

Predicting *C. difficile* recurrence from the microbiome using interpretable machine learning models

Principal Investigator: Georg Gerber, MD PhD. Brigham and Women's Hospital

Clostridioides difficile infection (CDI) is the most common cause of hospital-acquired infection in the U.S., responsible for ~500,000 cases of severe diarrhea and ~15,000 deaths annually. First-line treatment is ineffective in ~25% of CDI patients, creating substantial unnecessary morbidity with costs to the healthcare system of ~\$2.8 billion. There is thus an urgent need for accurate diagnostics to identify patients at risk for recurrent CDI in order to provide effective initial treatment. A causal relationship has been established between CDI and an impaired gut microbiome, suggesting a microbiome-based test could predict CDI recurrence. However, there are few prior studies and these suffer from serious shortcomings, including inadequate cohort sizes, limited longitudinal sampling, and no measurements of microbiome metabolic activities. In addition, due to the complexities of large, dynamic and multi-dimensional microbiome datasets, computational approaches are essential that can derive predictors from these data that are not only accurate, but also produce outputs that clinicians can understand. We therefore propose to take critical steps toward developing a microbiome-based diagnostic test for CDI recurrence by: (a) scaling up a novel human-interpretable machine learning approach that we have developed, to enable analysis of large multi-'omics datasets, and (b) applying our method to derive CDI-recurrence predictor models from our ongoing study that collects sequence-based and metabolome microbiome data from longitudinal fecal samples of 150 participants with CDI. This work is expected to provide critical preliminary results needed to obtain funding to develop and validate a diagnostic test for CDI recurrence.