particular emphasis on applications in medicine and public health. [4 cr.]

SPH BS 831 – Statistical Methods and Applications for Genomics

The goal of this course is for the students to develop a good understanding and hands-on skills in the design and analysis of 'omics' data from microarray and high-throughput sequencing experiments, including data collection and management, statistical techniques for the identification of genes that have differential expression in different biological conditions, development of prognostic and diagnostic models for molecular classification, and the identification of new disease taxonomies based on their molecular profile. These topics will be taught using real examples, extensively documented hands-on work, class discussion and critical reading. Students will be asked to analyze real gene expression data sets in their homework and final project. Principles of reproducible research will be emphasized, and students will become proficient in the use of the statistical language R (an advanced beginners knowledge of the language is expected of the students entering the class) and associated packages (including Bioconductor), and in the use of R markdown (and/or electronic notebooks) for the redaction of analysis reports. [2 cr.] -

SPH BS 849 – Bayesian Modeling for Biomedical Research & Public Health

The purpose of this course is to present Bayesian modeling techniques in a variety of data analysis applications, including both hypothesis and data driven modeling. The course will start with an overview of Bayesian principles through simple statistical models that will be used to introduce the concept of marginal and conditional independence, graphical modeling and stochastic computations. The course will proceed with the description of advanced Bayesian methods for estimation of odds and risk in observational studies, multiple regression modeling, loglinear and logistic regression, hierarchical models, and latent class modeling including hidden Markov models and application to model-based clustering. Applications from genetics, genomics, and observational studies will be included. These topics will be taught using real examples, class discussion and critical reading. Students will be asked to analyze real data sets in their homework and final paper. [2 cr.]

SPH BS 851 - Applied Statistics in Clinical Trials I

This is an intermediate statistics course, focused on statistical issues applicable to analyzing efficacy data for clinical trials. Topics include design and analysis considerations for clinical trials, such as randomization and sample size determination, and the application of statistical methods such as analysis of variance, logistic regression and survival analysis to superiority and non-inferiority clinical trials. This course includes lectures and computer instructions. Upon completion of the course, the student will be able to have a working knowledge of how to collect and manage clinical trial data; will be to analyze continuous, dichotomous, and time-to-event clinical trial data; and will be able to contribute to the statistical portions of a clinical trial study design. The student will also gain the overall knowledge required to interpret clinical trial statistical results. [4 cr.]

SPH BS 852 – Biostatistical Methods for Observational Studies

This course covers study design and intermediate-level data analysis techniques for handling confounding in epidemiologic studies. Confounding is carefully defined and distinguished from interaction. Course content covers stratification and multivariable techniques for controlling confounding in both matched and independent sample study designs, including analysis of covariance, logistic regression, and proportional hazards models. Model fit and prediction are discussed. Students are required to apply these methods with the aid of computerized statistical packages. The course will use statistical software R and SAS. [4 cr.]

SPH BS 853 - Generalized Linear Models with Applications

This course introduces statistical models for the analysis of quantitative and qualitative data, of the types usually encountered in health science research. The statistical models discussed include: Logistic regression for binary and binomial data, Nominal and Ordinal Multinomial logistic regression for multinomial data, Poisson regression for count data, and Gamma regression for data with constant coefficient of variation. All of these models are covered as special cases of the Generalized Linear Statistical Model, which provides an overarching statistical framework for these models. We will also introduce Generalized Estimating Equations (GEE) as an extension to the generalized models to the case of repeated measures data. The course emphasizes practical applications, making extensive use of SAS for data analysis. [4 cr.]

SPH BS 857 - Analysis of Correlated Data

The purpose of this advanced seminar is to present some of the modern methods for analyzing tricorrelated observations. Such data may arise in longitudinal studies where repeated observations are collected on study subjects or in studies in which there is a natural clustering of observations, such as a multi-center study of observations clustered within families. Students start with a review of methods for repeated measures analysis of variance and proceed to more complicated study designs. The course presents both likelihood-based methods and quasi-likelihood methods. Marginal, random effects and transition models are discussed. Students apply these methods in homework assignments and a project. [4 cr.]

SPH BS 858 - Statistical Genetics I

This course covers a variety of statistical applications to human genetic data, including collection and data management of genetic and family history information, and statistical techniques used to identify genes contributing to disease and quantitative traits in humans. Specific topics include basic population genetics, linkage analysis and genetic association analyses with related and unrelated individuals.

[4 cr]

Practical Training

Students will begin looking for a 400-hour practical training experience with the Career & Practicum Office usually in January. Once a practical training experience is secured, students will need to enroll in BS 910. Students will email a job description/summary of practical training experience to the Co-Directors with a signature form signed by the student and the mentor. The Co-Directors will then approve the experience and sign an [add/drop form](#) which the student will need to email to the SPH Registrar no later than June 15th. Students should also fill out the [Google Form](#) for their practical training experience.

At roughly the half-way point of the practical training experience (200 hours completed), students should submit a mid-point evaluation report that lists the work done, the work still to be completed, and an experience summary with a signature page completed by the student and mentor. This mid-term report should be submitted to the Co-Directors and the APA.

After the completion of the 400 hours of Practical Training, a final report must be submitted by August 15th along with a signature page signed by the student, the mentor, and at least one of the Co-Directors and submitted to the Co-Directors and APA.

BS910 - Practical Training

- Completion of a minimum of 400 hours (for instance: 40 hours per week for at least 10 weeks) of practical training will be required to obtain the MS in Applied Biostatistics degree. Students are expected to register for this practical training during the summer term, as a final requirement for degree completion. Practical training can be based on extension of the research rotations, industry-based internships or employment in the field of biostatistics. Students are required to write a midterm report and a research paper based on the practical training that must be submitted by August 15th to complete the degree a completed signature sheet signed by the student, the mentor, and at least one Co-Director must be submitted as well.

Appendix 5a: Practical Training Report Requirements

Practical Training Midterm Report Template:

To be completed at roughly 200 hours completed of the practical training experience

Student Name:

Mentor Name:

Mentor Contact Information:

Goals:

Aims that have been completed:

Experience summary (good experience or challenges encountered):

MSAB Practical Training Signature Form

Student Name:

Supervisor Name:

Supervisor Organization, Title, & Contact Info:

Date:

Practical Training Start Date:

Research Experience:

☐ Genetics

☐ Bioinformatics

☐ Clinical Trials/Clinical Research

☐ Observational Studies

☐ Other (please specify):

X

Student Signature

Date:

X

Supervisor Signature

Date:

X

Director's Signature

Date:

BS 910 Practical Training Report Requirements 2027

In order to complete your Practical Training Requirement to graduate by August, you will need to submit a report outlining your 400 hours of work. Please include the following in your final paper:

1. A cover page with title and name
2. A section titled “Introduction” that describes the setup of your internship
3. A section titled “Statistical Methods” that includes a detailed description of the statistical methods/analysis you used during your work
4. A section titled “Results” that describes the results of the analysis you conducted (figures are a great addition here)
5. A section titled “Discussion” that includes your thoughts about the internship experience, and some discussion of the results from your analysis.

Please include figures and tables as needed to summarize the results of your work. Please also attach your coding output as an appendix.

Example report on the following page.

MSAB Practical Training Signature Form

Student Name:

Supervisor Name:

Supervisor Organization, Title, & Contact Info:

Date:

Practical Training Start Date:

Research Experience:

☐ Genetics

☐ Bioinformatics

☐ Clinical Trials/Clinical Research

☐ Observational Studies

☐ Other (please specify):

X

Student Signature

Date:

X

Supervisor Signature

Date:

X

Director's Signature

Date:

Appendix 5b – Final Practical Training Report (an example)

Cover page:

MS Applied Biostatistics Practical Training Experience

Student Name

Date

Introduction:

For my internship experience, I worked on a new project at the Biostatistics and Epidemiology Data Analytics Center, where I am currently employed as a Statistical Programmer, that was beyond the scope of my normal responsibilities. This internship experience began in January 2021, and will run through May 2022. The client for this project was Upstream, a nonprofit organization that is focused on improving contraceptive care in the United States. For this project, I analyzed data from a virtual survey that Upstream distributed to patients of participating health centers. As a Statistical Programmer, my typical role on projects is data management and data cleaning. This experience gave me a chance to take on an analytical role.

Statistical Methods:

I cleaned data for the survey by excluding participants based on completion and eligibility criteria, combining each set of checkbox answers (initially represented as separate dichotomous variables) into a single multi-level variable, and recoding values based on open response answers where appropriate. I also derived necessary variables (such as a combined race and ethnicity variable).

The primary analysis for this experience was a validity and reliability analysis of the “Person Centered Contraceptive Care” (PCCC) measure, which is composed of four questions that were included in the patient survey. This measure is a short-form version of the Interpersonal Quality of Family Planning scale. Each PCCC item was scored from 1 to 5, with higher values indicating greater patient satisfaction. The overall PCCC outcome was represented by a dichotomous top-box score, which was 1 if the patient gave the highest score (“excellent”) for each item and 0 otherwise. I tested internal reliability using Cronbach’s alpha and item-total correlations. I used factor loadings to assess construct validity. Additionally, I analyzed logistic regression models using generalized estimating equations (GEE) to assess convergent validity between outcome PCCC top-box score and conceptually related metrics while accounting for clustering by agency and controlling for demographics.

Results:

Multiple psychometric indicators provided evidence of reliability and validity. A high Cronbach’s alpha value (0.936) and strong item-total correlations (all >0.8) demonstrated internal reliability. High factor

loadings (all ≥ 0.9) and a consistent single-factor analysis solution demonstrated construct validity. Significant associations between PCCC top-box score and conceptually related survey questions provided further evidence of construct validity. In a logistic regression GEE model, race/ethnicity and income were each significantly associated with PCCC top-box score after accounting for remaining demographic covariates and clustering by agency. Higher income was associated with greater odds of reporting the highest PCCC score. Relative to patients who were White alone, patients who were Hispanic or Latinx, Asian alone, or Black or African American alone each had significantly lower odds of reporting the highest PCCC score. Three additional models- constructed by adding voluntarism, shared decision making, and Method Information Index (MII) variables to the demographic model- each showed that these conceptually related constructs were significantly associated with PCCC top-box score. In the shared decision making model, those who reported deciding on birth control with clinic staff “together” and those who reported “mostly clinic staff decided” had significantly higher and lower odds of reporting the highest PCCC score respectively relative to those who decided “on their own”. In the voluntarism model, both those who experienced pressure from health center staff to choose to use or continue a birth control method and those who did not know whether they experienced such pressure had significantly lower odds of reporting the highest PCCC score relative to those who did not experience this pressure. In the MII model, patients whose provider conveyed information about contraceptive options and side effects had significantly higher odds of reporting the highest PCCC score relative to those whose provider did not convey all of this information.

Discussion:

The Upstream team and I are currently finalizing a “Brief Report” to submit to the Journal of Contraception that focuses on the reliability and validity analyses I conducted. Since the PCCC is a relatively new measure (published in 2021), this article would provide early support for its use in practice.

This experience has given me an opportunity to perform an independent analysis. This is especially helpful for my career goals, as I recently applied for a promotion and was able to reference this work to support my ability to perform a more analytic role. I also learned a lot about the concepts of validity and reliability as well as the methods used for their assessment.

In the Appendix I have included a descriptive table with details about the survey sample as well as output from the logistic regression GEE models I used for analysis.

Appendix:

Table 1. Survey Sample Demographics

Characteristic	n (%)
Language	
Arabic	4 (0.28%)
English	1282 (90.73%)
French	2 (0.14%)
Portuguese	24 (1.7%)
Spanish	100 (7.08%)
Vietnamese	1 (0.07%)

Race/Ethnicity	
Hispanic or Latinx	436 (34.01%)
Asian alone	84 (6.55%)
Black or African American alone	117 (9.13%)
White alone	582 (45.4%)
All other races/ethnicities	63 (4.91%)
Missing	131

Age	
2 = 18-19 years old	61 (4.32%)
3 = 20-24 years old	290 (20.52%)
4 = 25-29 years old	416 (29.44%)
5 = 30-34 years old	294 (20.81%)
6 = 35-39 years old	221 (15.64%)
7 = 40-44 years old	131 (9.27%)
Household size	
1	170 (13.03%)
2	378 (28.97%)
3	289 (22.15%)
4	256 (19.62%)
5 or more	212 (16.25%)
Missing	108
Adults in household	
1	186 (22.33%)
2	395 (47.42%)
3 or more	252 (30.25%)
Missing	580
Living children	
1	521 (62.85%)
2	124 (14.96%)
3 or more	184 (22.2%)
Missing	584
Annual household income	
Under \$25,000	330 (25.06%)
Over or equal to \$25,000	692 (52.54%)
I prefer not to answer	110 (8.35%)
I do not know	185 (14.05%)
Missing	96

Model 1. Demographics

	OR	CI	P
Race/Ethnicity			
Hispanic or Latinx	0.88	(0.8, 0.96)	0.007**
Asian alone	0.78	(0.68, 0.89)	<0.001**
Black or African American alone	0.75	(0.61, 0.92)	0.005**
All other races/ethnicities	1.12	(0.99, 1.26)	0.062
White alone	Reference		
Age			
18-19 years old	1.15	(0.88, 1.5)	0.315
20-24 years old	0.997	(0.9, 1.11)	0.956
30-34 years old	0.97	(0.85, 1.09)	0.587
35-39 years old	1.09	(0.99, 1.21)	0.085
40-44 years old	1.07	(0.93, 1.23)	0.350
25-29 years old	Reference		
Income			
Under \$25,000	0.83	(0.77, 0.89)	<0.001**
Over \$25,000	Reference		
Language			
Another language	0.91	(0.76, 1.09)	0.301
Spanish	0.62	(0.38, 1.02)	0.057
English	Reference		

Model 2. Demographics and Shared Decision Making (selected output)

	OR	CI	P
Shared Decision Making			
Clinic staff and I decided together	1.13	(1.05, 1.22)	0.001**
Clinic staff decided	0.91	(0.38, 2.2)	0.834
Mostly I decided	0.98	(0.87, 1.11)	0.757
Mostly clinic staff decided	0.5	(0.27, 0.9)	0.021*
I decided	Reference		

Model 3. Demographics and Voluntarism (selected output)

	OR	CI	P
Voluntarism			
Not sure	0.61	(0.42, 0.9)	0.011*
Yes	0.25	(0.13, 0.51)	<0.001**

No	Reference	
----	-----------	--

Model 4. Demographics and MII (selected output)

Model 4: Demographics and MII			
	OR	CI	P
MII			
All yes	1.17	(1.07, 1.28)	0.001**
Not all yes	Reference		

Code Output:

Section

7

Biostatistics Faculty

The Department of Biostatistics faculty are committed to the roles of teacher and mentor both inside and outside of the classroom. Their research brings depth and a real-life context to the classroom. The department's faculty has analyzed the multigenerational risk factors that contribute to heart disease, which led to a predictive tool for physicians to determine treatment strategies for patients with cardiovascular disease. In partnership with other academic institutions, members of the department have isolated and identified key factors that contribute to higher incidences of breast cancer and other diseases in African-American women. In addition, the department has designed and implemented an important comparative study that pinpoints risk factors for Alzheimer's disease. Members of the department are also actively engaged in clinical trials and methods for public health surveillance, designed to improve clinical treatments and aid the public health delivery system to identify disease hotspots. Faculty bios can be found [here](#) and information about faculty research can be found on the [Research](#) section of our website.

Biostatistics MS Degree Audit Sheet (34 credits)

I. AST: 3 week course (2 credits)	Semester Completed	Grade Earned	Credits Earned
SPH BS800: Accelerated Statistics Training			
II. MS Degree Required Courses: (32 credits)	Semester Completed	Grade Earned	Credits Earned
SPH PH 700 Foundations of Public Health			
SPH PH746 Professional Development Course			
SPH BS728 Public Health Surveillance			
SPH BS803 Statistical Programming for Biostatisticians			
SPH BS806 Statistical Learning with Applications in R			
SPH BS831 Statistical Methods and Applications for Genomics			
SPH BS849 Bayesian Modeling			
SPH BS851 Applied Statistics in Clinical Trials			
SPH BS852 Biostatistical Methods for Observational Studies			
SPH BS853 Generalized Linear Models			
SPH BS857 Analysis of Correlated Data			
SPH BS858 Statistical Genetics			
III. Practical Training	Date of Completion		Grade Earned
Research Rotation in Biostatistics (100 hours)			
BS910: Practical Training (400 hours)			

MS Requirements Checklist:

- ☐ Credits total: 32 MS graduate credits or approved transfer courses.
- ☐ Grade of B- or better in all courses applied to the MS
- ☐ All incomplete classes completed and grades posted
- ☐ Submitted [graduation application](#) to SPH Registrar by the deadline (usually in March)