HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Department of Biostatistics

2020-2021

Graduate Student Handbook

MASTER OF SCIENCE IN HEALTH DATA SCIENCE

Contents

1 INTRODUCTION

This handbook describes the Master's Degree Program in Health Data Science offered by the Department of Biostatistics at Harvard University. The SM program prepares graduates to handle "Big Data" in addressing the biomedical research questions that are becoming increasingly commonplace in hospitals and universities, research organizations, and the pharmaceutical and biotechnology industries. It can also provide the foundation for further doctoral studies.

The sections of this handbook include information and regulations concerning entrance requirements, program descriptions, degree requirements, and other program policies. Policies and official requirements of the School of Public Health are set forth in the *Harvard T. H. Chan School of Public Health Student Handbook* (https://www.hsph.harvard.edu/student-handbook/). Each graduate student is responsible for general knowledge of, and adherence to, the policies and requirements of the degree program in which the student is enrolled. Additional program information is available at the website https://www.hsph.harvard.edu/health-data-science/program/. Vitally important for our community is that all members demonstrate respect for each other and our discipline. For all members of the community, respect is demonstrated by attending all scheduled classes or meetings, and arriving on time, fully prepared, and ready to participate.

This handbook was prepared by the Executive Committee of the Program in Health Data Science. Dr. Heather Mattie is responsible for reviewing the student's program of study, and has the authority to consider exceptions to the rules and regulations established by the Executive Committee. The Program Directors welcome suggestions and comments.

PROGRAM LEADERSHIP AND ADMINISTRATION

Executive Committee

Dr. JP Onnela
Program Co-Director
Email: onnela@hsph.harvard.edu

Dr. Heather Mattie
Program Co-Director
Phone: 617-432-5308
Email: hemattie@hsph.harvard.edu

Dr. Christoph Lange
Dr. Jeffrey Miller
Dr. Brent Coull
Dr. Tamar Sofer

Administration

Jelena Follweiler
Senior Manager of Academic Services, Biostatistics
Phone: 617-432-1087
Email: jtillots@hsph.harvard.edu

Geryel Osorio Godoy
Graduate Program Coordinator, Biostatistics
Phone: 617-432-5061
Email: osoriogodoy@hsph.harvard.edu

Career Development

Dr. Erin Lake
Director of Student Development, Biostatistics
Co-Director, Master's Program in Biostatistics
Email: eklake@hsph.harvard.edu

Websites

Master of Science in Health Data Science
Department of Biostatistics

2 Master of Science in Health Data Science

2.1 Core Competencies

This 60-credit program, offered by the Department of Biostatistics, is designed to provide students with targeted skills and knowledge required for work in health data science. These specific skills and knowledge domains are:

- Apply experimental design in a research project
- Implement data wrangling by manually mapping data from one form to another
- Perform data visualization/communication
- Critically analyze and interpret data science
- Appropriately apply statistical inference/probability to make scientific conclusions from data
- Understand and employ linear models, regression and matrix algebra
- Apply methods for high-dimensional data
- Apply and interpret practical machine learning
- Be able to engineer software, including reproducible research
- Be proficient with high performance scientific computing

The SM in Health Data Science is intended as a terminal professional degree which will enable students to launch their careers in health-related data science. It can also provide the foundation for further doctoral studies in biostatistics or other quantitative or computational sciences with an emphasis in data science.

Students will receive training in quantitative methods, including applied regression, statistical inference, statistical computing, machine learning, statistical consultation and collaboration, and epidemiology.

2.2 Admissions Procedures and Requirements

2.2.1 Harvard T.H. Chan School of Public Health Requirements

Application for admission to the SM program is available online on the Admissions Office website (https://www.hsph.harvard.edu/admissions/admissions/how-to-apply/application-requirements/). For information on general requirements for admission, contact the Admissions Office by phone (617-432-1031) or through their website (https://www.hsph.harvard.edu/admissions/).

2.2.2 Program Requirements

All candidates for admission to the SM in Health Data Science program must have the following:

- An undergraduate degree in the mathematical sciences or allied fields (statistics, economics, etc.) or computer science, with a strong interest in health science,
- practical knowledge of computer scripting and programming, as well as experience with a statistical computing language such as R or Python,
- calculus through multivariable integration,
- one semester of linear algebra or matrix methods, and
- excellent written and spoken English.

Additional research or work experience would be considered beneficial, but not required.

Evidence that these requirements have been fulfilled should form part of the application.

2.3 Intra/Inter-departmental Biostatistics Degree Program Switch Protocol

The Department of Biostatistics, and several other academic departments at the School (such as Epidemiology), offer master's degree programs. These programs each have their own goals and requirements and make independent decisions about admissions. Students should carefully choose the program to which they apply and we expect students admitted to a program to meet the requirements of that program.

However, we recognize that the interests of some students may change during their time in graduate school. Therefore, the Department has established the following procedure for students applying for a change in program. This applies to students applying for a switch within the Department of Biostatistics, or between the Department of Biostatistics and another department at the School, such as Epidemiology. Please note that transfers between programs are not automatic and may not be approved, and if a student has received a scholarship or other funds from a degree program, that funding will not transfer to the new degree.

1. Students must complete at least one full semester of coursework before applying for a program transfer.
2. Students must enroll in and successfully pass any required coursework for their current program before beginning the transfer process. Please see your current degree program handbook for a list of required coursework. Students may also need to complete coursework in their proposed program so as to not fall behind in requirements to finish their program on time if approved for transfer.
3. Students must complete and submit a formal application and include an updated statement of purpose describing the reason for seeking a program transfer, current CV and one letter of recommendation.
4. The Directors of department master's programs, including the Directors of the student's current program and prospective program, will schedule an interview with the students to assess their application. At the meeting, students will provide a copy of their CV and statement of purpose to the program director(s), and will briefly explain their decision to apply for a transfer. The Directors will consider the applications and render a decision whether to approve or deny the transfer. These meetings will be scheduled in early January for students applying for transfer to start in the Spring semester, and in late May for students applying to transfer after the Spring semester.
5. The Directors will discuss and notify the student of their decision within three days. All decisions are final.
6. If approved, completed paperwork must be submitted to the Registrar's Office, and then to the Senior Manager of Academic Services.

Note that all program transfers are at the discretion of the program Directors and are not guaranteed. If approved, students will be assigned a new academic advisor affiliated with the chosen degree program. It is the responsibility of the students to ensure that they complete all requirements for their degree program.

2.4 Curricular Practical Training (CPT) Approval for Students with F-1 Visas

To be considered CPT, the work must not only be related to the student's major field of study but must also be an integral part of an established curriculum. Before seeking off-campus internship opportunities, students are required to discuss their plans with Elizabeth Capuano (see contact info below) from the Harvard International Office to determine their CPT eligibility. Please note that CPT eligibility may be impacted by the March 2020 guidance issued by the Department of Homeland Security - Student Exchange Visitor Program.

There are two ways in which students are eligible for CPT:

1. Employment that is a required part of a degree program, such as a required internship or practicum. This requirement must be formally documented in school publications, such as a student handbook.

2. Employment that is not required by a degree program, but for which a program will award academic credits. This could include training courses such as a field studies course, an independent study course that is based on an internship.

Our program does not offer the first option above. However, we do allow the second option with approval from either of the program directors. International students who wish to pursue this option **MUST** speak with Elizabeth Capuano elizabeth_capuano@harvard.edu, our representative at the Harvard International Office (HIO), **before beginning interview processes** at prospective internship sites to discuss the requirements for CPT authorization. Students should also speak to one of the directors about whether their employment would qualify for academic credits, as the academic credits are required for CPT authorization. If CPT eligibility is established, please note that students **MUST** obtain CPT authorization **PRIOR** to beginning the internship.

The most up-to-date information about CPT will be found here:
http://www.hio.harvard.edu/curricular-practical-training-cpt.

2.5 Advising and Degree Program Approval

2.5.1 Academic Advisor

All entering students are assigned an academic advisor to help plan course loads and explain program requirements. At the earliest possible date, the student and the academic advisor will develop a program of study. Should a student wish to change his/her academic advisor, they are encouraged to discuss this with Dr. Mattie. In addition, SPH provides services for all students with clinically documented learning and/or physical disabilities.

2.5.2 Departmental Approval of Program

The Master's degree program plan must be submitted to the student's academic advisor and Dr. Mattie, for approval by the Executive Committee, using the Master's Degree Program form provided. This plan should be submitted at least one semester prior to your expected graduation date.

2.5.3 Epidemiology Requirement

The School of Public Health requires that Master's students successfully pass one epidemiology course. The program requires that EPI 201 (Fall 1, 2.5 credits) be taken to satisfy this requirement in the first semester of the degree program.

2.5.4 Public Health Practice Requirement

Students may be required to take a public health course by the School of Public Health as part of their accreditation requirements. These requirements will be communicated to all incoming students by the School of Public Health directly.

2.5.5 Research Ethics Requirement

Students must satisfy a research ethics requirement by completing a course in responsible conduct of research or by completing an online training course during the first year in the program. Students who feel they have already completed an equivalent training program must submit adequate documentation to, and receive approval from, Dr. Mattie during the first semester in residence.

2.6 Satisfactory Progress Requirements

For students in the program, a total of 60 credits are required with a minimum of 55 ordinal credits from the core courses, computer science courses, and electives listed in Section 2.7.1, the epidemiology requirement in Section 2.5.3, and the Health Data Science Practice in Section 2.7.2. In addition, SPH students must remain in good academic standing, must complete program requirements within the designated time to degree, and must maintain a cumulative grade point average of 2.70 or above. All ordinal grades for courses used to satisfy program requirements specified in Section 2.7.1 and Section 2.7.2 must be at the level of B- or higher. Courses taken on a pass/fail basis cannot be used to satisfy ordinally graded program requirements.

A detailed presentation of SPH's regulations for Master's students is found at https://www.hsph.harvard.edu/student-handbook/. All Master's students and their advisors should make sure that SPH and HDS program requirements are met according to schedule.

2.7 Degree Requirements

A total of 60 credits are required for the SM60 in Health Data Science. Students with prior equivalent background to any of the required courses or strong reasons to take a different course can request permission from Dr. Mattie, for approval by the Executive Committee, for a substitution of one or more of the required courses. Students should submit a degree program form to Dr. Mattie at least one semester prior to their expected graduation date.

2.7.1 Course Requirements for the Health Data Science SM60 Degree

The degree requirements include a 20 credit ordinally graded core curriculum consisting of:

BST	222	Basics of Statistical Inference (Fall, 5 credits)
BST	260	Introduction to Data Science (Fall, 5 credits)
BST	261	Data Science II (Spring 2, 2.5 credits)
BST	262	Computing for Big Data (Fall 2, 2.5 credits)
BST	263	Statistical Learning (Spring, 5 credits)

An additional five credits must be taken in computer science from the following list:

BST	234.1	Introduction to Data Structures and Algorithms [Part 1] (2.5 credits)
BST	234.2	Introduction to Data Structures and Algorithms [Part 2] (2.5 credits)
BST	249	Bayesian Methodology in Biostatistics (5 credits)
or	STAT 220	Bayesian Data Analysis (5 credits)
BST	281	Genomic Data Manipulation (5 credits)
BST	282	Introduction to Computational Biology and Bioinformatics (5 credits)
APMTH	120	Applied Linear Algebra and Big Data (5 credits)
APMTH	207	Advanced Scientific Computing: Stochastic Methods for Data Analysis, Inference and Optimization (5 credits)
BMI	713	Computational Statistics for Biomedical Science (5 credits)
CS	105	Privacy and Technology (5 credits)
CS	124	Data Structures and Algorithms (5 credits)
CS	134	Networks (5 credits)
CS	164	Software Engineering Computer Science (5 credits)
CS	165	Data Systems (5 credits)
CS	171	Visualization (5 credits)
CS	182	Artificial Intelligence (5 credits)
CS	187	Computational Linguistics (5 credits)
CS	205	Computing Foundations (5 credits)
CS	207	Systems Development for Computational Science (5 credits)
STAT	171	Introduction to Stochastic Processes (5 credits)
MIT	6.869	Advances in Computer Vision (5 credits)

Twenty-five additional credits must be taken. Courses that would satisfy these requirements may come from the following list of elective courses:

BST	210	Applied Regression Analysis (5 credits)
BST	214	Principles of Clinical Trials (2.5 credits)
BST	216	Introduction to Quantitative Methods for Monitoring and Evaluation (2.5 credits)
BST	223	Applied Survival Analysis (5 credits)
BST	226	Applied Longitudinal Analysis (5 credits)
BST	227	Introduction to Statistical Genetics (2.5 credits)
BST	228	Applied Bayesian Analysis (5 credits)
BST	267	Introduction to Social and Biological Networks (2.5 credits)
BST	280	Introductory Genomics & Bioinformatics for Health Research (2.5 credits)
BST	283	Cancer Genome Data Science (5 credits)
EPI	202	Elements of Epidemiologic Research: Methods 2 (2.5 credits)
EPI	203	Study Design in Epidemiologic Research (2.5 credits)
EPI	204	Analysis of Case-Control and Cohort Studies (2.5 credits)
EPI	233	Research Synthesis & Meta-Analysis (2.5 credits)
EPI	271	Propensity Score Analysis (1.25 credits)
EPI	286	Database Analytics in Pharmacoepidemiology (2.5 credits)
EPI	288	Introduction to Machine Learning and Risk Prediction (2.5 credits)
EPI	293	Analysis of Genetic Association Studies (2.5 credits)
EPI	511	Advanced Population and Medical Genetics (5 credits)
ID	271	Advanced Regression for Environmental Epidemiology (2.5 credits)
RDS	280	Decision Analysis for Health and Medical Practices (2.5 credits)
RDS	282	Economic Evaluation of Health Policy and Program Management (2.5 credits)
RDS	285	Decision Analysis Methods in Public Health and Medicine (2.5 credits)
APCOMP	221	Critical Thinking in Data Science (5 credits)
APMTH	221	Advanced Optimization (5 credits)
BMI	701	Introduction to Biomedical Informatics (5 credits)
BMI	702	Foundation of Biomedical Informatics II (2.5 credits)
BMI	703	Precision Medicine I: Genomic Medicine (2.5 credits)
BMI	704	Data Science for Medical Decisions (2.5 credits)
BMI	705	Precision Medicine II: Integrating Clinical and Genomic Data (2.5 credits)
BMI	706	Data Visualization for Biomedical Applications (2.5 credits)
CI	722.0	Clinical Data Science: Comparative Effectiveness Research I (2.5 credits)
CS	109A	Data Science 1: Introduction to Data Science (5 credits)
CS	109B	Data Science 2: Advanced Topics in Data Science (5 credits)
HST	.953	Collaborative Data Science in Medicine (5 credits)
ME	530M.1	Clinical Informatics (5 credits)
MIT	6.864	Advanced Natural Language Processing (5 credits)
STAT	117	Data Analysis in Modern Biostatistics (5 credits)
STAT	131	Time Series & Prediction (5 credits)
STAT	260	Design and Analysis of Sample Surveys (5 credits)
SUP	135	Using Big Data to Solve Economic and Social Problems [aka ECON 1152] (5 credits)

Other courses may also be acceptable. EPI 201 (see Section 2.5.3) and HDSC 325 (see Section 2.7.2) will count as two of the 55 credit ordinal courses required. If you need cross-registration credit conversion, see: https://www.hsph.harvard.edu/r-o-student-knowledge-center/. Students are advised to consult with Dr. Mattie about any substitutions. To request a substitution, students should email Dr. Mattie with their reasons for the request (please cc the Graduate Program Coordinator, Geryel Osorio Godoy and the Senior Manager of Academic Services, Jelena Follweiler, so that we may file the information for the final degree audit).

2.7.2 Health Data Science Practice

The health data science practice will provide a culminating research experience that tests all competencies through a hands-on semester-long project-based research course (7.5 credits). This course will allow students to immerse themselves in multiple health data science projects in public health and biomedical science.

HDSC 325 Health Data Science Practice (7.5 credits)

The course will be co-taught by two faculty members with complementary expertise in statistics and computer science. The overall objective of this course is to allow students to integrate and apply various methods introduced in previous core coursework in the Health Data Science program to different real-world data sets across knowledge domains. This is a predominantly project-based course with lectures focusing on project planning and execution, career development and special advanced topics in data science.

Students will be randomly assigned to teams of 2-3 individuals. Each team will be assigned a Harvard affiliate mentor who will grant access to data at the mentor's respective institution and assign a semester-long project. Each team will be expected to work on the project outside of class and meet with their project mentor regularly. Course instructors will also regularly check-in with students and project mentors to ensure appropriate project expectations and progress. Teams will also be responsible for 2 presentations on the progress of the project before the final presentation of the project on the last day of class. Teams will be expected to complete a project analysis plan, project abstract and final project report including all code and visualizations.

3 ADMINISTRATIVE TIMELINE

Detailed requirements and deadlines for degree completion are given on the Harvard T.H. Chan School of Public Health webpage. All forms linked below are also located on the last page of this Graduate Student Handbook.

- **Summer Before Entering Program**
  - ☐ Take optional R/Python Programming Bootcamp/EdX online course.
- **Year One**
  - First Semester
    - ☐ Complete Research Ethics requirement by taking the free online CITI program or attending HPM 548 (See section 2.5.5). You will receive a reminder about where to find the online course before beginning the Fall term.
    - ☐ If planning to waive courses, ask for those waivers of Fall core courses by emailing Dr. Mattie **and** Jelena Follweiler with details about the course(s) taken or experience that may qualify for a waiver. To waive EPI 201 which is a school-wide core requirement, students must submit this Waiver of Core Courses Form (found at bottom of page) to the instructor teaching the course for approval, and then to the Registrar's Office.
  - Second Semester
    - ☐ Complete or waive Spring core courses.
    - ☐ Turn in the final program form by May 10.
- **Year Two**
  - Third Semester
    - ☐ Take HDSC 325 Health Data Science Practice.

4 PROGRAM FORMS

- **HDS SM60 Degree Program Form**
  https://content.sph.harvard.edu/biostats/publications/hds_handbook/SM60_Degree_Form_HDS.pdf