In July, a group of mathematicians, physicians and computer scientists will head off to boot camp. They won’t be marching in formation, but they’ll get their feet wet poring over such topics as natural language processing, visualization analytics and data mining. This boot camp is a primer on medical research in the era of big data.

The four-day Big Data Coursework for Computational Medicine (BDC4CM) is the brainchild of three researchers with Minnesota connections. They responded when the National Institutes of Health (NIH) asked for help getting people up to speed using massive amounts of information generated by such sources as gene sequencing machines, electronic health records (EHRs) and even smartphones. The NIH was offering grants as part of its Big Data to Knowledge (BD2K) initiative, launched with the Obama administration in 2013. The initiative’s aims include increasing integration of data science into biomedical research.

Collaborative coursework
When the NIH grants first became available, three Mayo Clinic colleagues—Christopher Chute, MD, DrPH; Claudia Neuhauser, PhD; and Jyotishman Pathak, PhD—came up with the idea of bringing together “fellows” from various fields for a crash course in computational medicine.

“We thought, between Mayo Clinic and the University of Minnesota, and with the computational side and the clinical side, we could put together a strong training program,” says Neuhauser, a mathematician who now directs graduate studies in bioinformatics and computational biology at the University of Minnesota. Chute is currently at Johns Hopkins University in Baltimore, and Pathak is at Weill Cornell Medicine in New York.

The NIH responded to the trio’s proposal with funding for three annual big data boot camps, the first of which took place in Rochester, Minnesota, in 2015. The second occurred last year in New York, where this year’s event also is scheduled.

Neuhauser says her team’s goal is to introduce participants to concepts and tools, and to make them aware of ethical issues that methods like data mining and gene sequencing raise. BDC4CM is not aimed at the small group of people who want to go “deep into the field,” she emphasizes. “That’s a whole degree program.” Instead, this experience is geared toward those—including physicians—within a much larger group of people, who need to understand what the smaller group is doing.

Data mining insights
Attendees at last year’s BDC4CM included Nishant Sahni, MD, MS, an assistant professor of medicine at the University of Minnesota and a hospitalist at University of Minnesota Medical Center (UMMC). Sahni says it was eye-opening to see how individuals from different fields, coming from different perspectives, think about the potential of big data. “It was almost like the five blind men with the elephant in the room, with everybody trying to feel what the field was about,” he says.

Sahni was drawn to the program by his interest in mining EHRs for information that can be used to improve clinical care. “All this data is sitting there and nobody is using it,” he says. Sahni has already tapped UMMC’s EHR for information about care of patients diagnosed with sepsis in the emergency room. He was
able to determine the proportion of patients who immediately received life-saving antibiotics, which has an impact on outcomes.

Sahni says data mining is very different from traditional research, where the goal is to generate data. “This is the other way around,” he says, explaining that the researcher’s task is to look for what might be found in data that already exists. “This is almost like dumpster diving.”

Individuals typically can’t do that work on their own, as it requires not only a clinical question but also the ability to query a database; organize data once it’s pulled; and then analyze it. “You typically won’t find single-author papers,” Neuhauser says. “It’s usually a group of people with different skill sets investigating something.”

Language lessons
The BDC4CM goals include bridging gaps in understanding between those who approach projects from the biomedical side and those who come from the computational side. Neuhauser says that when representatives from both groups attend the workshops together, it helps people from one group learn the “language” and understand the concerns of those in the other.

There’s evidence that such cross-communication is happening out in the field, as well. “Increasingly, people on the computation side have quite a bit of knowledge about the biomedical side,” Neuhauser says, noting that she’s learned a lot about cancer genomics, for example. Meanwhile, some clinicians are educating themselves about computers. Count Sahni among that group: He’s taught himself how to code.

Sahni believes that stepping outside of one’s professional silo will be important to those who wish to leverage big data in the future. “Increasingly,” he says, “this is a field that is going to be driven by people reaching out to the other side.” – CARMEN PEOTA

---

**Big Data Boot Camp fellowships**

The National Institutes of Health funds 20 fellows a year to attend Big Data Coursework for Computational Medicine (BDC4CM), defraying the costs of travel, meals and registration. The 2017 BDC4CM registration deadline is March 15.

**For more information, visit bdc4cm.org**