Using Electronic Health Records and Machine Learning to Assess Opioid Addiction Risk

Philip J. Freda, Ph.D. and Jason H. Moore, Ph.D.
Department of Biostatistics, Epidemiology, & Informatics, University of Pennsylvania, Philadelphia, PA 19104

**Goals**
- Use the Penn EHR and Penn Medicine Biobank (PMBB) to determine significant risk factors for Opioid Use Disorder (OUD).
- Create risk assessment protocols for clinicians when prescribing opioid-based analgesics.
- Reduce the number of individuals that progress from opioid misuse to opioid dependence and abuse.

**Introduction**
The U.S. opioid epidemic persists and has become an increasingly important issue worldwide. Each day in the United States, more than 130 people die from opioid-related overdose. Four factors that can act either additively or synergistically to increase the risk of developing OUD are psychiatric disorders, other drug use disorders, genetic predispositions, and environmental stressors, which may include chronic pain and past adverse experiences.

Data from genomic biobanks, electronic health records (EHR), and structured interviews/assessments can be utilized by researchers to identify these risk factors.

It is our goal to determine which factors and genetic variants are associated with OUD using the Penn EHR and the Penn Medicine Biobank (PMBB). This information will be used to develop risk assessment frameworks for opioid misuse, dependence, and abuse using machine learning and artificial intelligence.

Our hope is that this research will lead to clinicians being more informed about risk of addiction before administering opioid-based analgesics to patients.

**Workflow**
- **EHR**
- **PMBB**
- **SNPs eQTLs**
- **Diagnoses Procedures**
- **Clinic Notes**

**Implementation**
- **Informed Treatment**
- **Risk Assessment**

**First data pull:** 230,840 Individuals with drug abuse 2012-2014: 991 2014-2016: 2,029 2016-2018: 4,345

**Results & Discussion**
1. The rate of OUD prevalence in the Penn EHR has more than quadrupled from 2012 to 2018 despite greater awareness of the addictive qualities of opioid-based analgesics.
2. In this preliminary dataset, consisting of individuals with drug abuse and/or psychiatric phenotypes, the rates of several common psychiatric phenotypes are consistent between the total cohort and those diagnosed with OUD. This could mean that similar risk factors contribute to a host of psychiatric/drug abuse disorders.

**Future Research**
1. Use machine learning and NLP on all data points (including genomics data from the PMBB) from this dataset and a control cohort to determine risk factors of OUD.
2. Extend our findings to other types of data including structured psychiatric assessments.

**Literature Cited**