For office use only	Team Control Number	For office use only
T1 ________________	**1901679**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2019

MCM/ICM

Summary Sheet

After exploring the data provided to our team describing drug use and socioeconomic factors between 2010 and 2016 inclusive, we sorted the 69 types of opioid substances into four drug categories based on their synthesis and availability. Plotting use rates of each category of drugs over time revealed that use of mild painkillers and natural alkaloids have stayed relatively constant over time, semi-synthetic drugs have declined slightly, and synthetic drugs such as fentanyl and heroin have increased dramatically. These findings align with reports from the CDC. We also selected 54 out of 149 socioeconomic variables based on their variance inflation factor score (a common measure of multicollinearity) as well as their relevance based on the public health literature. To model the spread of the opioid crisis across Kentucky, Ohio, Pennsylvania, West Virginia, and Virginia, we took a dual-pronged approach, developing two completely different models and then comparing them at the end.

Our first model is founded on common modeling approaches in epidemiology: SIR/SIS models and stochastic simulation. We designed an algorithm from scratch which simulates a random walk between six discrete classes, each of which represent a different stage of the opioid crisis using thresholds for opioid abuse prevalence and rate of change. We penalize transitions between certain classes differentially based on realistic expectations. Optimization of parameters and coefficients for the model was guided by an error function which we also designed from scratch, and was inspired by the global spatial autocorrelation statistic Moran's I. Testing our model via both error calculation and visual mapping illustrated high accuracy over many hundreds of trials. However, this model did not provide much insight into the influence of socioeconomic factors on opioid abuse rates, because incorporating socioeconomic factors did not significantly change the model results.

Our second model made up for this deficiency in socioeconomic factor analysis. By running a collection of spatial regression models on our final collection of socioeconomic predictors (including total drug use rate), we explored characterising the spatial patterning of the opioid crisis as the result of a spillover effect, as the result of spatially-correlated risk factors, and as a combination of the two, using spatial lag, spatial error, and spatial Durbin models respectively. While all models confirmed significant spatial signals, the spatial Durbin model always performed the best. We also calculated the direct, indirect, and total impacts of each predictor variable on opioid abuse rate. Far and away the most important variable in all models was the total drug use rate in each county. The average result (across all seven years) was that, all else equal, a unit increase in total illicit drug use rate would raise the opioid abuse rate by 52%. This is quite realistic given a CDC statistic that in 2014, 61% of drug overdose deaths involved some type of opioid. By contrast, an ordinary linear regression reported only a 37% increase in opioid abuse rate per unit increase in total drug use rate. Statistical measures such as the Akaike Information Criterion and Likelihood Ratio Test verify the superiority of our spatial models.

To predict possible locations of origin of the opioid epidemic in each of the five states, we ran a Monte Carlo simulation of our random walk model from 2000-2010. We map out these counties and discuss their arrangement in the context of our other findings. The random walk finds that the opioid crises most likely starts in Montegomery, KY which aligns with our research that opioid abuse is more prevalent in rural communities than urban [10].

To forecast spread of the opioid crisis from 2017-2020, we used both our random walk and spatial regression models. The two models display surprisingly minimal deviance from one another, especially in 2019 and 2020. The random walk predicts that the number of counties above the illicit opioid use threshold will go down naturally within the next 7 years which also aligns with the idea that the opioid epidemic follows the spillover effect seen in epidemiology.

Due to the assumption that the SES indicators will change linearly, the second model's error will significantly increase after about 4-5 years. The random walk on the other hand, operates on a healthy tension between wanting to cluster together and randomly assigning classes. Near the initial date, it clusters more but the randomness starts to compound rather quickly. For this reason, the random walk has lowest errors near the 4-7 year mark. This means that the best strategy to predict the near future would be the spatial regression and the random walk for the 4-7 year range. Predicting anything beyond this point will have high error. Afterwards, we provide the the suggestion to the government that reducing general drug usage will help decrease illicit opioid usage.