ABSTRACT: 2019 ELAM Institutional Action Project

Project Title: Developing a Health Data Concentration within the Master's Program in Biostatistics:

Name and Institution: Bhramar Mukherjee, Department of Biostatistics, University of Michigan School of Public Health

Collaborators and Mentors: Peter X Song and Michael R Elliott, Professor of Biostatistics

Topic Category (choose 1): Administration   Clinical   <u>Education</u>   Faculty Development   Research

Background, Significance of project: There is a rapid growth in the volume, velocity, variety and veracity of biomedical data generated by modern high-throughput technology. We need to train the next generation of workforce to handle biomedical big data. The classical training in Biostatistics need to be supplemented by teaching of advanced computational methods.

Purpose/Objectives: To stay relevant with science and to stay competitive with peer institutions, the Department of Biostatistics, University of Michigan is launching a health data science concentration within its existing Master's program. As the new chair of the department and a researcher in the field of health big data, I co-led the process of developing the new curriculum and operationalized its implementation for a planned first roll out in the Fall of 2019.

Methods/Approach/Evaluation Strategy: (1) A needs assessment was conducted by surveying our current students regarding the structure and topics in data science that are of high priority to them and are not currently offered. (2) A committee was appointed consisting of faculty in Biostatistics with expertise in big data and student representatives with interest in this concentration to arrive at a proposed curriculum. The committee met for an hour every week for the summer and Fall of 2018 to arrive at a program structure and a plan for integration into our existing Master's program. (3) Each new course was then assigned initial instructors for first two years who developed a detailed outline, syllabus and a set of competencies. The individual course proposals were discussed, scrutinized and refined multiple times before presenting the ensemble to the Biostatistics faculty. (4) A retreat was organized for the entire Biostatistics faculty to solicit their ideas and garner their support for this proposed curriculum. (5) Resources were negotiated at the institutional level and faculty teaching assignments had to be expanded/modified to accommodate the additional courses.

Outcomes/Results: A proposal for the sub-plan was approved by the Biostatistics faculty in December and by the School of Public Health's Advisory Committee on Academic Programs in February. The proposal is undergoing final review at the Rackham Graduate School. We fully expect the initial offering to begin in Fall 2019 with an approximate initial cohort of 15 to 20 students.
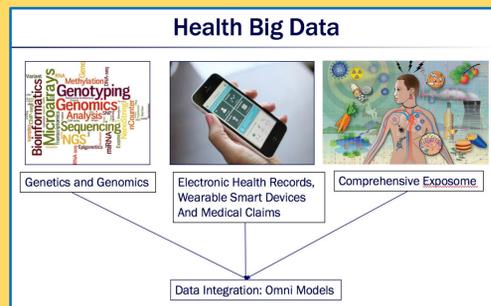
Discussion/Conclusion with Statement of Impact/Potential Impact: The field of health data science is changing rapidly with modern machine learning, artificial intelligence and data mining tools being used widely on massive datasets, and, often without careful consideration of sampling, selection bias, generalizability and inferential issues. A modern biostatistician is one who can harness the knowledge of big data into careful, credible and valid inference leading to actionable knowledge. The health data science plan within the Biostatistics Master's program will (a) offer hybrid training in both computational and stochastic modeling, (b) help attract a diverse body of students, (c) improve the training and experience of current students and (d) generate revenue plus recognition for the institution. We have plans to convert this concentration into online offering in partnership with Coursera. The impact of this program in intellectual, educational and financial terms are going to be profound for the department and the school.

Launching a Health Data Science Concentration Embedded in the Biostatistics Masters Program

Bhramar Mukherjee, PhD, Peter X-K Song, PhD and Michael R Elliott, PhD

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MICHIGAN

Background

Rapid advancement in high throughput technologies, increased access to continuous data streams from wearable smart devices and availability of patient records from large health care databases are leading to massive and messy datasets in biomedical research. Data are being generated in large volume, variety, velocity and veracity. These data sources are often not designed for population-based research but are being used for such purposes. Data linkage across domains has become common and critical.

Health Big Data

Genetics and Genomics | Electronic Health Records, Wearable Smart Devices And Medical Claims | Comprehensive Exposome

Data Integration: Omni Models

There is a tremendous need to train the next generation workforce in the field of health big data. Many institutions have launched a graduate curriculum in Data Science but the creation of data science programs with an exclusive focus on health is relatively new. As the incoming chair of the Department of Biostatistics at the University of Michigan (UM) School of Public Health, I wanted to lead and support the development of a concentration in health data science (HDS) within our existing Master's program.

Purpose

♦ To integrate data science and machine learning courses into a traditional Master's program in Biostatistics.

♦ To stay relevant in our training program, attract a more diverse body of students and remain competitive with our peers.

Methods

1 Stakeholder Meeting: Summer 2018

Met with faculty, students and key leadership to discuss the need for HDS training. This was done via retreats and town halls and one-on-one discussions.

2 Forming a committee : Fall 2018

Appointed a group of seven faculty members who are committed to this program to form a task force. Worked with them to draft a plan and timeline.

**3 Gap Analysis and Content Development: Fall 2018**

Studied existing data science programs at UM to minimize overlap and bridge the gap. Examined similar programs in the country. Solicited feedback via focused group with students.

4 Formal Submission: Winter 2019

Presented the proposal to the Biostatistics faculty and the School's Committee on Academic Programs for approval with a planned start in 2019.
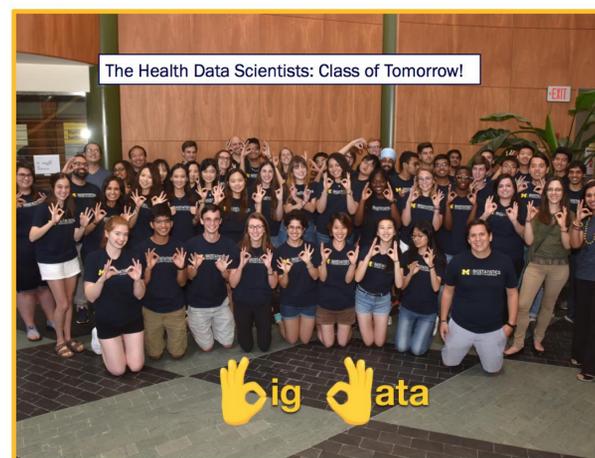
5 Approval, Advertisement and Launch

The proposal is in its final stages of approval and there is a strong interest in the pilot offering for the Fall of 2019. Marketing strategies are focusing on the exploding demand for data science.
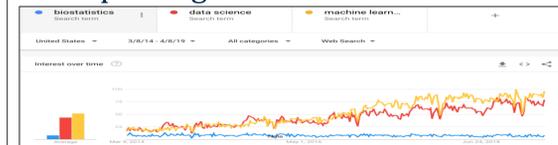
Results

➤ A new sub-plan in Health Data Science within the current Biostatistics Master's program was approved by the School-wide committee on February 15, pending minor revisions.

➤Resources were negotiated with the leadership based on teaching needs to recruit three new faculty working in the area of health big data: electronic health records, mobile health, precision medicine. These new recruitments will advance the research of the department in a modern direction.

➤Admitted MS students were informed about the program on the prospective student day. As of March 23, approximately 25 students enrolled in our inaugural cohort starting in 2019.

➤Michigan is the only Masters program besides Harvard University to offer a health data science concentration in 2019.

Nicholas Henderson | Xu Shi
Walter Dempsey | Welcome to our new Assistant Professors! M BIOSTATISTICS

New Faculty Recruits for HDS

Opportunities and Challenges

➤ The computing infrastructure of the department went through a major overhaul with expansion of the Linux cluster and recruitment of a new systems manager. This was an indirect effect of the HDS development that was welcomed by students, staff and faculty.

➤The enrollment in HDS courses may affect the enrollment in other more traditional courses, this creates a tension and competing interest between classic statistics and modern data science. We need to navigate this better to promote an inclusive research and training environment

The Health Data Scientists: Class of Tomorrow!

big data

Conclusions and Impact

♦ There is a growing need for trained professional in the field of health big data

♦ A brand new curriculum in health data science with more courses on advanced computing, visualization and machine learning is going to transform the existing landscape of Biostatistics education

♦ The program is going to increase revenue, recognition and relevance for the School and the department.

♦ Mindful attention should be given to integrate the foundational and the cutting edge material