Engaging Students in Health Informatics Research: Strategies for Success

Sorina Madison, Janusz Wojtusiak

Health Informatics Program, Department of Health Administration and Policy, George Mason University

Hands-on, research related experiences during academic training are known to "empower students as learners" and expose them to new ideas.[1] Overall, these experiences increase student satisfaction and retention, improve educational outcomes, and increase the likelihood to pursue further education and conduct research in the future.[2] Health informatics, in particular, is a field of study where specific, applied research skills and ability to work independently are expected for employment.[3] Regional groups are forming in the Washington, DC, area, to connect the industry with local academic partners and describe competencies and applied health informatics skills that students should possess at graduation. However, little is known about the best way to motivate participation and how to address barriers to health informatics student engagement in applied research.

At George Mason University, health informatics programs (undergraduate and graduate) require a practicum/capstone course. However, the limited time available (14 weeks) and the lack of well-established preexisting research skills may place too much pressure on some of the students and leaves them unable to fully take advantage of an industry placement, including opportunities for networking, job seeking, and refining of professional abilities. While the depth of skills and conceptual understanding of the issues varies between levels of academic work, we believe that all our health informatics students should be engaged in research activities throughout their academic programs, to better prepare for jobs in the field at graduation.

This project explores the feasibility of engaging health information students (graduate and undergraduate) in research projects, identify barriers to participation, and develop strategies to promote student engagement and collaboration, better preparing them for applied work in health informatics. Collaboration and engagement are promoted through: 1) weekly work meetings, and 2) the creation of a virtual Learning Community where participants interact, connect between meetings, post and discuss research results. Students were expected to take ownership of the various subprojects and engage in all major areas of the project: data collection, literature analysis, IRB application, data analysis, model construction, programming, testing, and preparation of manuscripts and presentations. Student engagement is measured through participation metrics such as: answering calls for support, participation in planning meetings, contributions to the online Learning Community, work on research subprojects,

collaborations with peers, research output (posters, abstracts, papers, reports), and research dissemination (presentations, web presence). In addition, student opinion about perceived engagement in relevant research, as well as the benefits for the student's learning, retention, and future professional career were collected using web-based surveys at two different points in time: at the beginning of the student's involvement with the project, and at the end of the project, given that students had participated for at least one semester.

Instead of assigning a limited number of students to different ongoing projects at different stages of completion, we started a new research project that could accommodate a large number or students and provide exposure to all stages of developing and conducting a large research project. The *Intelligent Patient Data Generator* will ultimately allow us to construct virtually unlimited artificial patient data to be used for software testing (e.g., EHR), algorithm development, education, or simulation purposes, while avoiding legal limitations on using and sharing real patient datasets, and the availability of large datasets that follow patients over time.[4] There are two steps involved in the creation of synthetic, longitudinal data: 1) building models based on real clinical and administrative patient information that capture patient trajectories from birth to death, and 2) generating artificial patient data using these models, while maintaining the diversity of real data.

For this project, we focused on the first step: building models which incorporate clinical and administrative knowledge related to A) the sequence and relationships between all data elements present in a given patient, and B) the statistical/probabilistic information related to the prevalence of values of specific data elements in a given population. We explored the possibility of using a series of predictive models, that given a patient history at time T, we can predict the probability of certain events occurring in the near future of T.

Project outcomes are two-fold: a list of recommended strategies that have been found to encourage student engagement in applied health informatics research throughout their academic program, and a library of models that capture patient trajectories from birth to death, which will be further used to generate synthetic health data. Both outcomes should be of interest to both academics and practitioners, as they promote development of a health informatics workforce that is better prepared for applied work at graduation, as well as take a step towards developing methods to generate artificial patient data to be used not only for education and training, but also for software testing and algorithm development.

[1] Association of American Colleges and Universities (2002). Greater expectations: A new vision for learning as a nation goes to college. Washington, DC: Association of American Colleges and Universities.
Association of American Colleges and Universities (2007). College learning for the new global century. Washington, DC: Association of American Colleges and Universities.

[2] Justice, C., Rice, J., Warry, W., Inglis, S., Miller, S., & Sammon, S. (2007). Inquiry in higher education: Reflections and directions on course design and teaching methods. Innovative Higher Education, 31, 201–214.
Nnadozie, E., Ishiyama, J., & Chon, J. (2001). Undergraduate research internships and graduate school success. Journal of College Student Development, 42, 145–156.
Pascarella, E. T., & Terenzini, P. T. (2005). How college affects students: A third decade of research. San Francisco, CA: Jossey-Bass.

[3] AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. J Am Med Inform Assoc 2012;19:6 931-938.

[4] Wojtusiak, J., Towards Intelligent Patient Data Generator, *Reports of the Machine Learning and Inference Laboratory*, George Mason University, 2014.