

**Causal Theory Underlying Analysis**

Demographic features such as poverty status, whether a person is middle aged, education level, employment status, income level, gender and health facility accessibility may be associated with higher incidences of opioid use in census tracts within Tempe, AZ and Cincinnati, OH. In our study, we extract these demographic features from the United States Census’ American Community Survey five-year (ACS5) estimates and respectively compare them to the number of opioid related EMS calls in Tempe, AZ and Cincinnati, OH. We will stop short of calling any relationship found, “causal”, and instead argue that these associations may represent risk factors that could lead to higher incidences of opioid use within the city.

**Pilot data demonstrating statistical relationship, including temporal precedence and covariance between proposed cause and effect**

|  |  |
| --- | --- |
|  |  |

**What are the threats to internal validity?**

*Potential Self-Selection*

The only potential selection-x threat we could face is selection-mortality. Subjects in high opioid use collection areas for example may have a greater incentive to move out of the sampling frame during our study than subjects in areas that do not face these issues.

While we do think that the experimental mortality and the selection-Mortality threat could be plausible, we do not expect it to have a significant impact on our study. Unless there is a major population shift from one area to another or a major health event (i.e. overdose spike) we do not expect our sample to change much during our study. We consider this particularly unlikely because we are looking at only a few years of data where we do not expect sample characteristics to change much absent a significant event.

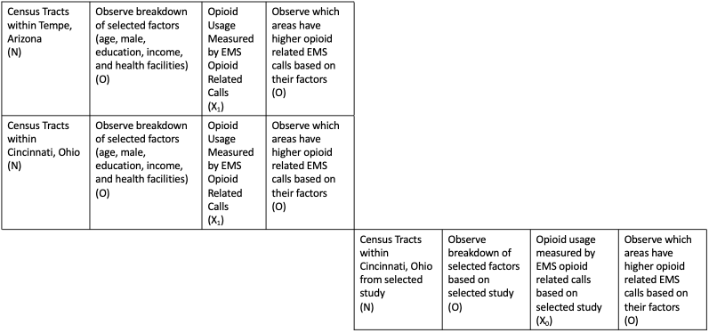
As for Selection-X threats as a whole, the areas are closely located and we can expect some spillover between and among areas, in which case the groups/areas will not have significantly different characteristics themselves, in which case, the difference in opioid concentration in different areas is not likely caused by the characteristics internal of the groups. We plan to address these effects by including spatial regressors in our analysis.

*Prescribed - Unprescribed Opioid Consumption Distinction*

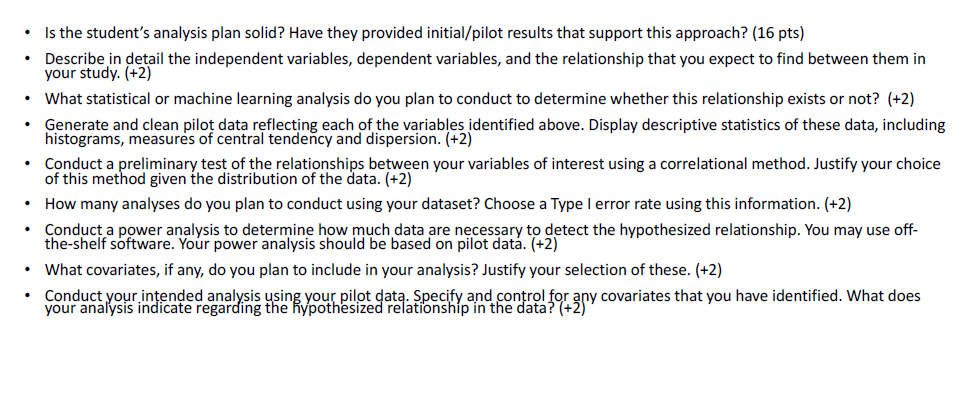
**Experimental design to rule out plausible threats assuming you could manipulate all units of analysis and conduct random assignment. Your design should allow you rule out all plausible confounds**

**Propose a combination of quasi-experimental designs you might use to mitigate the remaining threats to internal validity. Discuss how it would rule out other measures the size of remaining threats.**

Demonstrate your proposed experimental or quasi-experimental design on pilot data



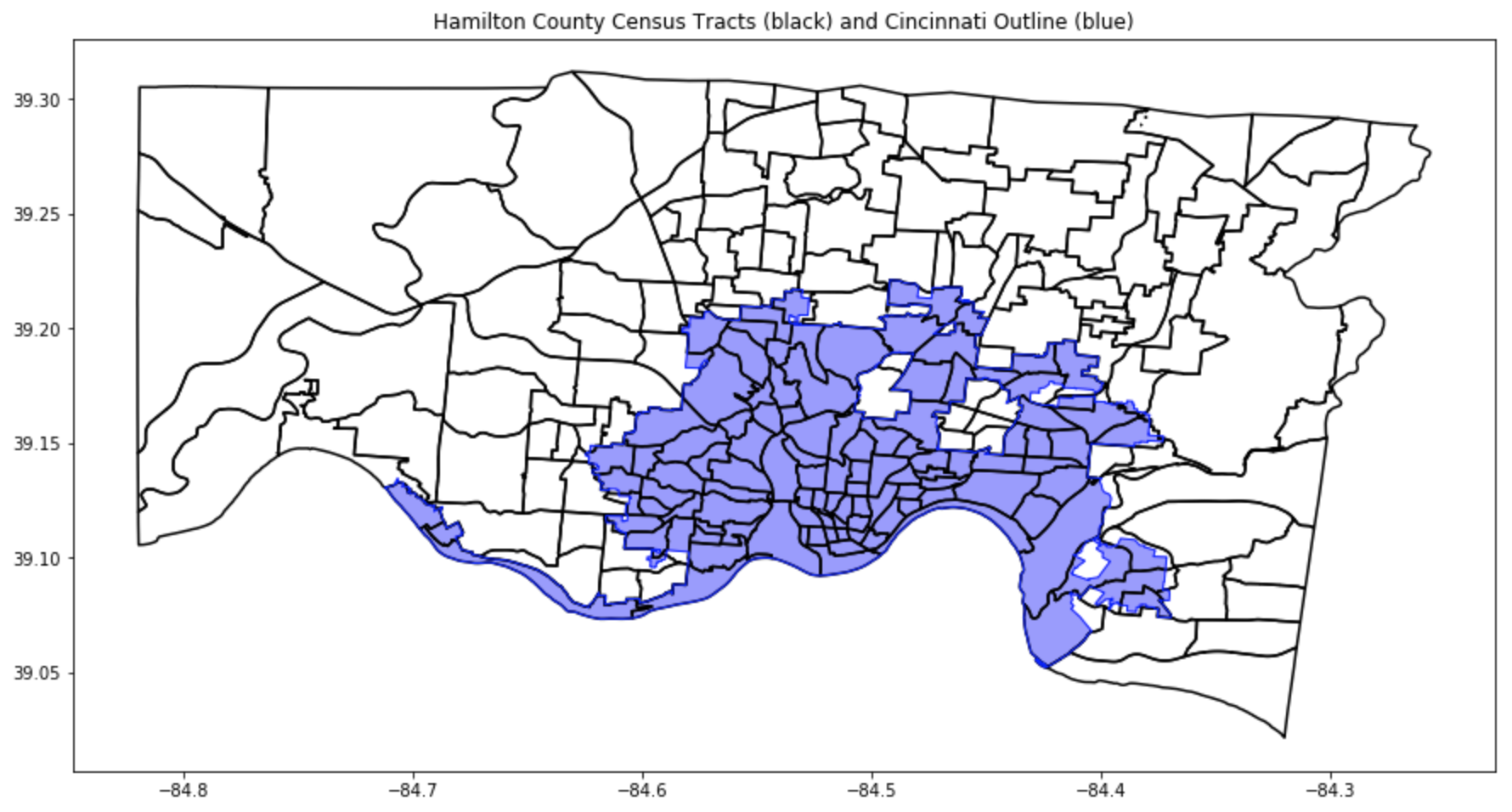
We propose using an equivalent materials quasi-experimental design to test the effects of demographic features on opioid use in both Tempe and Cincinnati. The design will seek to test the effects of the same socio-economic characteristics on opioid related calls in both cities. Our design does not have an explicit control group, a determination of a consistent relationship among these demographic factors compared against a selected study Li et al can bolster our claim that certain risk factors are enduring across both cities.



**Describe in detail the independent variables, dependent variables and the relationship that you expect to find between them in your study**

We select independent variables that represent the location of public health interventions across cities that could have an impact on the number of drug users in a given area. These include the number of medical facilities, pharmacies, drug dropoff locations, and naloxone distribution centers within two miles of each spatial unit in Tempe, Arizona and Cincinnati, Ohio.

We use census tracts as our spatial unit for Tempe and Statistical Neighborhood Approximations (SNA) for the spatial unit for Cincinnati. We used the SNA instead of the census tracts in this instance because the tracts do not overlay well with the border of the city of Cincinnati. Please see Figure 1.1 for reference. The white area is Hamilton County census tracts (black outlines) and the purple area is the union of the Cincinnati city and the census tracts. It shows that some census tracts within Cincinnati that also represent areas outside of Cincinnati.

Fig 1.1

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| EMS\_Calls | Dependent | The count of Opioid related EMS calls recorded per spatial unit |
| Population Normalized Mass Load (PNML) | Dependent | Average PNML per site collection per month per year |
| MedFacilitiesCount2mi | Independent | The count of medical facilities within 2 miles of each census tract |
| PharmCount2mi | Independent | The count of pharmacies within 2 miles of each census tract |
| DurgDropCount2mi | Independent | The count of drug drop off locations within 2 miles of each census tract |
| NaloxoneDistribCount2mi | Independent | The quantity of reported naloxone distribution count within 2 miles of each census tract |
|  |  |  |
|  |  |  |
|  |  |  |

Expected Relationships

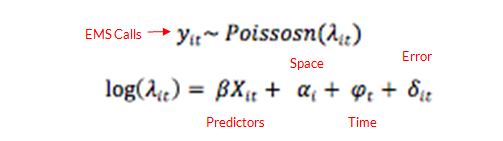
Spatial predictors we have included to have varying degrees of impact on our constructs of opioid use both in terms of magnitude and direction. For example, we expect a negative correlation between the number of medical facilities in an area and drug use because treatment is more readily accessible. However, because of the way our constructs are measured, the actual magnitude and sign of this relationship may not be as well represented. EMS calls could be negatively correlated with the number of medical facilities in an area if one believes that they act as depressants on overall drug use and represent treatment centers within an area. These same issues occur when hypothesizing the relationship between the medical facilities and the PNML measures. Medical waste that is dumped at a higher level near these areas may actually show a positive association with the PNML of the opioids in wastewater in an area, even though we expect these facilities to represent treatment.

Likewise, while we also expect naloxone distribution centers and drug dropoff centers to represent effective treatments that lower the amount of drug use in an area over time, there could be an association problem in our data. Areas with greater numbers of naloxone distribution centers or drug drop offs may have already had a high number of drug users in them to begin with, which led city officials to install these interventions at a greater rate. For the purposes of this analysis, we assume these types of treatments are effective and expect our model to detect the hypothesized long term effects of lowering drug use as measured by both EMS calls and PNML measurements in an area in a point in time. If this assumption proves to not hold, we may have to turn to intervention or time-series based (or both) analyses of when these interventions were introduced to understand this relationship.

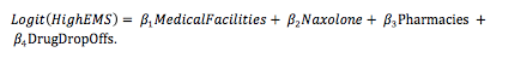
Finally, it is also difficult to guess how the number of pharmacies in an area will affect EMS calls. Because pharmacies may also produce medical waste, we believe they could have a positive effect on the PNML measurements of opioids in the area’s wastewater. The relationship is less clear for EMS calls. Pharmacies do act as treatment centers for opioids by administering naloxone and providing drug drop locations (depending on the city or state). But, pharmacies also prescribe and release more opioids into a given area to patients. Having more drugs in an area can be correlated with higher amounts of drug use.

What statistical or machine learning analysis do you plan to conduct to determine whether a relationship exists or not.

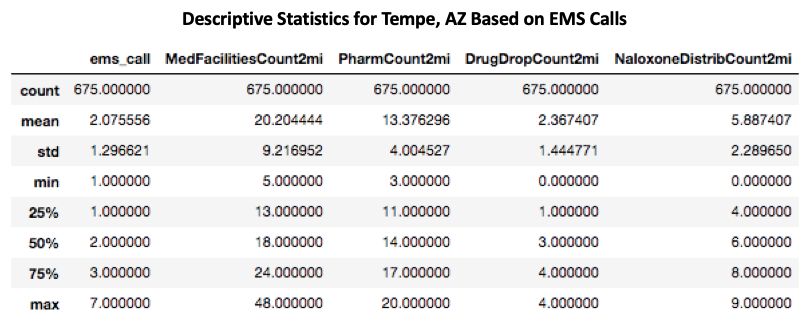
We will conduct a poisson regression for the EMS calls to accommodate the distribution of our dependent variables. Our analyses will attempt to determine the effect of the independent spatial predictors listed above on the number of EMS calls for each tract. All independent variables in this study are ratio measures, with the exception of the EMS calls, which we transform into a binary measure, as discussed below.

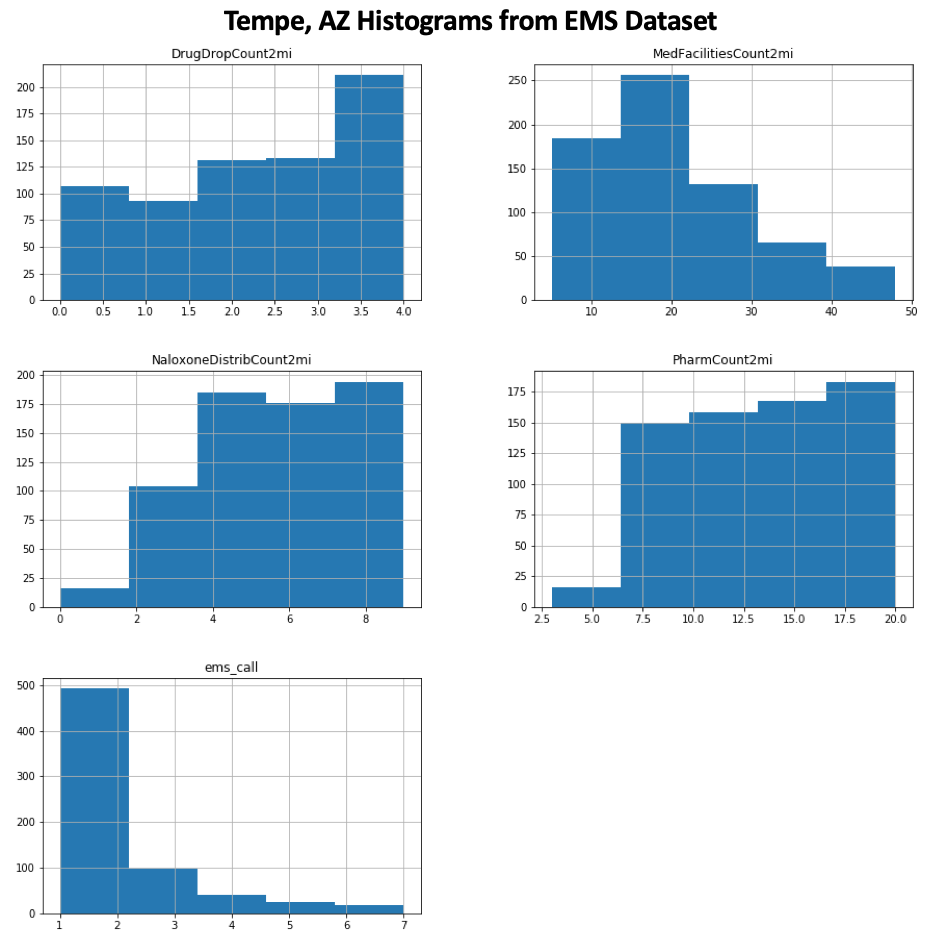


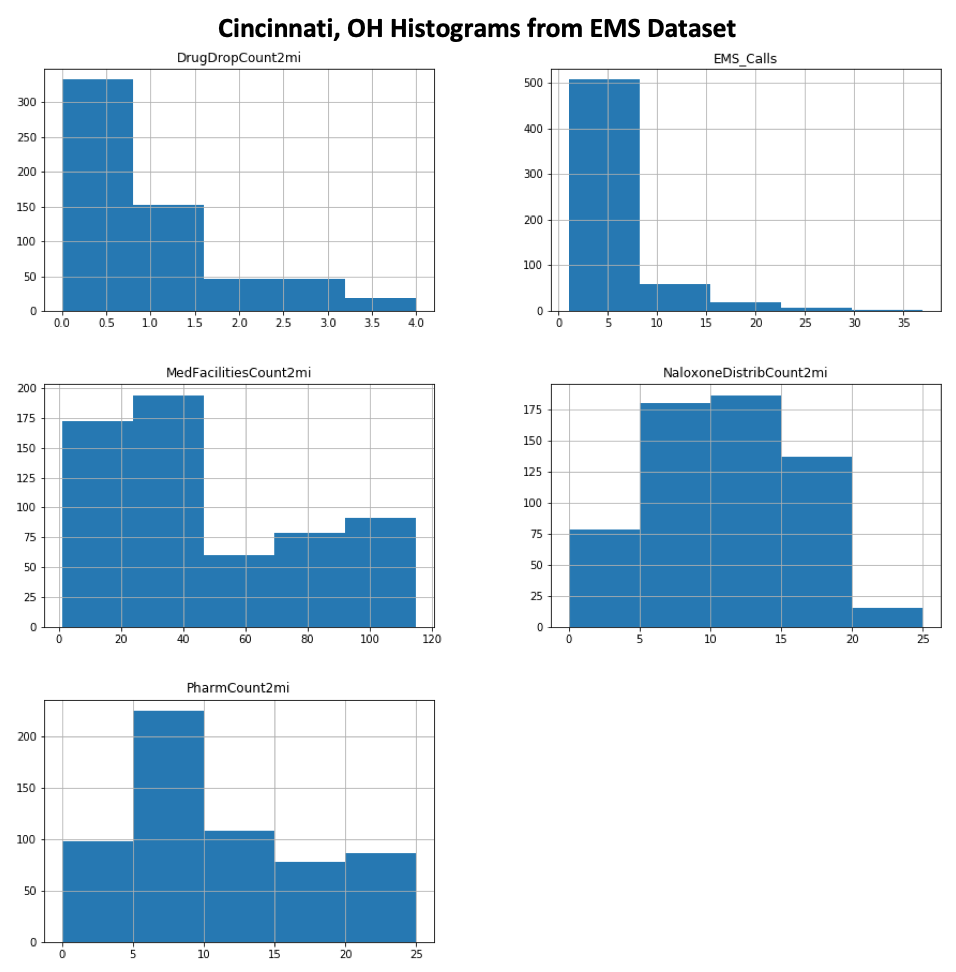
The count of EMS calls is power law distributed per census tract per year per month. This makes sense intuitively, since we can expect only a few calls a month, with long tails of high a number of calls happening in certain extreme cases (There’s more Cincinnati data does this hold true there too). As we will discuss below, we transform this variable into a binary classifier for a typical number of calls in an area at a point in time and an atypical high number of calls (i.e. above average per the dataset). As such, we use the following logistic regression model:



DESCRIPTIVE STATISTICS

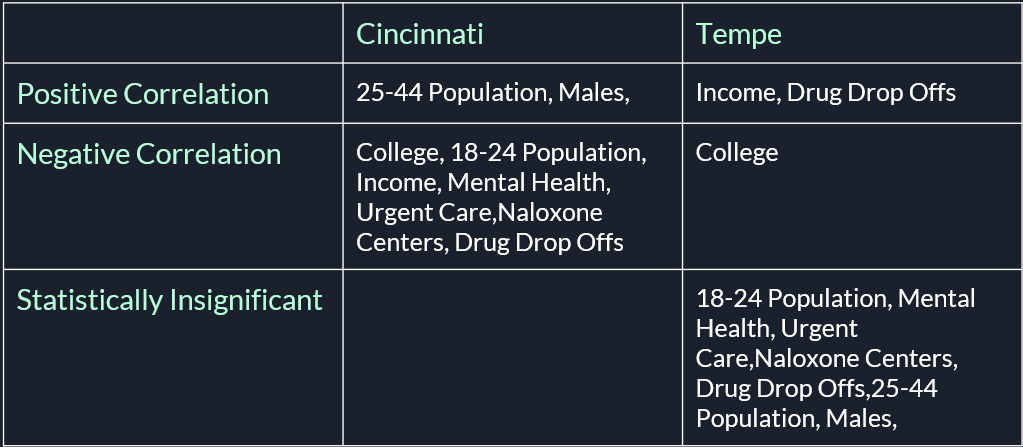






|  |  |
| --- | --- |
|  |  |

Generate and clean pilot data reflecting each of the variables identified above. Display descriptive statistics



**Conduct a power analysis to determine how much data is necessary to detect the hypothesized relationship. You may use off the-shelf software. Your power analysis should be based on your pilot data**