HEALTH DATA SCIENCE (MS)

Visit program website. (https://epibiostat.ucsf.edu/masters-degree-health-data-science/)

Degree Offered: MS
Program Leadership:
John Kornak, PhD, Program Director
Thomas Hoffmann, PhD, MA, Associate Program Director
Admissions Inquiries:
Eva Wong-Moy, Graduate Affairs Manager

Program Description

Data science plays a fundamental role in health sciences research: Learning from data is at the core of how we make advances in health research. Data science methods and tools are needed to deal with the expanding role of precision medicine, the widespread analyses of electronic health records, and the growing number of large and complex datasets.

The Master of Science (MS) Degree in Health Data Science (MiHDaS) is a two-year program in which students learn to apply data science, biostatistics, machine learning, and epidemiological thinking in clinical research settings.

The program is intended for:

- Quantitative science learners interested in studying data science with a focus on biomedical applications.
- Numerically able biomedical scientists interested in applying data science methods in clinical, epidemiological and biological sciences.

We also offer a one-year certificate program (CiHDaS) (https://epibiostat.ucsf.edu/certificate-health-data-science/), with condensed coursework and absent teaching and hands on capstone project experience, best suited for those already working in the biomedical or pharmaceutical industries.

Admission Requirements

- Bachelor's degree (BA/BS) or the equivalent from an accredited institution in a quantitative or biomedical science, or related field, with a minimum grade point average of 3.0.
- In addition to meeting the same admission requirements domestic students must meet, international applicants must also demonstrate proficiency in English. There are two ways to meet this English language proficiency requirement, which are outlined on the Graduate Division's International Admission Requirements (https://graduate.ucsf.edu/admission/intl-admission-requirements/) webpage. Please note that the Health Data Science program minimum internet based TOEFT iBT score is **100**.
- Transcripts
- Three letters of recommendation
- Resume or curriculum vitae
- Statement of Purpose
- Personal History Statement

Learning Outcomes

To complete the program, scholars must satisfy program objectives, which are to:

- Acquire a mastery of a broad set of data science research methods and in the techniques needed for the application of data science across biomedicine applications and research.
- Gain understanding of key issues that are particularly pertinent to the health sciences and evidence-based medicine, such as bias, confounding, interpretability, and causality.
- Plan and implement one or more health-related data science research projects.
- Write and submit a publication-quality research paper and a detailed methodology review.
- Present research results at a national or international meeting.
- Create a portfolio of data science skills and application areas.

Degree Requirements

- All core courses and required activities taken and passed with a grade C or higher.
- Maintain a cumulative GPA of 3.0 or higher (equivalent to a B average).
- Capstone project
- Educational apprenticeship (teaching assistant for one course)
- Unit requirement: 36 units

Core Courses

Course	Title	Units
Year 1		
Summer		
DATASCI 202	Opportunities and challenges of complex biomedical data (Opportunities and challenges of complex biomedical data)	3
DATASCI 213	Programming for Health Data Science in R (Programming for Health Data Science in R)	2
	Units	**5**
Fall		
BIOSTAT 200	Biostatistical Methods in Clinical Research I	3
DATASCI 214	Programming for Health Data Science in R II (Programming for Health Data Science in R II)	3
DATASCI 217	Introduction to Python and Data Science Tools	1-2
DATASCI 220	Data Science Program Seminar I	1
EPIDEMIOL 203	Epidemiologic Methods	3
	Units	**11-12**
Winter		
BIOSTAT 208	Biostatistical Methods II	3
DATASCI 216	Machine Learning in R for the Biomedical Sciences (Machine Learning in R for the Biomedical Sciences)	3
DATASCI 220	Data Science Program Seminar I	1
DATASCI 223	Applied Data Science with Python	2
	Units	**9**
Spring		
BIOSTAT 209	Biostatistical Methods III	3
DATASCI 220	Data Science Program Seminar I	1
DATASCI 224	Understanding Machine Learning: From Theory to Applications	3
EPIDEMIOL 201	Responsible Conduct of Research	0.5
HEALTH DATA SCIENCE ELECTIVE [1]		2-3
	Units	**9.5-10.5**
Year 2		
Fall		
DATASCI 221	Data Science Program Seminar II	1
DATASCI 222	Data Science Capstone Project	8

DATASCI 300	Data Science Educational Practice [2]	2
	Units	**11**
Winter		
DATASCI 221	Data Science Program Seminar II	1
DATASCI 222	Data Science Capstone Project	8
	Units	**9**
Spring		
DATASCI 221	Data Science Program Seminar II	1
DATASCI 222	Data Science Capstone Project	8
	Units	**9**
	Total Units	**63.5-65.5**

[1] Elective may be taken in any quarter.

[2] Educational Practice – may be any quarter through 2nd year.

Health Data Science Elective

Health Data Science (HDS) Elective selected from approved course list below (can be any quarter of the program). At least one must be completed.

Code	Title	Units
DATASCI 226	Bayesian Methods and Gaussian Processes	2-3
EPIDEMIOL 217	Molecular & Genetics Epidemiology I	2

Additional Information

Capstone Project

Students will begin developing a longitudinal capstone project as part of their requirements for the MiHDaS degree. Identification of the project will be encouraged in the first part of the program with the help of their UCSF faculty mentors (i.e. the members of their Graduate Committee), one of whom will be the Graduate Committee chair, one from the data science/biostatistics/bioinformatics faculty and one a clinical faculty member within UCSF.

The required capstone project encompasses four components:

1. Publication in a scientific journal
   a. Option 1: Submission of a first-authored publication in a scientific journal that is data science, general science or medical applications-based (this does not need to be accepted, but does need to be approved by the student's Graduate Committee); or
   b. Option 2: Submission of a non-first authored publication in a scientific journal where the student is generally expected to be within the first three authors; **and** a background and technical methodology report;
2. Master's committee oral defense;
3. Give an oral or poster presentation at a scientific conference; and
4. Compile a code and analysis portfolio for marketing the student's career skills.

These components were chosen to emphasize the crucial skills necessary to be a successful data scientist that go above and beyond purely technical skills. This includes but is not limited to:

- Carefully describing methodology used in a written format,
- Presenting work orally, and
- Conveying the importance of one's work in peer-reviewed publications and elsewhere.

This capstone element effectively provides students with an "apprenticeship" of sorts in the field of Data Science for the Health Sciences. By producing a submitted scientific paper approved by their committee, giving a presentation, and writing a methodological report, MiHDaS graduates will be able to clearly demonstrate that they are qualified to work in the field as part of a Health Sciences team.

Educational Apprenticeship

Students in the program will be expected to act as an educational apprentice (EA) for one course during their second year. This experience typically involves leading a weekly small-group discussion section of 10 to 15 students, holding office hours for students and grading homework assignments and projects. This requirement is designed to provide students with a valuable teaching experience without having a significant impact on the time needed for their capstone project work. In all cases, students will have taken during their first year the courses that they are asked to EA. Students will enroll in Data Science Educational Practice (DATASCI 300)