Question Write-Up (Plan 2)

EMSE 6577

Drew Gobbi, Olivia Wang, Olatunji Akinbule, Maddie Warndorf

1. **Specify the Problem**

This project will study the concentration of opiates in urban watersheds to estimate opioid usage patterns in cities. We intend to measure the relationship between estimated opiate use and access to care within certain communities.

The individuals who will care about this project are city policy makers, city health officials, and emergency medical services. This project can help better allocate resources to areas that may not have been identified through common opioid overdose measure. We believe that a geographic analysis and mapping of this public health problem provides an easily interpretable framework for non-technical audiences to better identify opportunities to improve treatment and identify risk factors.

1. **Three Generated Versions of the Question**
   1. Inductive

What relationships are present between opioid usage, access to treatment, and other demographic or geographic features? Can the presence of these relationships be used to form a testable hypothesis that can improve public health outcomes?

* 1. Deductive

Our theory from earlier literature is that communities with features X,Y,Z have higher levels of opioid usage. Can we confirm this hypothesis in our dataset?

* 1. Abductive

Given the opioid usage in a specific community, what is the most important factor which contributes to usage?

* 1. Best Version for Overall Goal

We believe that an inductive framework is the most appropriate for this question. We do not have any prior theory on what is driving opioid usage within a city, and will use our observations of patterns in the data to form hypotheses about the relationships between different features and health outcomes. This can help reduce confirmation bias from a deductive test of a prior theory.

1. **Different Types of the Question**
   1. Descriptive

Which communities are affected the most by the opioid epidemic, where are they, and are they able to access treatment?

* 1. Relational

Which communities are affected the most by the opioid epidemic, where are they, are they able to access treatment, and what might they have in common with one another?

* 1. Causal

Does the set of features X,Y,Z cause higher opioids usage?

* 1. Best Suited Type of Question for Overall Goal

We think that a relational question is best suited to our overall goal because we don’t have enough information about the data or prior literature to form a causal hypothesis, nor can we design control groups or experiments to test the causal relationships ourselves. What we can do is to figure out the communities that are most affected and the traits that they have in common, which could be the basis for future research.

1. **Unit of Analysis**

Our unit of analysis will be at the community level (i.e. watershed as one community level and/or census tract within a given watershed as another community level). The data we collect is representative of a given watershed and the inhabitants of that watershed, so we believe that this is the appropriate level of generalization for this analysis. Additionally, an analysis at the community level gives us an anonymized view of usage and health outcomes on average. While observations at the individual level may be biased by certain features (e.g. race or income with different usage patterns or access to treatment), we believe we can adequately control for these by including the aggregate composition of these features within the community.

1. **Cross-Sectional or Longitudinal**

Our question has both cross sectional and longitudinal components in understanding and tracking the concentration of opiates across sewage drainage collection areas.

The cross sectional component of this study stems from the analysis of multiple features at a given point in time including the income level, proximity to health care facilities and other available census data that can help establish potential relationships between opiate consumption and socio-economic characteristics of the population.

The longitudinal component of this study occurs because we are studying the repeated cross sections over time.

1. **Different Scope Versions of the Question**
   1. Too Broad

* Can the presence of opiates in wastewater explain opioid related deaths?
* How are demographic factors related to opiate concentration?
  1. Too Specific
* Can the detected presence of fentanyl in wastewater from the City of Tempe be linked to fentanyl related deaths within the past year?
* Is there an increasing trend across different collection areas in Tempe?
  1. Just Right and Justification
* How does the detected presence of opiates in wastewater contribute to the explanation of opioid related deaths within a specific city. This is the best version of the question because it cannot be answered in a simple “yes” or “no” way. It also gives more explanation on what is being discussed as well as what data should be collected.
* How does income level correlate with the opiate concentration levels from waste water within a certain collection area? This question has specific data points for which information can be collected to determine the relationship between income and narcotics usage.

1. **Feasibility**
   1. How Long Will it Take?

This analysis can be completed after a couple of months of research. The bulk of the work required will involve merging and cleaning the wastewater opioid measures, demographic features, and geographic features so that they can be integrated with one another to analyze potential correlations. Once this procedure is well established, we believe that it can be scripted and automated for real-time updates of our relational map.

* 1. Is it Ethical?

The analysis is performed on community-aggregated and anonymized data points to understand drug use patterns. This allows for an anonymous analysis that can identify opportunities to improve treatment risk and factors. Consequently, we do not anticipate any ethical concerns with the study.

* 1. Who is on the Critical Path?

Public health professionals and local and federal government agencies are on the critical path to completing this analysis for two distinct reasons. First, health professionals and government agencies will control the accessibility of some features necessary for this analysis. Second, as consumers of this information, we need to demonstrate the value proposition of showing relationships between drug use, treatment, and community level features in determining policy interventions that can effectively improve public health outcomes.

* 1. How Much Will it Cost?

This depends on whether or not we collect the data ourselves. The Tempe data is freely available from the project’s dashboard and much of the demographic and geographic data can be acquired from the city government’s open data portal and the US Census Bureau API. If we were to collect data ourselves, we would have to purchase and/or design collection units to sample around a given city. White papers from other pilots in Cary, NC and Mathematica Policy Institute suggest that this would cost between $20,000 and $100,000. Accordingly, we would require grant backing (as in the case of the Cary study) or a partnership with a local government agency to complete this type of analysis.

* 1. Is it Appropriately Scoped?

The question has a necessarily broad scope on which relationships are present in the dataset because we do not rely on any prior theory. Restricting to a given city helps to constrain this analysis to a set of observation points that are bound to an at least somewhat similar environment and associated set of features. Data availability and the detectability of certain relationships will also help control the scope of this project.

1. **“Can’t Lose” Formulation of Question**

The point of our analysis will be to detect the presence of certain relationships between opioid use, treatment levels, and certain community level features. Once we establish the presence of these relationships, we believe that we will be able to form causal questions in future analyses that are beneficial regardless of whether or not the answer is yes or no.

For example, if it is shown that there is a negative relationship between treatment and opioid usage, we could test the effectiveness of a public health intervention on reducing drug use in certain areas. If this is effective, then we can establish that the same type of intervention should be leveraged in other areas to improve health outcomes. If the intervention is ineffective, we can test for other confounding relationships amongst features in the community that may be contributing to the opioid problem.

Similarly, if it is shown that certain communities with different features have different usage patterns despite exhibiting the same access to treatment, we could test why existing public health interventions are ineffective in those communities and what new interventions may increase health outcomes over time.