Multi-institutional validation of machine learning algorithms to predict and reduce acute care during cancer therapy

Julian Hong, MD MS[1]; Jean Feng, PhD[1]; Ryzen Benson, PhD[1]; Isabel Friesner, BA[1]; Daphna Spiegel, MD MS[2]; John Boyle, MD[3]; Olivier, Morin, PhD[1]; Jesse Alexander, MD, PhD[1], Nitin Ohri, MD MS[4]; Manisha Palta, MD[5]

[1]University of California, San Francisco; [2]Beth Israel Deaconess Medical Center; [3]Essentia Health; [4]Montefiore Medical Center; [5]Duke University

An estimated 650,000 patients with cancer receive systemic therapy or radiation therapy (RT) annually in the United States. Many of these patients undergoing outpatient cancer therapy will require acute care with an emergency department visit or hospital admission due to symptoms from treatment, disease, or comorbidities. This can impact cancer outcomes, patient treatment decisions, and costs to patients and the healthcare system. While there has been much enthusiasm for artificial intelligence and machine learning (ML) to improve healthcare delivery, high quality prospective data are lacking, especially across diverse clinical practice settings.

We previously completed one of the first randomized controlled studies in healthcare ML, demonstrating that ML based on EHR data can accurately generate personalized predictions (prospective AUC 0.851) and guide supportive interventions to decrease acute care visits (22% to 12%) and costs in patients undergoing RT and chemoradiotherapy (CRT) (NCT04277650). We have also recently developed a ML model for predicting hospitalizations based on prospective clinical trials of daily step counts collected in patients undergoing CRT (temporal validation AUC 0.85).

Our research objective is to leverage a geographically, racially, socioeconomically, and technically diverse network of healthcare settings and patients to assess and maximize how accurately and equitably these approaches generalize. Our team includes the University of California, San Francisco (UCSF), Duke University, Beth Israel Deaconess Medical Center, Essentia Health in Duluth, MN and Ashland, WI, Washington Hospital in Fremont, CA, Duke Regional Hospital in Durham, NC, and Duke Raleigh Hospital in Raleigh, NC. Specifically, we seek to: (1) prospectively evaluate the validity of an EHR-based acute care prediction ML algorithm across our network and establish a framework for equity, generalizability, and portability and (2) validate our existing patient-generated health data (PGHD; step count) models that predict hospitalization during CRT at UCSF and integrate with our EHR-based ML algorithm to enhance prediction of acute care needs.

A pipeline to collect and harmonize the data across the multiple institutions is currently in progress. The daily step count-based prediction model was recently completed and the external prospective validation trial is in preparation and anticipated to be open to accrual soon.

The long-term goal is to develop informatics-based tools that can be broadly and equitably deployed to improve the delivery of cancer care and subsequent treatment outcomes. This research will generate data regarding the generalizability and fairness of EHR- and PGHD-based approaches and a platform for a future multi-institutional randomized controlled trial.