

# Identification of Subphenotypes of Opioid Use Disorder Using Unsupervised Machine Learning

Fatemeh SHAH-MOHAMMADI<sup>1</sup> and Joseph FINKELSTEIN  
*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

**Abstract.** This paper aimed to detect the latent clusters of patients with opioid use disorder and to identify the risk factors affecting drug misuse using unsupervised machine learning. The cluster with the highest proportion of successful treatment outcomes was characterized by the highest percentage of employment rate at admission and discharge, the highest percentage of patients who also recovered from alcohol and other drug co-use, and the highest proportion of patients who recovered from untreated health issues. Longer participation in opioid treatment programs was associated with the highest proportion of treatment success.

**Keywords.** Opioid use disorder, Machine learning, Subphenotyping.

## 1. Introduction

Optimal opioid use disorder (OUD) treatment in opioid treatment programs (OTP) should be tailored to individual patient profiles [1]. The goal of this study was the identification of subphenotypes of patients with OUD using heterogeneous data sources generated by OTP clinics and mapping the uncovered latent classes to OTP outcomes.

## 2. Methods

The study cohort comprised patients with OUD undergoing treatment at OTP clinics at the Mount Sinai Health System (MSHS) in New York City. The study dataset was generated by aggregation, and harmonization of patient information from the New York State Office of Addiction Service and Supports (OASAS), OTP clinic software for patient management (AVATAR), and MSHS electronic health records (Epic). The analytical dataset comprised patients' socio-demographic information, clinician notes, drug urine toxicology screens, employment status, admission, transfer, and discharge records, and state forms. The discharge form contained information on the patient's drug and alcohol recovery status, and data about the medical and legal issues. OTP outcome was defined based on discharge summary, which included discharge status as follows: "completed treatment: all treatment goals met", "completed treatment: half or more goals met", "treatment not complete: some goals met," and "treatment not complete: no goals

---

<sup>1</sup> Corresponding Author: Fatemeh Shah-Mohammadi, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, 2<sup>nd</sup> Fl, New York, NY, USA, 10035; E-mail: fateme.sh.mohammadi@gmail.com.

met". The patients with the first two discharge statuses were considered as patients with successful treatment, and patients with the last two statuses were labeled as failed treatment cases. We used the factor analysis for mixed data (FAMD) method for preprocessing [2].

### 3. Results and Discussion

Three clusters have been identified. There was a significantly larger number of male patients compared to females. Most patients in all clusters were 45 years old and older. All patients who succeeded in treatment belonged to cluster # 2. Compared to the patients in clusters # 0 and 1, patients in cluster # 2 had the highest employment rate at admission and discharge (18.47% and 19.79%, respectively). The percentage of patients who recovered from alcohol and drug abuse (43.47% and 86.97%, respectively) was significantly higher in cluster # 2. The same scenario was valid in meeting the legal and medical goals for patients in cluster # 2. Compared to other clusters, the rate of patients who adhered to the treatment for over 2 years was almost twice as high in cluster # 2. Since the lower percentage of employment at admission and discharge, achievement of alcohol, drug, medical and legal goals belonged to the clusters with the largest percentage of failed patients (i.e., clusters # 0 and 1), patients in these clusters need more rigorous treatment, social support, and dedicated personal attention. This finding conveys the fact that OTP patients with alcohol co-use are at risk of worse treatment outcomes. Due to the increase in employment rate from admission to discharge among the patients in the cluster with the highest successful treatment rate, it can be assumed that employment acts as a facilitator for adhering to treatment for opioid use disorders and can result in better treatment outcomes. This finding is aligned with many studies claiming that unemployment is a significant risk factor for substance use and the subsequent development of substance use disorders. Due to the highest rate of patients with more than 2 years of participation in OTP in cluster #2, it can be concluded that the longest enrollment in OTP is associated with the highest successful treatment rate.

### 4. Conclusion

Three clusters have been identified for patients who participated in OTPs. The largest percentage of successfully treated patients belonged to the second cluster, with patients mainly employed. This conveys that unemployment acts as a risk factor for the treatment failure of opioid use disorders. This cluster also has the highest percentage of patients who recovered from untreated health issues from alcohol and other drug co-use. The longest adherence to the treatment is associated with the highest successful treatment rate. Subphenotyping of patients with OUD may help personalize their treatment.

### References

- [1] McCarthy M. Opioid prescribing rates in US vary widely between states, CDC reports. *BMJ*. 2014 Jul 4;349:g4424.
- [2] Mori Y, Kuroda M, Makino N Multiple Correspondence Analysis. In: *Nonlinear Principal Component Analysis and Its Applications*. Springer Briefs in Statistics. Springer, Singapore.2016.