Modeling Opioid Abuse Indicators and Interventions in Appalachia
AI for Social Good Workshop

Savannah Thais and Chase Shimmin

- **Topics**: Which of the following categories does your problem belong to?
  - Health
  - Social welfare and justice
- **Problem**: What problem do you want to investigate and why? If known, what are the root causes of the given problem? What are some existing solutions? (max 200 words)

  We want to investigate the issue of opioid abuse in Appalachia. Although the opioid crisis is a pressing national issue, Appalachia has higher rates of opioid use than any other area of the US [1]. Coupled with limited medical care access (there are 65% fewer providers per 100,000 people than elsewhere in the US [1]), under-serving education programs [2], and a rapidly growing depression rate [1], this has led to an overdose death rate up to 5 times higher than the national average [3]. A complex network of policy and infrastructure issues unique to this region (economic distress due to the decline of the coal industry, widely-spread and difficult to access rural communities, and disconnected, apathetic local politicians [4] [5] [6]) necessitate a location specific intervention. Existing solutions such as Prescription Drug Monitoring Programs and Naloxene distribution have proven inadequate as the rate of opioid related death in the region continues to grow [3].

- **Proposal**: Describe your proposed solution. How does it address the shortcomings of current approaches? (max 200 words)

  We plan to build an integrated model which accounts for economic, education, public health (mental and physical), infrastructure, and political factors as they relate to both perpetuating and mitigating the opioid crisis in Appalachia. This work will proceed in two stages: first, we will build a predictive model to estimate markers of opioid abuse severity such as fatal and non-fatal overdose rates, opioid prescription rates, and drug related arrest rates. We will analyze this model using widely accepted deep-learning interpretation techniques [8] to gain insight into what factors within the described cause areas are most correlated with drug abuse, as well as whether these factors vary by community type within the region (extremely rural, small town, or city). We also plan to implement causal inference techniques, as described in [7], to better understand possible limits on causal reasoning regarding our findings and potential shortcomings in our initial dataset (in terms of a lack of support/overlap for certain indicators). After this initial analysis, we will explore existing small-scale interventions currently implemented in certain Appalachian areas. The goal of this second analysis is to better understand a generalizable, multifaceted, proactive, community specific intervention plans. We will begin this study by employing unsupervised clustering methods to better understand existing success indicators and which community indicators (indicated by the first phase of the project) are not being addressed by current interventions. Collectively, these two studies will allow us to gain insight into both successful intervention tactics and region specific risk factors, and together these insights will inform more effective opioid crisis response.

- **Impact**: What is the expected social impact in the short, medium, and long-term of the solution to the problem? (max 150 words)

  The impact of the first project phase would be a more robust understanding of the unique and interacting factors contributing to the Appalachian opioid epidemic. This information would be useful to community organizers and health professionals looking to design more localized and community tailored intervention programs. Long term, through modeling existing programs and community structures, we want to provide additional guidance to policy makers on which types of programs are likely to be most successful in their individual community. We hope this information will influence resource distribution, funding, infrastructure development, and law making decisions.

- **Evaluation**: How would you quantify success? Are there smaller-scale environments in which you can test your proposal? How might a larger-scale deployment fail to reflect the initial experiments? (max 150 words)

  The success of our predictive model could be measured using traditional ML metrics like test-set accuracy and MSE. Our analysis of indicating factors could be compared to existing community analyses (eg [10]), although these only exist for certain areas and do not account for all the factors we are considering. This method of testing is also limited to the individual communities in which we make these comparisons, which may allow our model to ultimately perform better for certain communities. After constructing our models, we hope to work with a local Appalachian organization to use the insights gained from our analysis in their programmatic design. Success in such a program (measured in terms of improved local outcomes) would lend credibility to our model and encourage other organizers and policy makers to utilize our findings. This is subject to the same risks discussed above: a community specific test is not necessarily a generalizable proof. This could be mitigated by partnering with multiple organizations in different community types. It is important to consider the distribution of benefit and harm as evaluation metrics as well, to ensure that redesigned intervention programs are benefiting all sub-populations.

- **Risks**: Could your solution lead to any unintended harmful consequences or risks? Describe them. How could the resulting system be abused? Are there vulnerable populations that might be put at risk? What checks could you introduce to prevent these potential bad actors? (max 150 words)

  We see the biggest risk of our system to be incomplete or inaccurate intervention modeling. If policy makers and organizers utilize our recommendations, this could result in inefficient funding allocation and program design. We believe all of the interventions included in our model would be positively impactful or neutral, and that our location specific, multifaceted approach to modeling indicators is novel and addresses region specific factors that are not considered in other work. Regardless, the possibility remains that we are not accounting for all contributing components of the Appalachian opioid epidemic. This crisis is, in many ways, a social problem which consequently requires a social solution, and social factors are not always easy to model numerically. Nevertheless, we believe that by incorporating information on community structures, legal influences, location specific infrastructure, and economic development, we will progress beyond previous work in this area.

- **Data**: Describe the dataset(s) available for your project (i.e. amount of data, measurements granularity, data collection frequency, way of accessing the data). Who is responsible for data collection? Are there privacy concerns, and what is the license? (N.B.: In the absence of privacy concerns, we encourage data that can be shared publicly). How have these datasets been used previously? (max 200 words)

  Much of our initial data comes from the County Health Rankings & Roadmaps Program (CHRRP) [11]. CHRRP is a collaboration between UW Population Health Institute and the Robert Wood Johnson Foundation to provide researchers and community leaders with local data around multiple health-influencing factors. This data is updated annually, is public, and contains data on all 3,007 US counties. Relevant information includes income distribution, education levels, employment statistics, accessibility of healthcare resources, community safety metrics, community structure descriptors, and drug related deaths. We intend to verify the drug related death information from CHRRP by using the CDC Detailed

Mortality data [12]. We also plan to utilize county level opioid related arrest records as an additional abuse behavior indicator [13]. In order to account for political and infrastructure related factors, we plan to analyze records of pharmaceutical company lobbying (see eg [16]), infrastructure development spending [14], and local voting records on proposed laws like Kentucky Senate Bill 6 [15]. The data regarding opioid abuse interventions in Appalachia is taken from a variety of sources including grants from the US Department of Agriculture and Rural Development [17], resources from the Office of National Drug Control Policy [18], and intervention strategy publications for the National Academy for State Health Policy [19].

- **Labels**: Would your data require any additional annotation before it could be incorporated into your solution? If so, how do you plan on obtaining these labels? Are there different approaches to annotation, and how do they compare in terms of level of detail and ease of preparation? (max 150 words)

  Much of our data is already labeled by county and all the data from CHRRP is constructed using metrics that allow for easy normalization and comparison between communities. However, some of the more policy related data like voting records and lobbying would need to be constructed by hand. We would need to assess the similarity of policy proposals and lobbying efforts across different counties and states to ensure that our final dataset is self-consistent. Ideally, we would be able to construct this portion of our dataset with a legal expert. The intervention information for the second stage of our project is more heterogeneous. We would likely need to design quantifiable success metrics and program descriptors to make the various interventions directly comparable. We will depend heavily on our public policy focused team members for this design process, and hope to also be able to work with public health experts.

- **Social System**: Describe your team's skills and backgrounds. What are other resources (i.e. stakeholders, scientists, and funders) would you like to add to your team? (max 150 words)

  Our team consists of four members with a variety of project related backgrounds. One member is a ML researcher who has worked primarily on generative models. Another member is a post-doc who works on scientific research applications of ML including spatio-temporal modeling and particle identification. A third member is a former resident of Appalachia who has worked with a variety of community outreach organizations; they are currently finishing a physics PhD and will start a ML focused post-doc in a few months. Our final member is completing their Masters Degree in Public Policy and researches community focused intervention. We feel our ML experience and expertise is quite robust, but would greatly appreciate working with someone with experience in causal modeling. Additionally, we hope to work with a public health researcher and a legal expert who can help with some of our data aggregation needs.

- **Technical System**: If applicable, please share any technical elements of your proposed solution that have already been explored. What would your baseline system look like, how well do you imagine it will work, and what extensions have you imagined? (max 150 words)

  A baseline system would combine an opioid abuse indicator predictive model with an unsupervised learning analysis of existing community based intervention programs. The predictive model, likely a deep neural network, would be trained on county-level socioeconomic and health data to calculate prevalence levels of opioid abuse behaviors. If we can access the necessary data, we imagine building an individualized predictive model also. We would then employ a variety of interpretation techniques like activation maximization and layer-wise relevance propagation [8] to better understand risk indicators and their relationship to community type. The investigation of existing intervention programs would begin with clustering and component analysis to better understand success indicators and critical elements of impactful programs. By understanding commonalities of existing interventions and how completely they address the indicators we understand to be important from our predictive model, we can better provide guidance on developing new programs. We also hope to explore model-based Reinforcement Learning to aid in the design of new interventions [9], but would need to add additional expertise to our team.

References

[1] PDA Inc, The Cecil G Sheps Center for Health Services Research, and Appalachian Regional Commission, *Health Disparities in Appalachia*, August 2017

[2] T. Shaw, A. DeYoun, and E. Redemacher, *Educational Attainment in Appalachia*, Journal of Appalachian Studies. Volume 10 Number 3.

[3] *Overdose Mapping Tool*, `https://overdosemappingtool.norc.org/`, Accessed March 2018

[4] C. Hurst, *Inequality in Appalachia*, 6th Edition, Pearson Education, pp 62-68

[5] D. Narcisco, *Appalachian Spring: Will it Ever Come?*, Christian Science Monitor, 08827729, Vol. 90, Issue 215

[6] T. Burris, *Appalachian consequences of the War on Poverty*, American Psychological Association

[7] MIT-IBM Watson AI Lab, *Fighting the Opioid Epidemic with Interpretable Causal Estimation of Individual Treatment Effect*, https://medium.com/@MITIBMLab

[8] G. Montovan, W. Samek, K. Muller, *Methods for interpreting and understanding deep neural networks*, Digital Signal Processing, Volume 73 (2018), pp 1-15

[9] K. Doya, K. Samejima, K. Katagiri and M. Kawato, *Multiple Model-Based Reinforcement Learning*, Neural Computation (2002), Volume 14 Issue 6

[10] *Appalachian Opiate Epidemic: Visualizing Overdoses*, https://heroin.net/think/appalachian-opiate-epidemic/

[11] University of Wisconsin Population Health Institue and Robert Wood Johnson Foundation, *County Health Rankings & Roadmaps*, http://www.countyhealthrankings.org/

[12] Center for Disease Control and Prevention: CDC Wonder, https://wonder.cdc.gov/

[13] Opendata PA, *Opioid Arrests by County*, https://data.pa.gov/Opioid-Related/Opioid-Arrests-by-County-PSP-/tpie-wrnx J. Scura, *Pharmaceutical Firms Continue to Inundate Towns with Opioids*, https://www.freedommag.org/blog/pharmaceutical-firms-continue-to-inundate-towns-with-opioids-e813f1

[14] US Census Bureau, *Public Infrastructure Spending for State and Local Governments*, https://www.governing.com/gov-data/state-local-government-construction-spending.html

[15] Kentucky Senate Bill 6 (2018) https://legiscan.com/KY/bill/SB6/2018

[16] J. Scura, *Pharmaceutical Firms Continue to Inundate Towns with Opioids*, https://www.freedommag.org/blog/pharmaceutical-firms-continue-to-inundate-towns-with-opioids-e813f1

[17] United States Department of Agriculture and Rural Development, *Program 101 Fact Sheets*, https://www.rd.usda.gov/publications/fact-sheets

[18] Office of National Drug Control Policy, *Federal Resources for Rural Communities to Help Address Substance Use Disorder and Opioid Misuse* (2018), https://www.rd.usda.gov/files/RuralResourceGuide.pdf

[19] C. Corso and C. Townley, *Intervention, Treatment, and Prevention Strategies to Address Opioid Use Disorders in Rural Areas: A Primer On Opportunities For Medicaid-Safety Net Collaboration*, (2016), https://nashp.org/wp-content/uploads/2016/09/Rural-Opioid-Primer.pdf

[20] C. Louizos et al, *Causal Effect Inference with Deep Latent-Variable Models*, Conference Paper at NIPS 2017