Modeling HIV Incidence Levels

Eric C. Logue
Galvanize Denver

elogue01@gmail.com
linkedin.com/in/elogue01
Github.com/elogue01

Background

- 971,524 HIV+ individuals in the US in 2015[1]
- 37,600 new HIV diagnoses in 2014[1].
- Relatively high rates in communities of color, among men who have sex with men, and other vulnerable populations[2].
- Lower socioeconomic status is a significant risk factor[2].
- US opioid crisis has further complicated the efforts to combat HIV

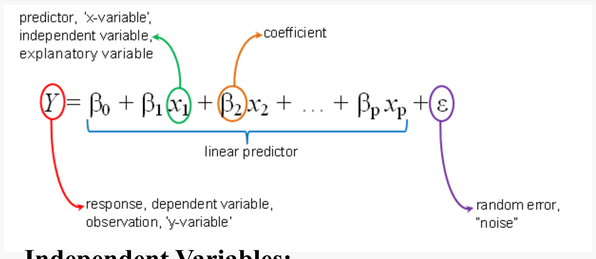
Objectives

1. To accurately model the HIV incidence (new infections per 100,000) in US counties using a linear regression model
2. Identify factors that are significant drivers of HIV incidence levels and lunderstand how these drivers differ between regions

Data

- Data was sourced from the amfAR opioid database, AIDSVu, 5-year American Community Survey, the Emory CAMP.
- This analysis focused on the year 2015.
- Data from the Scott County, IN outbreak was excluded from the dataset.
- Data suppression narrowed the dataset to 746 counties

Multiple Linear Regression



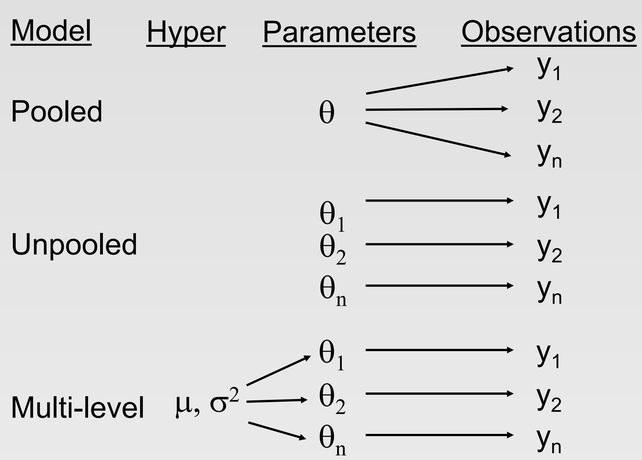
Independent Variables:

1) HIV prevalence (HIV+ individuals per 100,000)
2) % of population African American
3) % of population uninsured
4) log transformed household income

Dependent Variable:

HIV incidence (new HIV diagnoses per 100,000)

Bayesian Models



Evaluation and Results

Model	WAIC Score	LOO Score	RMSE
Pooled Model	5143.71	5149.12	7.48
Unpooled Model	5295.42	5304.53	7.15
Multilevel Model	5066.49	5092.10	6.30

Table 1: Evaluation for pooled, unpooled, and multi-level models

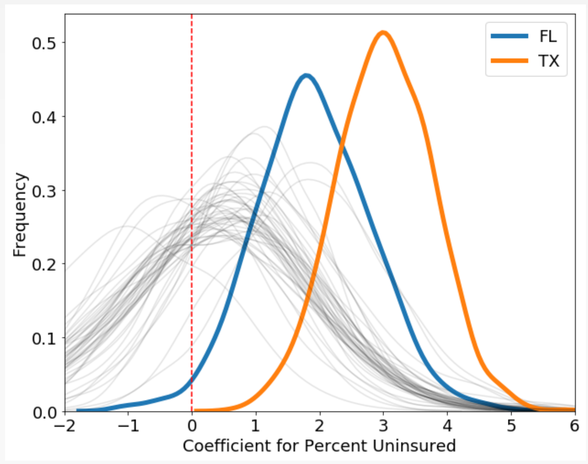
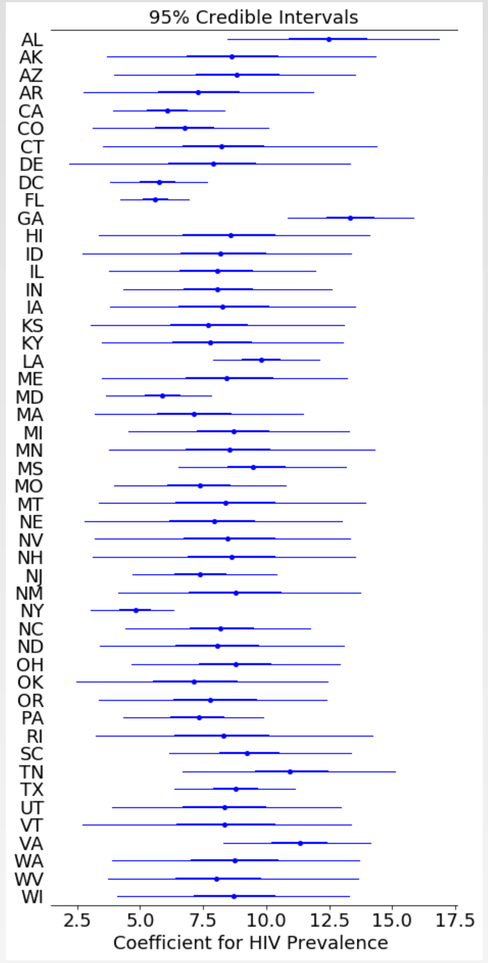


Figure 1: Florida and Texas have a positive coefficient for the percent uninsured variable. Probability distributions for the percent uninsured coefficient are displayed . Distributions with 95% CI above zero are displayed as bolded colored lines.

Evaluation and Results Continued

A



B

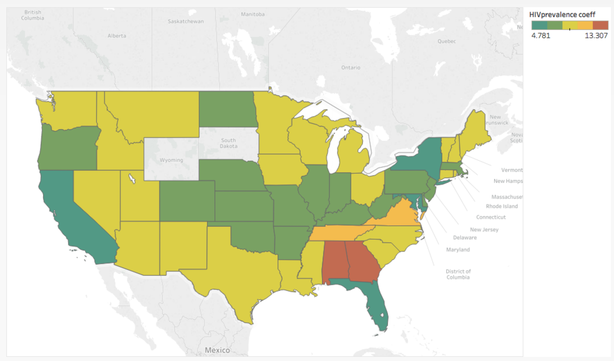


Figure 2: The effect of the HIV prevalence level on HIV incidence varies by states. A) Probability distributions for the HIV prevalence coefficient are displayed with interquartile range (bold line) and 95% CI (thin line) for each state. **B)** Mean HIV prevalence coefficient values for each state are displayed on a map of the US as a color spectrum represents values from low to high (blue to red).

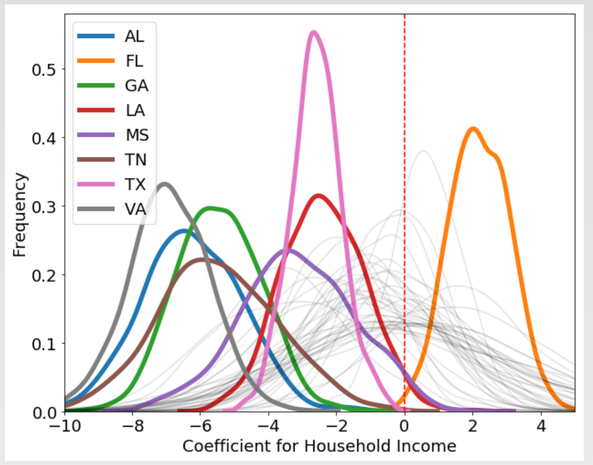


Figure 3: Seven southern states show negative coefficients for household income levels. Probability distributions for the $\log_{10}$ transformed household income coefficient are displayed for each state. Distributions having significant positive or negative distributions are displayed as bolded colored lines.

Discussion

- Reductions in the rates of the uninsured in FL and TX could reduce their HIV incidence levels (i.e. Medicaid expansion).
- AL and GA appear relatively ineffective at combating the spread of HIV infection.
- Economic shocks may be harmful to HIV incidence levels in southern states
- Potential for model refinement
  - Add data from additional years
  - Additional features need to be explored in the multi-level model

References

1. *HIV in the United States: At A Glance https://www.cdc.gov/hiv/statistics/overview/ataglance.html*

2. *CDC fact sheet:Todays HIV /AIDS epidemic https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/todaysepidemic-508.pdf*