Name: Alvin Jeffery Email: alvin.d.jeffery@vanderbilt.edu

Detecting Opioid Use Disorders in the Electronic Health Record to Support Genetic Studies: Adaptation of the Addiction Behaviors Checklist

Eli Bradley[1]; Angus Hayes Chatham[1]; Lori Schirle[1]; Sandra Sanchez-Roige[2,3]; David C. Samuels[4]; Alvin D. Jeffery[1,5]

[1]Vanderbilt University School of Nursing; [2]Department of Psychiatry, University of California San Diego; [3]Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center; [4]Department of Molecular Physiology and Biophysics, Vanderbilt Genetics Institute, Vanderbilt University; [5]Department of Biomedical Informatics, Vanderbilt University Medical Center

Individuals whose chronic pain is managed by opioids are at high risk of developing opioid use disorders. Because large data sets (e.g., electronic health records) are required for conducting well-powered genetic studies, we need high-throughput phenotyping methods. The most used approach (i.e., administrative billing codes) has a high degree of misclassification. Our prior work leveraged the natural language processing approach of concept extraction, which resulted in more accurate identification of opioid use behaviors. Given the computational complexity of concept extraction, in this work, we sought to examine whether a simpler approach that identifies keyword/pattern matches in text notes would provide similar improved performance while being more readily implemented and interpreted. The Addiction Behaviors Checklist (ABC) is a validated instrument designed to identify prescription opioid-related addiction behaviors among people with chronic pain. The objective of this pilot study was to leverage natural processing techniques to automate the ABC in order to expedite research or clinical chart reviews. Using clinical notes in a retrospective cohort of 5,697 individuals with chronic pain, we converted the Checklist into regular expressions with iterative input from subject matter experts. The automated approach identified individuals with problematic opioid use that were missed by diagnostic codes (area under the curve = 0.72 vs. 0.52 and F1 score = 0.66 vs. 0.08). This automated data extraction technique can facilitate the identification of people at-risk for, and suffering from, opioid use disorders, which will create new opportunities for more accurate high-throughput phenotyping to support genetic studies.