

Name: Alvin Jeffery

Email: a.jeffery@vumc.org

Precision Phenotyping of Opioid Use Disorder for Multiple Populations

Chen Jin, BS¹; Yang Cao, BS¹; Dagmawi Negesse, MS²; Parker Christy, BS²; Lori Schirle, PhD¹; Sandra Sanchez-Roige, PhD^{3,4}; David C. Samuels, PhD³; Alvin D. Jeffery, PhD^{1,2,3}

¹School of Nursing, Vanderbilt University; ²Department of Biomedical Informatics, Vanderbilt University Medical Center; ³School of Medicine, Vanderbilt University; ⁴Department of Psychiatry, University of California San Diego

To support opioid use disorder (OUD) genetics studies, we need computable phenotypes due to the resource-intensive nature of manual chart reviews. Traditional phenotyping relies on administrative billing codes (which have poor recall and precision). In this study, we built on our prior work of developing OUD phenotypes from clinical notes by exploring whether gender or racial differences in documentation content could affect phenotyping performance.

We analyzed text notes using the natural language processing package ScispaCy for 7,999 individuals with chronic pain at Vanderbilt. We calculated the salience of all extracted entities using the Term Frequency-Inverse Document Frequency formula and compared the highest-ranked entities between sub-groups. We focused our analysis on Female and Male genders and Black and White racial groups, as listed in the electronic health record.

We found many differences in entities' rank-ordered salience between sub-groups. For example, both "sickle cell disease" and "cocaine dependence" were more salient for Black individuals with OUD compared to White individuals. "Human immunodeficiency virus" was more salient in Males than Females. "Pregnancy" was not found in the charts of Males. "Suboxone" was highly salient for all sub-groups except Black individuals. "Gastric bypass" was more salient for Females and White individuals while "coronary artery bypass graft" was more salient for Black individuals.

While some discrepancies have a biological rationale (e.g., "pregnancy" absent in Males' charts), other discrepancies are more concerning. Variances might arise from socioeconomic factors and/or clinicians' implicit biases. We are updating our prior scoring system to evaluate whether performance is improved.