



Stanford Center for  
Biomedical Informatics Research

# Connection

## What's Inside

Treating COVID-19 Patients  
A Leadership Role in Research  
My Day: Manisha Desai, PhD  
Faculty Focus  
Knowledge Graphs  
Attacking Cancer

Fall 2020

*We Connect Data to Health*

## Diving into COVID—and Looking Beyond

As you will read in this issue of *Connection*, BMIR faculty are leading many activities to address the most pressing biomedical issue of our time, COVID-19. BMIR's solid relationship with the clinical enterprise makes us well situated to address the serious challenges that the pandemic presents. Because our faculty are either clinicians themselves or work closely with peers in the clinical arena, we are naturally positioned to make great progress here.

Not only that, but the benefits of our participation in translatable, actionable work will extend well beyond the immediate clinical problem. COVID-19 is stimulating considerable basic research in informatics, which means that the results of our efforts today will apply not only to COVID, but also to health matters far into the future. For example, as you will read on Page 4, Olivier Gevaert is refining his image analysis techniques. That work could help someone

to diagnose COVID pneumonia, but it also will have ramifications for Dr. Gevaert's work on the characterization of other respiratory diseases, of cancer, and of scourges yet to emerge.

Mark Musen, MD, PhD

Director, Stanford Center for Biomedical Informatics Research

## CLINICAL CARE Treating COVID-19 Patients: What We Can Learn



Jonathan Chen, MD, PhD (above), and Lance Downing, MD, have treated some 30 COVID-19 patients to date.

*Physician-scientists Jonathan Chen, MD, PhD, and Lance Downing, MD, have taken care of about 30 COVID patients in the hospital and in clinic since February, 2020. They were recently asked how their experience in treating those patients relates to BMIR's mission in informatics.*

**How does your work with COVID-19 patients connect with your informatics research interests?**

**Downing:** COVID has made many other patients delay visits to health care facilities

and receive services remotely. An increase in care delivery away from health centers requires extending the reach of health care from the hospital into the home. One project I'm working on is Deepstroke, which aims to improve stroke diagnosis through machine learning algorithms. We think this technology can eventually be used to alert high risk patients that they're having a stroke so they can seek treatment sooner.

Another project is CARE-IT (Care Itinerary), which was developed in collaboration with Stanford Biodesign. It's a mobile app that allows patients to learn about, record, and share advanced care planning wishes. It's especially timely with COVID-19 because both the risk of severe illness is elevated and the ability for patients to have these conversations in person with their doctors is greatly impaired.

**Is there another example of how informatics work connects to patient care?**

**Chen:** While informatics tools are essential to enable things like telemedicine, I'm trying to take it to the next level, electronic consultation. In-person visits were already inefficient, and now with COVID-19, they aren't even possible in some cases. So we're

working on how to facilitate on-the-spot medical decision making to reduce delays in patients' access to care for non-COVID matters. There's also our work in advanced clinical research infrastructure that's enabling the CDC to improve their ongoing public health efforts [see "Leadership" article on page 2].

**Can you describe the breadth of the BMIR division?**

**Chen:** We have intensive data analytic and statistical capabilities for informatics work, but we are also active in applying that clinically. In fact, understanding how informatics work interacts with real human health care is rare, but it's a unique value that our division offers.

While we do deal with abstract concepts, high technology, and advanced ideas, we also bring that to the bedside. Lance and I see patients, so we are literally on the front lines. That helps us formulate questions and inspires us to think about how to tackle things that couldn't be done in isolation.

Inquiries about the work of Drs. Chen and Downing may be sent to [contact-BMIR@stanford.edu](mailto:contact-BMIR@stanford.edu).

## COVID-19 BMIR Assumes Leadership Role in Research

BMIR faculty have been involved in no fewer than 27 of the scores of research projects that Stanford Medicine scientists are performing in response to COVID-19.

The efforts of Nigam Shah, MBBS, PhD, and his team have used de-identified patient data for [15 published reports](#) that examine trends in presenting symptoms of patients and that monitor the diagnostic performance of tests for the novel coronavirus, the age distribution of positive cases, and hospitalization rates. The Shah lab also provides “data curation” for other scientists working to create better models of the COVID-19 pandemic, and it participates in COVID-19 studies in the Observational Health Data Sciences and Informatics collaborative.

Utilizing the [STanford medicine Research data Repository \(STARR\)](#), a team led by Jonathan Chen, MD, PhD, has been able to give near real-time updates on clinical processes and outcomes surrounding

COVID-19 to enable the CDC’s ongoing public health efforts.



Purvash Khatri, PhD, is applying multi-cohort systems immunology approaches to better understand the host response and its regulation in viral infections including SARS-CoV-2, Ebola, dengue, influenza, and other viruses to develop novel diagnos-

tics to keep the “curve flat” by identifying patients at lower risk of severe infection who can recover at home.

Quantitative Sciences Unit Director Manisha Desai, PhD, established a shared infrastructure for facilitating clinical trials as part of her team’s leadership of data coordinating centers for seven studies involving COVID-19. Dr. Desai also leads the data science behind a collaborative [initiative](#) that involves estimating the incidence and prevalence of COVID-19 in the Bay Area. In a recent [podcast](#), Dr. Desai provided insights into the challenges and progress of COVID-19 clinical trials.

Tina Hernandez-Boussard, PhD, is completing several manuscripts about a multi-institute study related to the clinical manifestations of COVID-19. She is also investigating the disproportionate impact of the pandemic on minorities, with a focus on health disparities in terms of infection rates, hospitalizations, and clinical manifestation.

### MY DAY Manisha Desai, PhD



Teaching is a primary activity for Manisha Desai, PhD.

What’s a typical day for Manisha Desai, PhD? A recent one began with a planning meeting to discuss the role of the [Quantitative Sciences Unit \(QSU\)](#) as the data coordinating center for current and future trials taking place in the outpatient setting.

Dr. Desai then attended a joint Steering/Executive Committee meeting for ARREST,

a clinical trial involving hospitalized patients with pneumonia and hypoxemia, with the QSU serving as a data coordinating center. The agenda focused on recent findings about dexamethasone and how to adjust the trial in view of those findings.

As a champion of team science, it’s not surprising that Desai’s lunch took place during the QSU’s weekly two-hour team meeting, where faculty and staff discussed how to design studies in novel ways.

Next, the QSU director met with a group of data scientists who work with Tom Robinson, MD, a pediatrician who’s developed a very successful program to reduce obesity in children. While Dr. Desai normally works with Dr. Robinson in the clinical trial setting, that day’s meeting focused on how to

overcome methodological gaps in pediatric research, such as how to bring real world data sources like accelerometers into clinical trials at the risk of missing data if children don’t wear the instruments. Coincidentally, Dr. Desai was also serving as a principal investigator of a study of how accelerometer data can be processed and analyzed for integration into clinical trials.

Education is a priority for Dr. Desai, which is why she also found time to meet with a bioinformatics fellow and a cardiovascular fellow, whom she mentors to build their capability in research methods.

The long day extended into the night and the weekend, about the only time Dr. Desai can work on manuscripts, the lifeblood of the QSU. That day she wrote about two current projects: one paper described the model for her team and how they collaborate; the other addressed the infrastructure her team created to facilitate trials in the current pandemic.



Tina Hernandez-Boussard, PhD, is working to ensure that no unintended consequences come from the reporting of machine learning results.

Tina Hernandez-Boussard, PhD, understands that most clinical assertions are developed through randomized controlled trials, and those don't always generalize into routine treatment. That's why she and her team are using data from everyday care

to verify how generalizable some of those clinical assertions are and how evidence generated from routine patient care can [guide policy and clinical workflow](#). For instance, she and her team are using real-world evidence to understand the effectiveness of novel treatments in a broader population, not just the effect among patients in a controlled study.

The [Boussard Lab](#) is also using natural language processing and machine learning technologies to understand clinical data sets. Dr. Hernandez-Boussard says it's relatively simple to search for discrete data fields in electronic health records, but she and her colleagues are using [biomedical informatics technology](#) to dig for nuances in the data – such as clinical notes and patient surveys of reported outcomes – to get a holistic view of the patient. Importantly, they are searching to understand not just whether a patient has a symptom or outcome, but the [severity of that outcome](#), to help guide clinical treatment.

A third area of focus is how the information generated from machine learning and artificial intelligence algorithms can be applied without bias across populations. In that regard, Dr. Hernandez-Boussard is active with [AcademyHealth](#), [AMIA](#) (the American Medical Informatics Association), and other organizations to publish [standards](#) regarding the reporting of machine learning results to ensure that no unintended consequences come from the work in data-driven medicine.

Collaboration, with other School of Medicine departments as well as entities beyond Stanford, figures prominently in Dr. Hernandez-Boussard's activities. In fact, many industry partners have approached Dr. Hernandez-Boussard in recent months to ask if she can shift her research to look at COVID-19 [see "Leadership" article on page 2] as they connect her with other global researchers working in that domain.

## COLLABORATION Knowledge Graphs Give Real World Uses to Collected Data

BMIR is collaborating with two prominent companies to build knowledge graphs to improve their ability to organize biomedical knowledge.

### ELSEVIER

Elsevier, a major Dutch publishing and analytics company, is working with BMIR Director Mark Musen, MD, PhD, and his lab to enhance the knowledge graph infrastructure used by Elsevier to store machine-usable clinical knowledge and to index its vast supply of clinical publications. The goal is to improve search capabilities and to drive decision support for physicians.

In another case, BASF, one of the world's largest chemical companies, has sought the expertise of the Musen Lab to structure information from multiple, disparate databases and to improve development of pharmaceuticals and other compounds.

BASF and Elsevier are building their knowledge graphs using WebProtégé, the web version of the [Protégé](#) tool, a knowledge graph editor developed by Stanford.

Knowledge graphs may incorporate ontologies—logic-based knowledge-representation formalisms that are used to structure and define information gathered from a variety of sources. Ontologies define the concepts in a given application area—for example, all the compounds needed for BASF's chemical engineering work. Humans may understand these concepts intuitively, but a computer requires a formal representation of the knowledge about such concepts. One area of BMIR's expertise is in building infrastructure to encode all those disparate bits of knowledge in a way that is interpretable both by humans and computers. That infrastructure allows companies like Elsevier and BASF to construct knowledge graphs.

Physicians will be able to use an Elsevier system to help them make more informed

clinical decisions. The knowledge graph will also facilitate searches of Elsevier's enormous collection of books and journals. Currently, searching this expansive collection of publications can be hard when terms are used in different ways, such as saying either heart attack or myocardial infarction. The Musen Lab will help Elsevier extend and verify the quality of its knowledge graphs to make the content in the Elsevier systems more useful and accessible.

### **BASF** We create chemistry

BASF is collaborating with the Musen Lab to develop infrastructure for engineering massive knowledge graphs that include definitions for millions of chemicals and their properties. The enormous scope of the BASF graph makes this project particularly challenging.

## SPOTLIGHT Attacking Cancer with Machine Learning and Software Tools

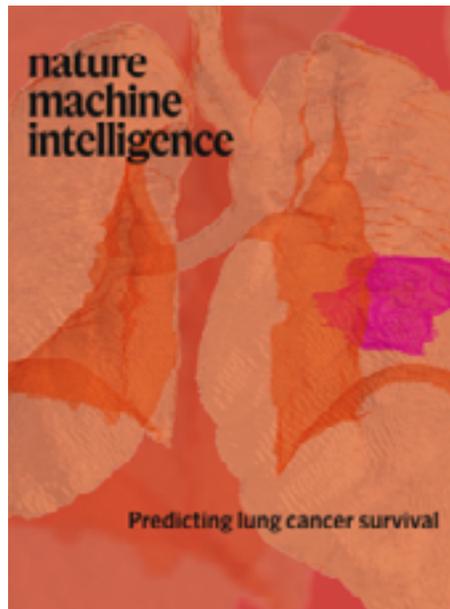
A translational informatics specialist who is an expert in biomedical data fusion and quantitative imaging, Olivier Gevaert, PhD, looks at cancers across all specialties. A recent [paper](#) in *Nature Machine Intelligence* focused on Dr. Gevaert's lung cancer work.

While CT imaging is an important diagnostic tool in guiding therapeutic decisions for patients with lung cancer, analysis of CT images is limited to what is visible to the human eye. Because of different analyses of images among radiologists, there is a wide discrepancy in clinical care decisions among oncology centers.

Dr. Gevaert led a team of researchers in biomedical informatics, radiology, data science, electrical engineering, and radiation oncology to create a machine learning neural network called **LungNet**. Using LungNet with CT images of lung cancer patients, the team developed a model to predict the prognosis of those patients.

They tested the model using data from several hundred cases each from Stanford and three other hospitals in Florida, the Netherlands, and Germany.

"The tests showed that prognosis predictions based on imaging outperformed other types of clinical data. They also showed that the use of deep learning improves prediction of prognosis from CT images compared to a standard computer vision approach using radiomics features," Dr. Gevaert said.



Dr. Gevaert's lung cancer paper was featured on the cover of the May 2020 issue of *Nature Machine Intelligence*.

He added that LungNet illustrates how machine learning technology can be a cost-effective approach to advance disease detection, diagnosis, and treatment. He expects that LungNet can be used to classify benign versus malignant tumors as well as to stratify patients into low, medium, and high risk groups. That will foster intensified treatment for patients in higher risk groups while reducing unnecessary treatments for patients in lower risk groups.

The work was supported by the NIH's National Institute of Biomedical Imaging and Bioengineering.

Dr. Gevaert, the chair-elect of the Biomedical Imaging Working Group of the American Medical Informatics Association, also focuses on brain cancer, including glioma, a type of tumor found in the brain stem and spinal column.

"These are very aggressive tumors," he said. "Treating this devastating disease is a difficult problem for medical oncologists and tumor boards. In summary, our team recognized that with the rapid advancement in high-throughput DNA sequencing, digital pathology, and quantitative analysis of radiographic images, more quantitative data is now available, potentially enabling more accurate and effective treatment decisions. However, no efforts we knew of had been undertaken to integrate these disparate data modalities: molecular biomarkers, digital pathology, and multi-modal MR images."

A recent [article](#) in *JCO Clinical Cancer Informatics* describes how Dr. Gevaert's team applied Imaging-AMARETTO algorithms and software tools to brain tumors. Imaging-AMARETTO is a data-driven platform for diagnostic, prognostic, and therapeutic decision making in cancer. In the article, they showed how the formulation and development of computational imaging genomics methods can link multiomics, imaging, and clinical data to predict clinical outcomes. Specifically, they applied Imaging-AMARETTO to integrate three patient studies of brain tumors.



# BMIR

Stanford Center for  
Biomedical Informatics Research

WE CONNECT DATA TO HEALTH

The Stanford Center for Biomedical Informatics Research (BMIR) uses advanced research techniques to discover, apply, translate, and organize data that make a difference for health and health care. With its expertise in clinical and translational informatics research and biostatistics, the division works to uncover new ways to ad-

vance personalized medicine and to enhance human health and wellness.

Collaboration is in our DNA. We are excited about the prospect of working with other experts who share our goal to connect data to health and medicine. We encourage you to contact Mark Musen, Director of BMIR ([musen@stanford.edu](mailto:musen@stanford.edu)), to learn more.

Connection  
Fall 2020

Stanford Center for Biomedical Informatics Research  
1265 Welch Road, Stanford, California 94305-5479  
<https://bmir.stanford.edu>