conditions, but little has been done to explore these issues in large electronic health records data sources that include dental health records. Here we report on our exploration of data readiness and completeness of three of these data sources in the Clinical and Translational Science Awards (CTSA) network. METHODS/STUDY POPULATION: Three CTSAs from the Consortium of Rural States (CORES) with diverse geographies, demographics, and data ecosystems can integrate medical and dental records, but it is unknown if the target population having both dental and medical records have sufficient completeness and similarity to enable dental/medical health studies. Here we use descriptive analytics to characterize the demographics, and the "complete data" approach presented by Weber et al. to evaluate differences between the completeness of the general populations and the one having both dental/medical records. We accomplish this by identifying patients with dental records in commonly used research networks and performing empirical patient statistics in comparison to the entire population available at the three institutions. RESULTS/ANTICIPATED RESULTS: This poster will present the results of using the Weber et al. approach to compare the completeness of records of the general patient population in the Iowa, Kentucky, and Utah medical/dental health care systems to those for which they have also dental records. The completeness of the records of these two subpopulations is also associated with different demographic characteristics, as it has been established that the populations served by the dental clinics is biased by dental insurance considerations. The work will show what retrospective studies can (or not) be done using these populations when taking into account that it is well established that studies of populations with different level of completeness can be inconsistent. DISCUSSION/SIGNIFICANCE OF IMPACT: This study provides an informatics framework to assess similarity and completeness of patient records with and without dental records. Establishing the level of similarity and completeness in these patient populations is critical to justify the validity of studies that utilize a combined record.

358

Rare disease study identification (RDSI): A natural language processing assisted search and visualization tool for clinical studies of rare diseases

Michael Lin[1], Jennifer Weis[1], H M Abdul Fattah[2] and Jungwei Fan[1]
[1]Mayo Clinic and [2]University of Arizona

OBJECTIVES/GOALS: Identifying and indexing rare disease studies is labor intensive, especially in research centers with a large number of trials. To address this gap, we applied natural language processing (NLP) and visualization techniques to develop an efficient pipeline and user-friendly web interface. Our goal is to offer the rare disease study identification (RDSI) tool for adoption by other sites. METHODS/STUDY POPULATION: The RDSI retrieves study information (short and long titles, study abstract) from the IRB system. These descriptive fields are then processed by the MetaMap Lite NLP program for identifying disease terms and standardizing them to UMLS concepts. By terminology identifier mapping, the diseases intersecting with concepts in rare disease databases (Genetic and Rare Disease program and Orphanet) are further scored to pinpoint studies that focus on a rare disease. The web interface displays a scatter bubble chart as an overview of all the rare diseases, with each bubble size proportional to the number of studies for that disease. In addition to the visual navigation, users can search studies by disease name, PI, or IRB number. Search results contain detailed study information as well as the evidence used by algorithms of the pipeline. RESULTS/ANTICIPATED RESULTS: The RDSI identification results and functions were verified manually and spot-checked by several study investigators. The web interface is a self-contained solution available to our staff for various use cases like reporting or environment scan. We have built in a versioning mechanism that logs the date of each major result in the process. Therefore, even as the rare disease data sources evolve over time, we will be able to preserve any historical context or perform updates as needed. The RDSI outputs are replicated to Mayo Clinic's enterprise data warehouse daily, allowing tech-savvy users to leverage any useful intermediate results at the backend. We anticipate the performance of the rare disease identification to be further enhanced by employing the advancements in AI technology. DISCUSSION/SIGNIFICANCE OF IMPACT: The RDSI represents an informatics solution that offers efficiency in identifying and navigating rare disease clinical studies. It features the use of public databases and open-source tools, manifesting return on investment from the broad translational science ecosystem. These considerations are informative and adoptable by other institutions.

359

Differentiating opioid use disorder from healthy controls via ML analysis of rs-fMRI networks

Ahmed Temtam[1], Megan A. Witherow[1], Liangsuo Ma[2], M Shibly Sadique[1], F. Gerard Moeller[4], C. Kenneth, Dianne Wright[5] and Khan M. Iftekharuddin[6]
[1]Vision Lab, Dept. of Electrical Engineering, Old Dominion University, Norfolk, VA, USA; [2]Institute of Drug and Alcohol Studies, Richmond, VA, USA, Department of Psychiatry; [3]Virginia Commonwealth University, Richmond, VA, USA; [4]Institute of Drug and Alcohol Studies, Richmond, VA, USA, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA, Department of Neurology, Virginia Commonwealth University, VA, United States; [5]Center for Clinical and Translational Research, Virginia Commonwealth University, VA, United States and [6]Dept. Of Electrical Engineering, Old Dominion University, Norfolk, VA, USA, Data Science Institute, Old Dominion University, Virginia Beach, VA, USA

OBJECTIVES/GOALS: This work aims to identify functional brain networks that differentiate opioid use disorder (OUD) subjects from healthy controls (HC) using machine learning (ML) analysis of resting-state fMRI (rs-fMRI). We investigate the default mode network (DMN), salience network (SN), and executive control network (ECN), as well as demographic features. METHODS/STUDY POPULATION: This work uses high-resolution rs-fMRI data from a National Institute on Drug Abuse study (IRB #HM20023630) with 31 OUD and 45 HC subjects. We extract rs-fMRI blood oxygenation level-dependent (BOLD) features from the DMN, SN, and ECN. The Boruta ML algorithm identifies statistically significant features and brain activity mapping visualizes regions of heightened neural activity for OUD. We conduct fivefold cross-validation classification experiments (OUD vs. HC) to assess the discriminative power of functional network features with and without incorporating demographic features. Demographic features are ranked based on ML classification importance. Follow-up Boruta analysis is performed to study the medial prefrontal cortex (mPFC), posterior cingulate

https://doi.org/10.1017/cts.2024.984