STANFORD
SCHOOL OF MEDICINE
Stanford University Medical Center

A Knowledge Graph-based Approach for Exploring the U.S. Opioid Epidemic

Maulik R. Kamdar, Tymor Hamamsy, Shea Shelton, Ayin Vala, Tome Eftimov, James Zou, Suzanne Tamang

Department of Biomedical Data Science AT STANFORD MEDICINE

Motivation

- The opioid abuse epidemic is one of the most challenging US public health challenges, with 64,000 overdose deaths in 2016 and 2 million people currently addicted.
- To understand US opioid prescription trends and to predict possible opioid abuse behavior in patients, statistical models should use clinical data from multiple healthcare sites in the US.
- Challenges: Vast semantic heterogeneity exists between different clinical systems, and the lack of use of standard terminologies to encode clinical features (e.g., patient medications).
- Current Methods: Raw patient data, extracted from legacy databases, and transformed under a uniform representation format (e.g., FHIR, OMOP) for use in machine learning models. Burden for the clinical centers, and creates multiple copies of private and secure patient data.
- Knowledge Graphs (large directed networks of entities and relations, with a fixed set of semantic classes and properties) can aid in the task of normalization of similar entities encoded using different identifiers and enable integration of data from multiple heterogeneous sources.

Materials

- Terminologies: (retrieved from the BioPortal repository of biomedical ontologies)
  $\mathcal{T}_1$: ATC - Drug ingredients classified by anatomical, therapeutic, and chemical properties.
  $\mathcal{T}_2$: RxNorm - Standard names for clinical drugs and dosage forms, as well as relations between clinical drugs to their active ingredients, drug components, and brand names.
- Unified Medical Language System (UMLS): Concept Unique Identifier (CUI) mappings between classes having similar meanings in different ontologies.
- Electronic Medical Record (EMR) Data Warehouse: De-identified EMR data collected from more than 400 hospitals and healthcare facilities (87% Urban and 13% Rural) from across 42 states in US (59% South, 17% West, 13% Midwest, 12% Northeast) during 2009-2016, and aggregated and stored in the Google BigQuery Analytics data warehouse.

Opioid Drug Knowledge Graph (ODKG): Methods and Characteristics

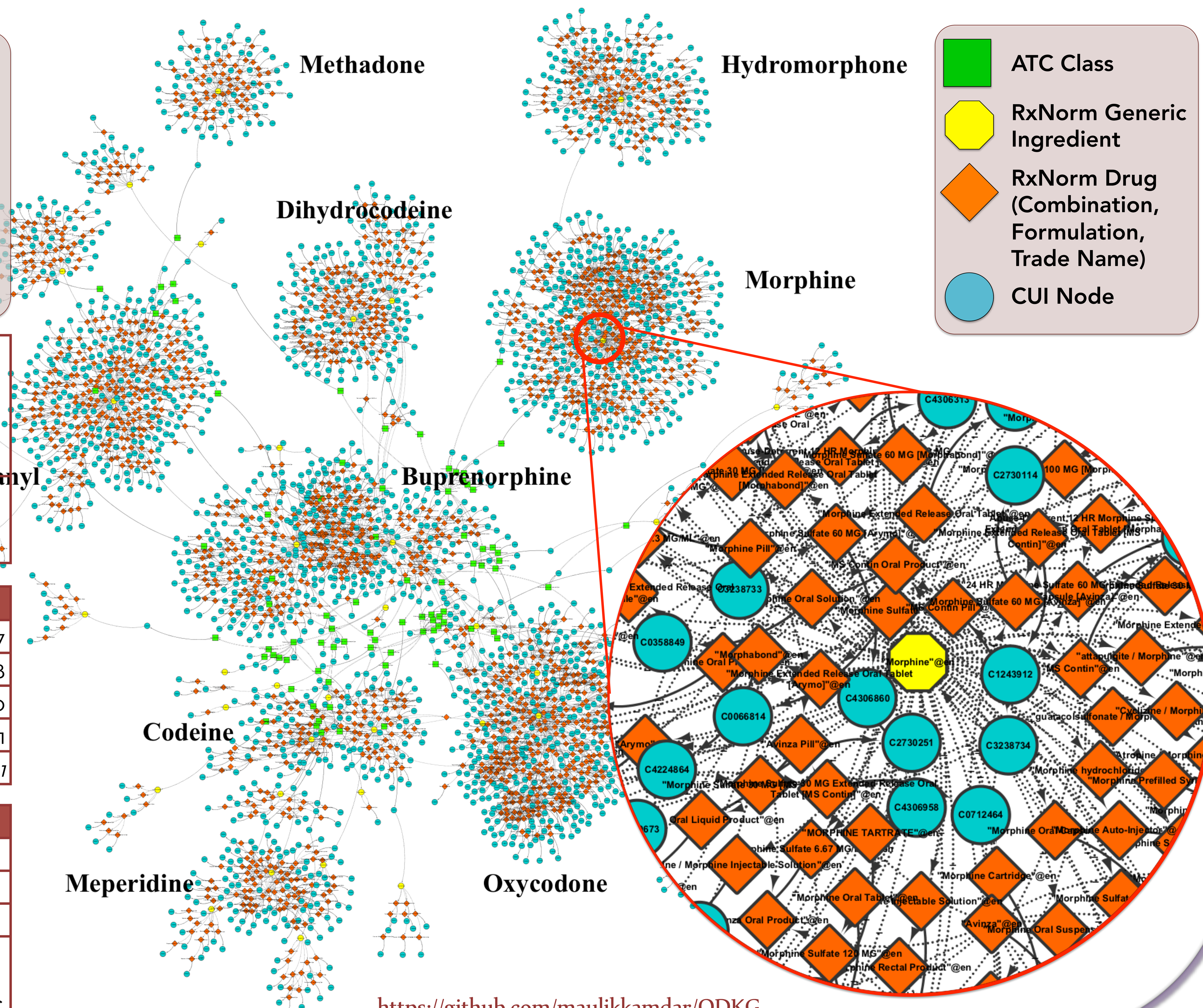
Steps to generate the ODKG

- ❑ Use Hierarchical Reasoning to retrieve all descendants (i.e., drug active ingredients) of the Opioid-related base classes in the ATC terminology.
- ❑ Use UMLS CUI mappings to retrieve similar active ingredients of opioid drugs from the RxNorm terminology.
- ❑ Use a fixed set of properties to retrieve semantic relations between opioid generic ingredients and other classes (e.g., drugs, combinations, formulations, and tradenames) from the RxNorm terminology.
- ❑ Use MedEx to parse drug strings in the aggregated EMR database and instantiate the strings under different drug classes using CUIs and RxCUIs.

ATC Base Classes	(N02A) Opioid analgesics, (N01AH) Opioid anesthetics, (R05DA) Opium alkaloids and derivatives, (N07BC) Drugs used in opioid dependence (A06AH) Peripheral opioid receptor antagonists
ODKG Properties	*Has CUI, Has RxCUI, SubClass Of, Ingredients Of, Has Form, Form Of, Part Of, Ingredient Of, Consists Of, Constitutes, Has Tradename, and Precise Ingredient Of*

		Count
Nodes	*ATC Class*	97
	RxNorm Generic Ingredient	48
	RxNorm Drug (e.g., Combination, Tradename)	4960
	Concept Unique Identifier	5051
Edges	*hasCUI, or any of the above ODKG properties*	13581

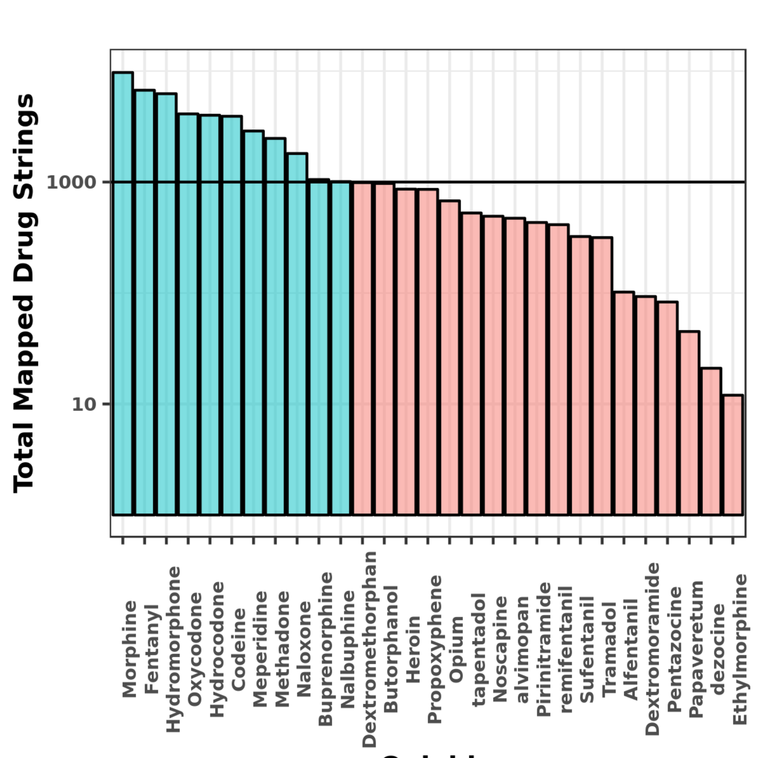
Property Type	Classes Linked to Morphine at First Degree Hops
Part Of	*Atropine / Morphine, Cyclizine / Morphine*
Has Tradename	*MSContin, EMbeda, Avinza, Duramorph, Kadian*
Has Form	*Morphine Hydrochloride, Morphine Sulphate*
Ingredient Of	*Morphine Injectable Solution, Morphine Prefilled Syringe, Morphine / Naltrexone Extended Release Oral Tablet, Morphine Sulfate 20 MG/ML, Morphine hydrochloride 40 MG*

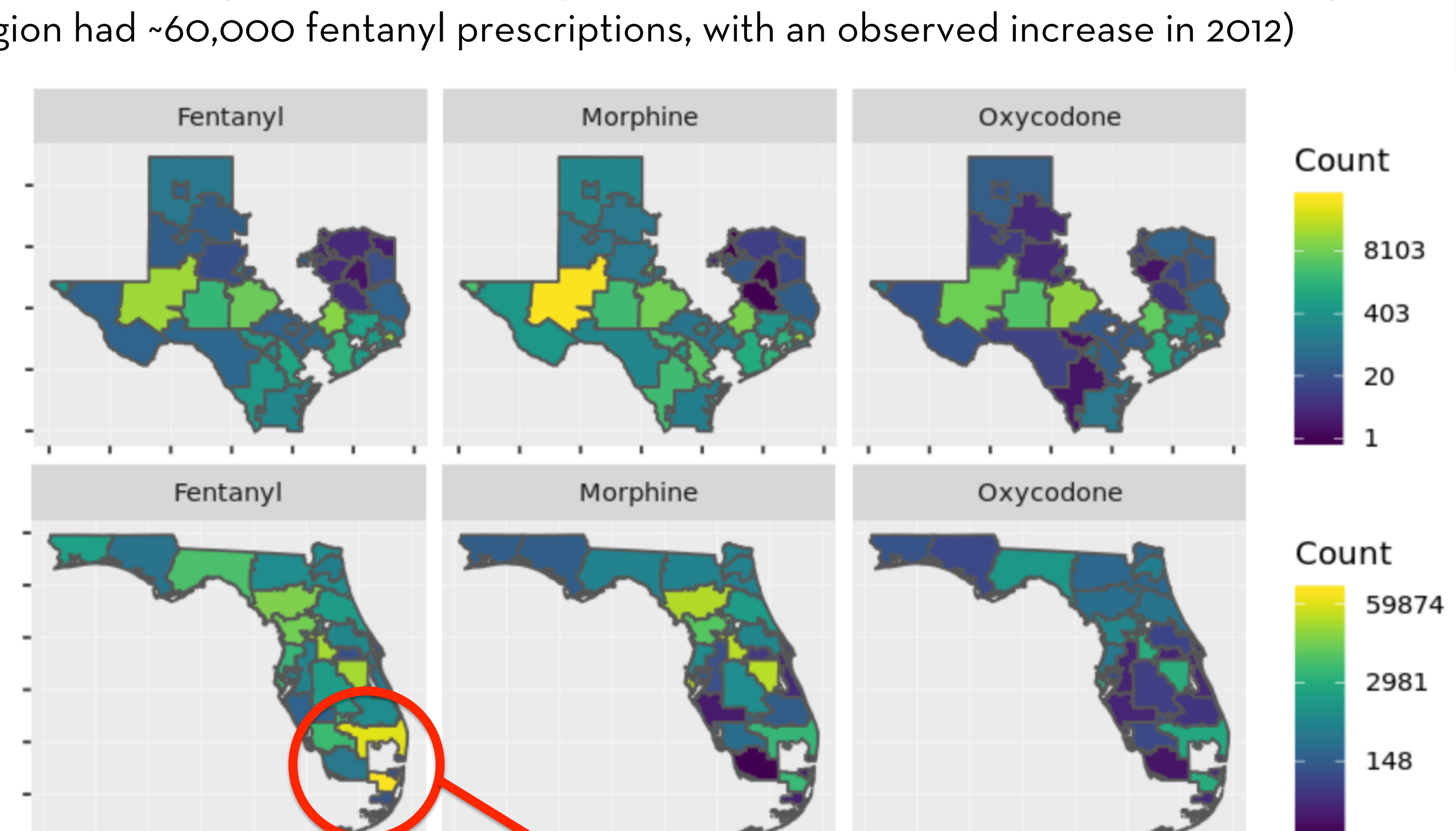


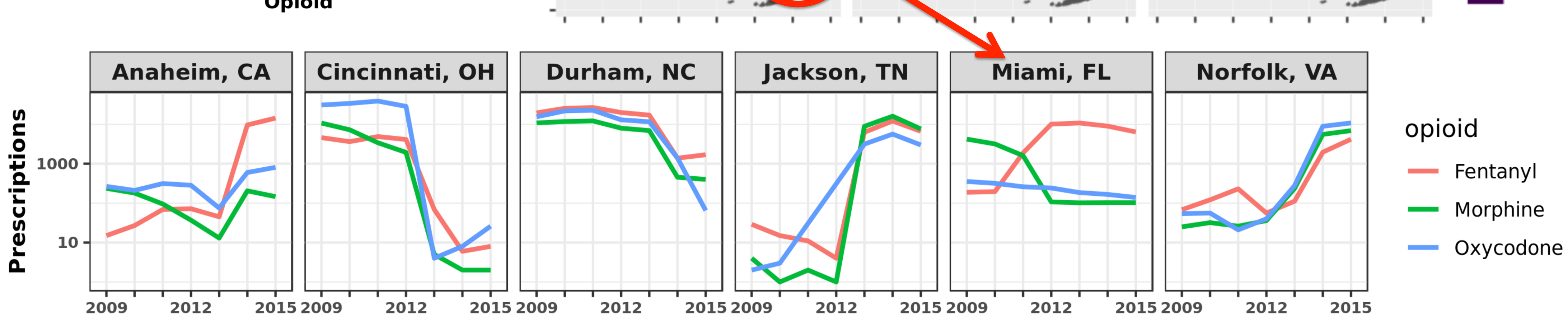
https://github.com/maulikkamdar/ODKG

Opioid Summary Statistics over Location and Time

- 425,059 unique drug strings in EMR data warehouse are parsed using MedEx, out of which 288,983 drug strings are mapped to at least 1 CUI (68% coverage), and 374,208 drug strings are mapped to at least 1 RxCUI (88% coverage).
- 29 opioid-related active ingredients in the ODKG have more than 10 drug strings from the EMR data warehouse.
- Opioid prescriptions can be categorized according to different US regions and time periods, as determined through patient admission year (e.g. Miami region had ~60,000 fentanyl prescriptions, with an observed increase in 2012)







Future Work

- The ODKG will be used to develop a Web-based tool that can facilitate visualization of historical patterns and can enable comparisons across opioids, time, and US regions.
- We hope to further refine the ODKG by consulting with a domain expert and tailor it for specific use cases and end users.
- We plan to compare our approach against the OMOP-based approach of transformation of clinical data.
- We will explore other potential areas of application, including:
  - Develop dynamic phenotyping methods to visually analyze individual pain medication use profile.
  - Identify potential risk factors for long-term use of opioids.
  - Detect adverse outcomes for incident user of prescription opioids, specific to the surgical setting.

Conclusion

- To identify the best strategies to reduce opioid over-prescription and misuse, a better understanding of country and regional consumption patterns, pharmaceutical industry influences, and sociopolitical factors that impact consumption, is needed.
- Heterogeneous drug names and drug composition pose a significant challenge in developing machine learning models over heterogeneous clinical data from multiple centers to study the opioid epidemic.
- We develop Opioid Drug Knowledge Graph to capture how opioid drugs relate to each other, and to translate medications from diverse electronic medical records into a common set of chemical-dosage features to enable a large number of prediction and modeling tasks.