



U.S. Department of Veterans Affairs  
Veterans Health Administration



# THE BIG DATA SCIENTIST TRAINING ENHANCEMENT PROGRAM

Developing the Next Generation of Healthcare Data Scientists

## MISSION

### BD-STEP

The Big Data Scientist Training Enhancement Program (BD-STEP) is a two-year fellowship program that uses data science to advance cancer research and patient care. A Veterans Health Administration (VHA) advanced fellowship launched in 2015 in collaboration with the National Cancer Institute (NCI), the program provides well-rounded training and unparalleled access to VA data resources and NCI cancer research expertise. Competitively selected postdoctoral fellows work with VA clinicians and cancer researchers to gain valuable clinical exposure and oncology domain knowledge. Fellows use comprehensive health data to pursue patient-centered research questions, improving basic understanding of cancer while also improving clinical care for Veterans.

### VHA

The mission of the VHA is to honor America's Veterans by providing exceptional health care that improves their health and well-being. BD-STEP connects talented early career data scientists with VA researchers and clinicians to advance healthcare for our Veterans. Through clinically-oriented operational projects and research that harnesses the VA's big data resources, fellows' research projects can inform healthcare administrators and empower clinicians to translate findings to improve patient care. VHA provides program leadership, VA Medical Center oversight, and fellow salaries and benefits for BD-STEP.

### NCI

NCI leads, conducts, and supports cancer research across the nation to advance scientific knowledge and help all people live longer, healthier lives. NCI's charge to support workforce development includes training and mentoring the next generation of cancer researchers. The development of data scientists in oncology is particularly important to harness the massive generation of data across the cancer continuum and answer fundamental questions in cancer research and care. Research guidance and support for BD-STEP fellow travel, training, and curriculum development are provided by the NCI Center for Strategic Scientific Initiatives.



# BD-STEP and the Value of the VA Health Data

The Big Data Scientist Training Enhancement Program (BD-STEP) was launched in 2015 to train the next generation of healthcare data scientists capable of interpreting and gaining insights from large clinical datasets.

The Veterans Health Administration (VHA) is America's largest integrated healthcare system, providing care at 1,250 health care facilities and serving 9 million enrolled Veterans each year. The long-term care Veterans receive within this centralized healthcare system provides a rich source of longitudinal patient data—covering patients through periods of health and illness. This is unique to the VHA, as the care patients receive in other US healthcare organizations is often fragmented among different clinical sites, making it difficult to obtain a complete patient profile through the aggregation of medical records.

Within the integrated VA healthcare system, there are many untapped opportunities to gain insights from patient data to advance cancer research and care. BD-STEP provides an avenue to access the rich, diverse data available in the VA Electronic Health Record (EHR), including longitudinal clinical patient data and diagnosis and treatment information from the VA Central Cancer Registry. BD-STEP utilizes the expertise of early-career data scientists to analyze these data and facilitate the execution of large-scale system changes in clinical care.

Fellows are placed in four VA medical centers across the country to work with clinicians and interdisciplinary researchers to address important patient-centered health challenges. The sites are guided by an advisory council with VHA and NCI membership, including the NCI's Center for Strategic Scientific Initiatives, Center for Cancer Training, and Center for Biomedical Informatics and Information Technology.

Over the course of their research, fellows network with healthcare and data science experts across government, industry, and academia. They receive research mentorship from VA healthcare providers and academic researchers and curriculum oversight by VHA and NCI program leadership. This equips BD-STEP graduates with the skills and connections they need to pursue careers in healthcare data science after graduation.

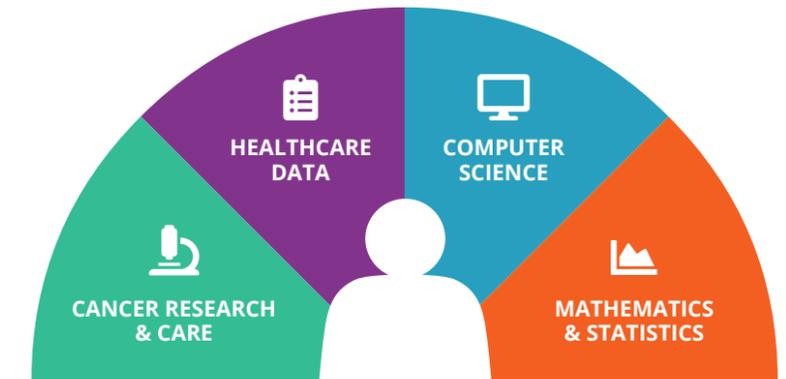
Since the launch of the program, BD-STEP fellows have initiated diverse studies using VA healthcare data resources. These including predicting hepatocellular carcinoma in hepatitis C patients using a cohort of more than 180,000 Veterans, comparing frailty assessment via clinical teams and machine learning to predict mortality in patients with congestive heart failure, and characterizing dynamic biological changes associated with prostate cancer progression in obese patients.

## Fellowship Milestones

- APPOINTMENT**  
 Fellows are appointed to selected VA medical centers across the country.
- TRAINING**  
 Fellows learn how to access and navigate the VA data and shadow clinicians to understand real-world needs and the clinical environment.
- COLLABORATIVE RESEARCH**  
 Fellows work with academic and clinical advisors to develop and address important cancer research questions.
- CAREER LAUNCH**  
 Fellows continue in careers at the intersection of big data and health.
- IMPACT ON RESEARCH & CARE**  
 Research outcomes are published and projects are continued in NCI-supported academic institutions and VA medical centers, advancing cancer research and improving care.

## Profile of a BD-STEP Graduate

Fellows who graduate from the program are more than data scientists; they are interdisciplinary researchers who use data science to make a difference in patient care. Fellows use computer science and mathematics to gain insights from healthcare data and solve real-world clinical cancer problems, launching their careers in healthcare data science. Connections made during BD-STEP last beyond the end of the fellowship, providing a network of support and collaboration for early-career scientists.



### CONTACT INFORMATION:

**Frank Meng, PhD**  
 BD-STEP Program Director, VHA  
[frank.meng2@va.gov](mailto:frank.meng2@va.gov)

**Michelle Berny-Lang, PhD**  
 Program Director, NCI Center for Strategic Scientific Initiatives  
[michelle.berny-lang@nih.gov](mailto:michelle.berny-lang@nih.gov)

## Data from America's largest integrated healthcare system

### BD-STEP Network: Fiscal Year 2020

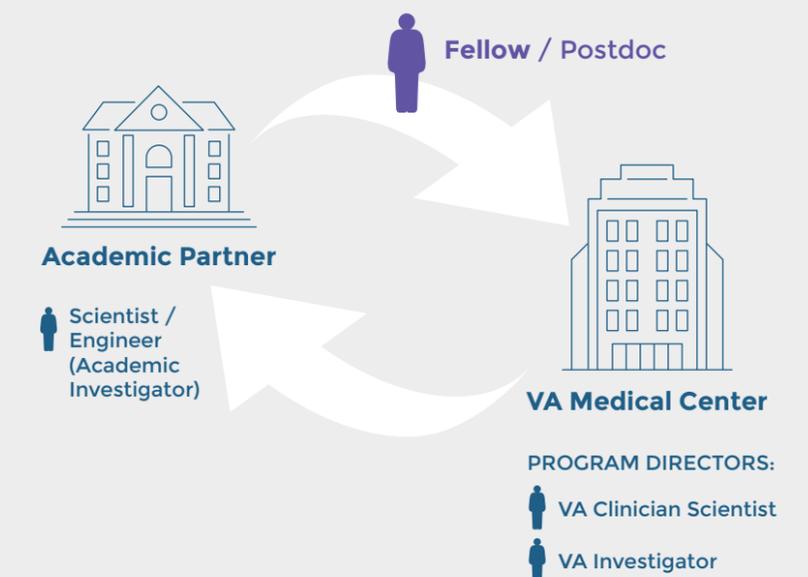
**33**  
 FELLOWS  
 ACROSS  
**4**  
 COHORTS



## Applicant Backgrounds

Fellows come from diverse academic and research backgrounds in fields including:

- MATHEMATICS
- ENGINEERING
- COMPUTER SCIENCE
- INFORMATICS
- CHEMISTRY
- PHYSICS



## HOW OUR FELLOWS HAVE BENEFITED



*“BD-STEP allows clinicians and researchers to solve real-world problems with an immediate impact on our nation’s Veterans. Using my education to help others is one of the greatest career successes that I could have ever asked for.”*

### **JEREMY MASON**

Assistant Professor of Research Urology, Keck School of Medicine, University of Southern California / Class of 2016 / Modeling Liver Disease Progression



*“I developed skills in machine learning, data analytics, and statistics, and enhanced my healthcare domain knowledge through collaboration with VA physicians. The data is incredibly extensive and there is still so much to learn.”*

### **JOANNA SYLMAN**

Data Scientist – Analytics, Komodo Health / Class of 2018 / Changes Associated with Prostate Cancer Progression in Obese Patients



*“Through BD-STEP, I learned the role of advanced data analysis in the setting of a nationwide healthcare system and accessed the VA’s large real-world healthcare datasets.”*

### **NATHANAEL FILLMORE**

Associate Director for Machine Learning and Predictive Analytics, VA; Instructor in Medicine, Harvard Medical School / Class of 2017 / Clinical and Genomic Factors of Multiple Myeloma Progression



*“The experience I gained from BD-STEP allowed me to move into a career with a true impact on the care that Veteran patients receive.”*

### **HANNAH GELMAN**

Research Health Science Specialist, VA / Class of 2018 / Using MVP data to assess genotype-guided Warfarin dosing algorithms



*“I have the privilege to work with the largest electronic health record in the United States to build my future career.”*

### **JAVAD RAZJOUYAN**

BD-STEP Fellow / Class of 2019 / Assessing Frailty to Predict Mortality in Patients with Congestive Heart Failure



*“BD-STEP connected me with leaders in clinical care at the VA, which led to my current VA position. It’s fulfilling to use my career to bring cutting-edge treatments to Veterans nationwide.”*

### **BRADLEY HINTZE**

Data Scientist, National Oncology Program, VA / Class of 2017 / Data infrastructure for precision oncology

FOR MORE INFORMATION VISIT:

[cssi.cancer.gov/bd-step](https://cssi.cancer.gov/bd-step) and [va.gov/oa/specialfellows/programs/sf\\_bdstep.asp](https://va.gov/oa/specialfellows/programs/sf_bdstep.asp)