Biomedical Informatics Grand Rounds

Wednesday, May 7, 2025
3:00 pm – 4:00 pm

Deriving Study-Ready Datasets from Observational Health

Johanna Loomba, ME

Director of Informatics, iTHRIV CTSA, UVA

Remote Access
Join Zoom Meeting https://stonybrook.zoom.us/j/95617197636?pwd=KytzZ2pVRG9SZGpKZUtpNXJISjNjZz09
Meeting ID: 956 1719 7636 Passcode: 924293

Bio: Johanna is committed to leading the design of data systems, in particular those related to clinical data, and identification of applicable analytic tools and methods. Johanna has a bachelors in Symbolic Systems: Neural Systems from Stanford University and a master's in Systems Engineering from the University of Virginia. She previously worked as the Director of Neurosurgery and Neuro-oncology clinical research, managing dozens of drug and device trials as well as outcomes research and providing me with bedside research experience that richly informs my understanding of patient care workflows and Real World Data (RWD) as represented in electronic health record systems. As the Director of Informatics for the integrated Translational Health Research Institute of Virginia (iTHRIV) CTSA and as the lead of the National COVID Cohort Collaborative (N3C) Logic Liaison team, Johanna is responsible for the design and development of a variety of processes and systems that accelerate translational data science by facilitating feature extraction, integrating public datasets, and transforming complex source data into machine learning ready data frames. She was also an instructor for the 2024 AIM AHEAD "Traineeship in Advanced Data Analysis Using NCATS Data and the N3C Enclave", and is passionate about supporting researchers who are new to observational health research.

Abstract: Real-world observational health data can accelerate translational science, transform healthcare and inform policy. However, due to the inherent complexity of electronic health record data, the scientific community must still overcome major hurdles when they try to derive meaningful feature sets from this data. Indeed, converting heterogeneous and complex raw data to analysis-ready tables is time intensive, error prone, and often done one project at a time, resulting in data pipelines that can mask assumptions and be difficult to reuse. This presentation will establish why research teams working with this data must grapple with data collection, mapping, and harmonization before determining whether or not their scientific question can be answered or informed by analysis of real-world data. We will uncover key considerations related to medical vocabularies, harmonization, logical feature derivation, and quality assessments. Some best practices will be shared, enabling research teams to develop their own rigorous methods and approaches to analysis of real-world health data.

Educational Objectives:
After participating in this lecture, attendees will learn and understand:

1. Limitations of observational health data ("Real World Data")
2. Advantages of harmonization to OMOP Common Data Model
3. Approaches to complex feature derivation and the value of standardized pipelines
4. How to evaluate quality and work with missing data in derived feature sets

Disclosure Statement: The faculty and planners have no relevant financial relationship with ineligible companies, whose primary business is producing, marketing, selling, reselling, or distributing health care products used by or on patients.

Continuing Medical Education Credits: The School of Medicine, State University of New York at Stony Brook, is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians. The School of Medicine, State University of New York at Stony Brook designates this live activity for a maximum of **1** ***AMA PRA Category 1 Credits™.*** Physicians should only claim credit commensurate with the extent of their participation in the activity.