**Introduction**

Prescription opioid abuse plays an essential role in public health issues. The price of prescription opioids indicates the supply-demand relationship of drugs. This study case aims to explore the relationship between drugs’ unit price and other factors. More specifically, our group’s interest is to explore the factors related to the cost per milligram and the heterogeneity in the region. The dataset is provided by StreetRx, a reporting tool for people at large to anonymously report the price they paid or heard for diverted prescription drugs.

Our drug interest is Morphine. Morphine is used to “relieve moderate to severe pain and maybe habit-forming,” especially with prolonged use (MedlinePlus).

**Data cleaning**

The dataset (Morphine) contains 9268 observations with 13 variables. There are several missing values in the dataset. For example, the feature `primary reason` has 5061 empty cells. Due to the high missing rate ($5061/9268 \approx 54.61%$), our group decided to remove `primary reason`. Moreover, we deleted all observations containing NA.

Price per milligram (ppm)

Whether in a hierarchical model or linear regression, the response variable has to be normally distributed. From the histogram on the left, the `ppm` distribution has a long tail on the right (extremely right-skewed). And the Box-Cox transformation suggests doing a log transformation. The distribution looks roughly normal therefore we will use log(ppm) as the dependent variable.

Date

Two ways