

Submitter Name: Lori Schirle  
Submitted email: [Lori.schirle@vanderbilt.edu](mailto:Lori.schirle@vanderbilt.edu)

## **A Novel Data-Driven Approach to Evaluate the Spectrum of Opioid Use in Electronic Health Data**

Lori Schirle<sup>1</sup>, Alvin Jeffery<sup>1,2</sup>, Ali Yaqoob<sup>1,3</sup>, Sandra Sanchez-Roige<sup>4,5</sup>, David Samuels<sup>6</sup>

<sup>1</sup>School of Nursing, Vanderbilt University; <sup>2</sup>Department of Biomedical Informatics, Vanderbilt University; <sup>3</sup>Data Science Institute, Vanderbilt University; Department of Psychiatry, University of California, San Diego; <sup>4,5</sup>Department of Medicine, Division of Genetic Medicine, Vanderbilt University; <sup>6</sup>Department of Human Genetics, Vanderbilt University

**Background:** Problematic opioid use does not occur in a simple case-control manner, but on a continuum. Early work to identify problematic opioid use has primarily focused on extremes of opioid use. There is an urgent need to expand this work to capture the *spectrum* of problematic opioid use.

**Rationale/significance:** Electronic health record (EHR) data offer untapped opportunities to extract risk severity. Administrative data contain information on clinical diagnoses that coexist with problematic opioid use, particularly psychiatric disorders and pain conditions, along with invaluable information from clinical notes.

**Hypothesis:** We hypothesized that combining data-driven opioid use disorder (OUD) comorbidities and EHR text converted into standardized medical vocabulary can serve as a new framework to identify a continuum of problematic opioid use.

**Methods:** We compared two novel data-driven methods to gold-standard manual review in a high-OUD risk cohort of 5,697 chronic pain patients from the Vanderbilt Biobank: 1) a logistic regression-based OUD comorbidity risk score generated from OUD ICDs, and 2) an innovative text-based risk score capturing word importance in clinical notes.

**Results:** We achieved an area under the curve (AUC) of .79 for the comorbidity risk score, and .80 for the text-based risk score. The Spearman's Correlation Coefficient between these 2 scores was .53.

**Discussion:** Both approaches give a continuous measure of problematic opioid use severity. The moderate correlation suggests unique information in each risk score that can be combined in future studies to build a robust risk prediction model to enhance OUD detection and prevention research.