**INTRODUCTION**

Distributions of health outcomes take many different forms. The shape of the distribution depends on the particular health outcome and the specific population amongst other factors. In thinking about how to improve the health of populations it is particularly important to understand what influences health and specifically whose health it influences. Health outcomes also have an important geographical component; populations in certain areas often share certain risk factors and disease distributions. As urbanization continues and expands in the coming decades, understanding how place of residence impacts health will become increasingly important.

Here I look at several specific health outcomes and their relationship with socioeconomic status (SES) and density, which represents a rough measure of urbanicity.

**METHODOLOGY AND DATA**

In order to quantitatively analyze the impact of urbanicity on health, I selected several specific health outcomes (CHD & breast cancer) and looked at their distribution over a range of densities using county and city data. I then used multiple regression to test the relationship between the specific health outcomes, density and a range of measures of socioeconomic status.

While health is a composite variable, an overall measure of health like average life expectancy encompasses a wide range of factors. Choosing specific health outcomes allowed me to parse out the risk factors associated with each outcome and further investigate the nonlinear relationship seen here.

**RESULTS AND ANALYSES**

**CORONARY HEART DISEASE**

A scatter plot of log-density and coronary heart disease mortality divided by city (purple) and county (grey) data with a blue locally weighted smoother of the county data and a yellow line of best fit for the city data.

The regression model indicates a highly significant nonlinear poly-nomial that describes the relationship between CHD mortality, density and a range of socioeconomic factors ($F(11, 3087) = 188.98$, $RMSE = 73.12$; $p < 0.0001$). Compared to a similar model without density, log density explains an additional 7% of the variance ($R^2 = 0.40$ compared to an $R^2$ of 0.33).

**BREAST CANCER**

A scatter plot of the relationship of log density with breast cancer mortality divided by city (purple) and county (grey) data with a blue locally weighted smoother for county data and a line of best fit in yellow for city data.

The regression model indicates a significant nonlinear polynomial that describes the relationship between breast cancer mortality and log-density ($R^2 = 0.31$ compared to an $R^2$ of 0.18) relative to a model without log density as a predictor variable.

**DISCUSSION**

- The relationships between coronary heart disease (CHD) mortality and breast cancer mortality and density is distinctly nonlinear.
- In the mid-low ranges of density there is a precipitous drop in both breast cancer and CHD mortality followed by a subsequent increase/ leveling off.
- This needs further attention and explanation; what other variables are influencing this relationship given that many aspects of SES are controlled for?
- City data are situated at the dense end of the county distribution and resemble the distributions of both health outcomes at the county level.
- Even though county and city data does not describe the same populations, these cities seem to be experiencing a similar overall phenomenon; after a density of 1,000 people/mile², density has little to no effect on CHD and breast cancer mortality.

**THANKS & ACKNOWLEDGEMENT**

First and foremost, thanks to Nick and Pad for all their help and guidance relative to the project. I would also like to thank MACED for their资助 of the Big City Health Inventory dataset and the Foundation Trust for their support of this research.