Contents

1 Introduction 1

2 Master of Science in Health Data Science 3
  2.1 Core Competencies ................................................. 3
  2.2 Admissions Procedures and Requirements .......................... 3
    2.2.1 Harvard T.H. Chan School of Public Health Requirements ................................................. 3
    2.2.2 Program Requirements ........................................... 3
  2.3 Intradepartmental Biostatistics Degree Program Switch Protocol ................................................. 4
  2.4 Advising and Degree Program Approval ................................ 4
    2.4.1 Academic Advisor .............................................. 4
    2.4.2 Departmental Approval of Program ................................ 4
    2.4.3 Epidemiology Requirement ....................................... 5
    2.4.4 Research Ethics Requirement ..................................... 5
  2.5 Satisfactory Progress Requirements ................................ 5
  2.6 Degree Requirements .................................................. 5
    2.6.1 Course Requirements for the Health Data Science SM60 Degree ................................................. 5
    2.6.2 Health Data Science Practice ..................................... 6

3 ADMINISTRATIVE TIMELINE 8

4 PROGRAM FORMS 8
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. The Executive Director 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 Director and the Executive Director welcome suggestions and comments.
Program Leadership and Administration

Executive Committee

Dr. JP Onnela
Program Director, Master’s Program in Health Data Science
Email: onnela@hsph.harvard.edu

Dr. Heather Mattie
Executive Director, Master’s Program in Health Data Science
Phone: 617-432-5308
Email: hemattie@hsph.harvard.edu

Dr. Christine Choirat
Email: cchoirat@gmail.com

Dr. Christoph Lange
Email: clange@hsph.harvard.edu

Dr. Jeffrey Miller
Email: jwmiller@hsph.harvard.edu

Dr. Brent Coull
Email: bcoull@hsph.harvard.edu

Administration

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

Eric Creighton
Graduate Program Coordinator, Biostatistics
Phone: 617-432-5061
Email: ecreighton@hsph.harvard.edu

Career Development

Dr. Erin Lake
Director of Student Development, 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
- Be able to 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 you to launch your career 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 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 package 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 Intradepartmental Biostatistics Degree Program Switch Protocol

The Department of Biostatistics offers a number of 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. 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 each Director, 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 with three days. All decisions are final.

6. If approved, completed paperwork must be submitted to the Registrar’s Office, and then to the 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 Advising and Degree Program Approval

2.4.1 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, he or she is encouraged to discuss this with the Executive Director. In addition, SPH provides services for all students with clinically documented learning and/or physical disabilities.

2.4.2 Departmental Approval of Program

The Master’s degree program plan must be submitted to the student’s academic advisor and the Executive Director, 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.
2.4.3 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.

2.4.4 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, the Executive Director during the first semester in residence.

2.5 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 2.6.1 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 average of 2.70 or above. All ordinal grades for courses used to satisfy program requirements specified in Section 2.6.1 must be at the level of B- or higher. Courses taken on a pass/fail basis cannot be used to satisfy ordinarily 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.

2.6 Degree Requirements

A total of 60 credits are required for the SM60 in Health Data Science. A minimum of 55 ordinal credits of coursework is required from the courses listed below. Students with prior equivalent background to any of the required courses or strong reasons to take a different course can request permission from the Executive Director, 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 the Executive Director at least one semester prior to your expected graduation date.

2.6.1 Course Requirements for the Health Data Science SM60 Degree

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

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BST 222</td>
<td>Basics of Statistical Inference (Fall, 5 credits)</td>
<td></td>
</tr>
<tr>
<td>BST 260</td>
<td>Introduction to Data Science (Fall, 5 credits)</td>
<td></td>
</tr>
<tr>
<td>BST 261</td>
<td>Data Science II (Spring 2, 2.5 credits)</td>
<td></td>
</tr>
<tr>
<td>BST 262</td>
<td>Computing for Big Data (Fall 2, 2.5 credits)</td>
<td></td>
</tr>
<tr>
<td>BST 263</td>
<td>Statistical Learning (Spring, 5 credits)</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BST 234</td>
<td>Introduction to Data Structures and Algorithms (5 credits)</td>
<td></td>
</tr>
<tr>
<td>BST 281</td>
<td>Genomic Data Manipulation (5 credits)</td>
<td></td>
</tr>
<tr>
<td>APMTH 120</td>
<td>Applied Linear Algebra and Big Data (5 credits)</td>
<td></td>
</tr>
<tr>
<td>BMI 713</td>
<td>Computational Statistics for Biomedical Science (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 105</td>
<td>Privacy and Technology (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 124</td>
<td>Data Structures and Algorithms (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 164</td>
<td>Software Engineering Computer Science (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 165</td>
<td>Data Systems (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 171</td>
<td>Visualization (5 credits)</td>
<td></td>
</tr>
<tr>
<td>CS 187</td>
<td>Computational Linguistics (5 credits)</td>
<td></td>
</tr>
<tr>
<td>STAT 171</td>
<td>Introduction to Stochastic Processes (5 credits)</td>
<td></td>
</tr>
</tbody>
</table>
Twenty-five additional credits must be taken. Courses that would satisfy these requirements may come from the following list of elective courses:

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BST 210</td>
<td>Applied Regression Analysis</td>
<td>5</td>
</tr>
<tr>
<td>BST 216</td>
<td>Introduction to Quantitative Methods for Monitoring and Evaluation</td>
<td>2.5</td>
</tr>
<tr>
<td>BST 223</td>
<td>Applied Survival Analysis</td>
<td>5</td>
</tr>
<tr>
<td>BST 226</td>
<td>Applied Longitudinal Analysis</td>
<td>5</td>
</tr>
<tr>
<td>BST 228</td>
<td>Applied Bayesian Analysis</td>
<td>5</td>
</tr>
<tr>
<td>BST 254 Sec 3</td>
<td>Measurement Error and Misclassification</td>
<td>2.5</td>
</tr>
<tr>
<td>BST 267</td>
<td>Introduction to Social and Biological Networks</td>
<td>2.5</td>
</tr>
<tr>
<td>BST 280</td>
<td>Introductory Genomics &amp; Bioinformatics for Health Research</td>
<td>2.5</td>
</tr>
<tr>
<td>BST 282</td>
<td>Introduction to Computational Biology and Bioinformatics</td>
<td>5</td>
</tr>
<tr>
<td>BST 283</td>
<td>Cancer Genome Analysis</td>
<td>5</td>
</tr>
<tr>
<td>EPI 202</td>
<td>Elements of Epidemiologic Research: Methods</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 203</td>
<td>Study Design in Epidemiologic Research</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 204</td>
<td>Analysis of Case-Control and Cohort Studies</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 233</td>
<td>Research Synthesis &amp; Meta-Analysis</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 271</td>
<td>Propensity Score Analysis</td>
<td>1.25</td>
</tr>
<tr>
<td>EPI 286</td>
<td>Database Analytics in Pharmacoepidemiology</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 288</td>
<td>Data Mining and Prediction</td>
<td>2.5</td>
</tr>
<tr>
<td>EPI 293</td>
<td>Analysis of Genetic Association Studies</td>
<td>2.5</td>
</tr>
<tr>
<td>ID 271</td>
<td>Advanced Regression for Environmental Epidemiology</td>
<td>2.5</td>
</tr>
<tr>
<td>RDS 280</td>
<td>Decision Analysis for Health and Medical Practices</td>
<td>2.5</td>
</tr>
<tr>
<td>RDS 282</td>
<td>Economic Evaluation of Health Policy and Program Management</td>
<td>2.5</td>
</tr>
<tr>
<td>RDS 285</td>
<td>Decision Analysis Methods in Public Health and Medicine</td>
<td>2.5</td>
</tr>
<tr>
<td>APMTH 207</td>
<td>Advanced Scientific Computing: Stochastic Methods for Data Analysis, Inference and Optimization</td>
<td>5</td>
</tr>
<tr>
<td>APMTH 221</td>
<td>Advanced Optimization</td>
<td>5</td>
</tr>
<tr>
<td>BMI 701</td>
<td>Introduction to Biomedical Informatics</td>
<td>5</td>
</tr>
<tr>
<td>BMI 702</td>
<td>Foundation of Biomedical Informatics II</td>
<td>2.5</td>
</tr>
<tr>
<td>BMI 703</td>
<td>Precision Medicine I: Genomic Medicine</td>
<td>2.5</td>
</tr>
<tr>
<td>BMI 705</td>
<td>Precision Medicine II: Integrating Clinical and Genomic Data</td>
<td>2.5</td>
</tr>
<tr>
<td>BMI 706</td>
<td>Data Visualization for Biomedical Applications</td>
<td>2.5</td>
</tr>
<tr>
<td>CI 722.0</td>
<td>Clinical Data Science: Comparative Effectiveness Research</td>
<td>2.5</td>
</tr>
<tr>
<td>ME 530M.1</td>
<td>Clinical Informatics</td>
<td>5</td>
</tr>
<tr>
<td>STAT 260</td>
<td>Design and Analysis of Sample Surveys</td>
<td>5</td>
</tr>
</tbody>
</table>

Other courses may also be acceptable. EPI 201 (see section 2.4.3) will count as one of the 55 credit ordinal courses required. Students are advised to consult with the Executive Director about any substitutions.

### 2.6.2 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 projects, which will include investigations from areas such as statistical genetics and genomics, computational biology, comparative effectiveness research, analysis of electronic medical records, remote sensing environmental data, and network analysis, among others, will be developed and organized by the instructors, based on current data science projects being directed by faculty in the depart-
ment. The overall objective of this course is to help the student integrate and apply data science methods presented in coursework to real world, different massive datasets. For each project, the data are likely to be messy and students need to develop their analytic methods appropriately. With faculty input, the students will write a paper that is of the same standard as publishable papers in health science journals and present the results in class. These faculty instructors (mentors) will grade the project papers and also give feedback on how to improve presentations. Students then incorporate this feedback on paper writing and presentation into the next project.

Students will learn to be aware of problems that arise in study design, data collection, data wrangling and visualization, inference and computation, and develop practical problem solving skills. The student will learn to communicate through presentation of oral and written reports, following the standard of health science journal articles for each project completed in the class, and through student and faculty critiques of these reports at both the intermediate and final stages of projects. In summary, the course represents a research-based class experience, and will allow students to generate novel findings from current big data in health.
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/Pyton 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 2.4.4). 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 the Executive Director and the Manager of Academic Studies in Biostatistics with details about the course(s) taken or experience that you have that may qualify you for a waiver. To waive EPI 201 which is a school-wide core requirement, you 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 your final program form by May 1.

• 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