Analysis of Chronic Disease risk estimates for different cities in the state of Virginia

Venkata Dhanush Kikkisetti

2023-12-08

Introduction and Backgound:

Chronic deceases effecting many people lifes and introduced many health problems.There are different types of chronic diseases like asthma, diabetes, Arthritis, kidney disease, pulmonary disease etc. It is very difficult to find how it is effected but it can prevented when it is traced earlier. To analyze the risk factor, I gathered 50 states of dataset with different cities and different chronic measures estimates which consist of nearly 800k rows. As it is very complex to analyze the complete dataset, I want to analyze the risk factors for the state of Virginia. The main goal of this project to analyze the chronic disease risk factor for varying cities in the state of Virginia.

Dataset and data transformation:

Source of data : "https://cdcarcgis.maps.arcgis.com/home/item.html?id=ea8b721cf9034814bce067ddefd21ecc" The dataset provides 500 Cities Project 2016 data release based on 2014, 2013 model-based small area estimates(SAE) for 27 measures of chronic disease related to unhealthy behaviors (5), health outcomes (13), and use of preventive services (9). Data were provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. It represents a first-of-its kind effort to release information on a large scale for cities and for small areas within those cities. It includes estimates for the 500 largest US cities and approximately 28,000 census tracts within these cities. These estimates can be used to identify emerging health problems and to inform development and implementation of effective, targeted public health prevention activities.

As said, I only considered health outcomes(13) and virgina state for this analysis which resulted in 6000 rows, which is good for my small scale analysis. Here each observation is the estimates of small area(may be county) with in that city with its population count for that small area. Here are the main variables we have for our analysis.

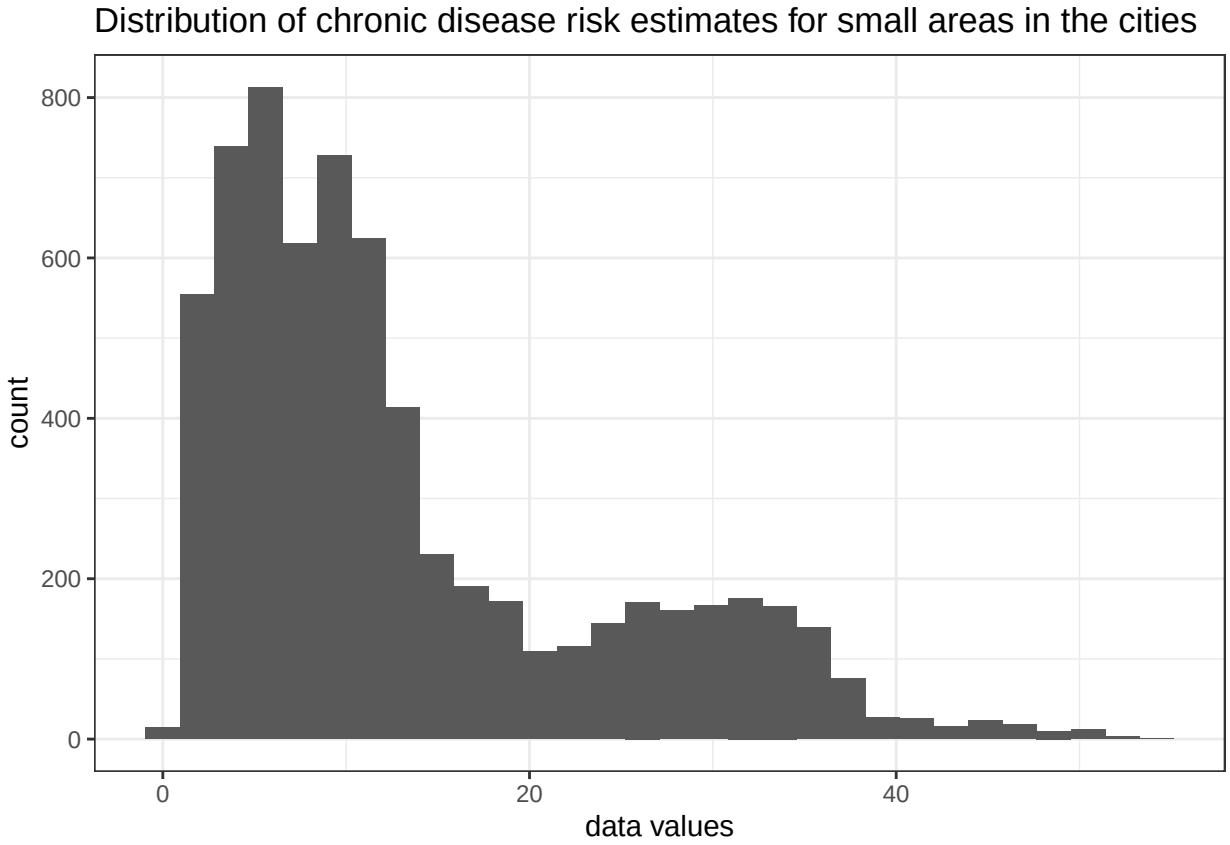
Table 1: Data variables and it's description

Variables	Data Description(total number of level for each categorical variable)
City	11 Different cities in Virginia (Alexandria, Chesapeake, Hampton, Lynch-burg, Richmond, Virginia Beach, Roanoke, Suffolk, Norfolk, Portsmouth, Newport)
Health outcomes	13 different health outcome for CH disease(asthma, diabetes, Arthritis, kidney disease, pulmonary disease ETC)
SAE	Risk estimates for small area within those cities
Population Count	Total number of people in that small area
HighCI	97.5 percentile of risk estimate for SAE

Variables	Data Description(total number of level for each categorical variable)
LowCI	2.5 percentile of risk estimate for SAE

There are 28 NA values in the SAE, HighCI,LowCI variables in the data set. There is a possibility that we can remove this Null data points as 0.5% in complete data. However, I used multiple imputation to check the relation ship between Null value columns to other column to check any relationship and predict the null values with multiple imputation.

Lets look at the distribution of small area estimates



Statistical Approach:

Now the dataset is clean and it is transformed into its appropriate form. We consider only health outcomes observation in the state Virginia as we want analyze how small area estimates differ for each city in Virginia with varying health outcomes, i.e We want to analyze different risk estimates for different cities in the state of Virginia and analyze the risk factor for varying health outcomes with in the cities and across all the cities.

Using Bayesian inference, we create a varying intercept-varying slope and non-nested multi-level models to compare the small area estimates for risk factor. Here varying intercept and varying slope(VIVC) is for prediction of SAE that vary with cities and vary health outcomes within those cities. We use this to see how chronic disease estimates vary for each cities and impact of different chronic diseases in cities have in prediction. Here is the model diagnosis after fitting VIVC model.

Table 2: VIVS model diagnostics in Bugs model

Variable	Estimate	SD	RHat	neff
Alexandria	10.24	0.61	1.01	150
Chesapeake	12.96	0.54	1.01	180
Hampton	13.6	0.59	1.005	400
.....				
Portsmouth	14.69	0.61	1.08	280
sd_a	0.14	0.47	1.008	640
sigma_a	28.73	27.12	1.08	33

$$y_i \sim \mathcal{N}(\alpha_{j[i]}^{city} + \beta_{j[i]}^{city} \times ChronicMeasures_i, \sigma_y^2)$$

$$\alpha_j^{city} \sim \mathcal{N}(0, \sigma_\alpha^2)$$

$$\beta_j^{city} \sim \mathcal{N}(0, \sigma_\beta^2)$$

The bugs model is runned for 3 chains and 4000 iterations each. We observe that all the city estimate and super population,finite population standard deviation reached to its convergence level at Rhat values less than 1.1.

Non-nested multi level model is for prediction of SAE that vary with cities and vary though health outcomes. Here we will predict the SAE with varying cities and varying chronic disease across all the observations. I am interested in analysis how well the model can able to predict the estimates with both varying cities and varying chronic diseases instead of fitting a model that vary chronic disease with in the cities.

Table 3: Non-Nested model diagnostics in Bugs model

Variable	Estimate	SD	RHat	neff
Alexandria	9.40	0.70	1.07	39
Chesapeake	12.39	0.69	1.08	36
Hampton	13.18	0.69	1.08	34
.....				
Portsmouth	14.39	0.70	1.07	39
sd_a	1.50	0.05	1.001	3000
sigma_a	0.67	0.14	1.01	3000

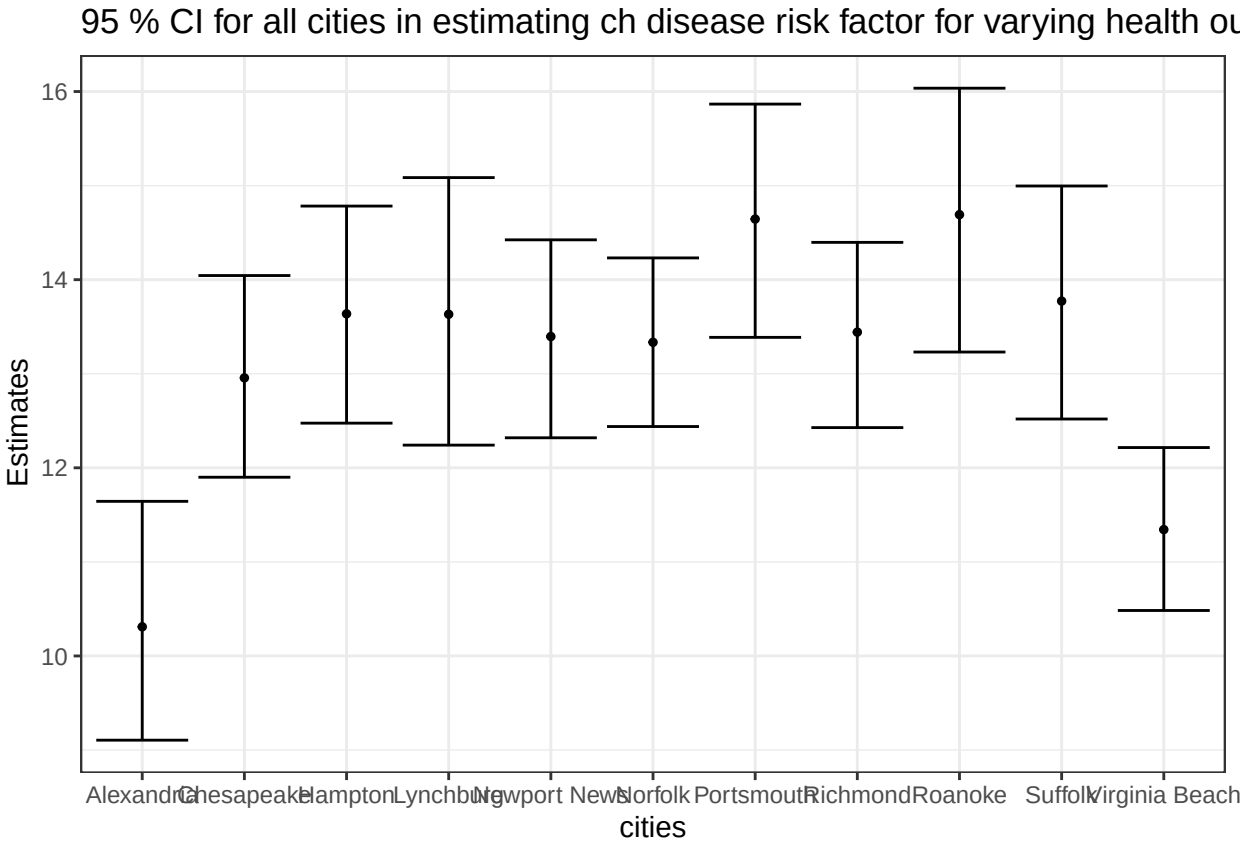
$$y_i \sim \mathcal{N}(\mu_0 + \alpha_{j[i]}^{city} + \beta_{k[i]}^{ChronicMeasure}, \sigma_y^2)$$

$$\alpha_j^{city} \sim \mathcal{N}(0, \sigma_{city}^2)$$

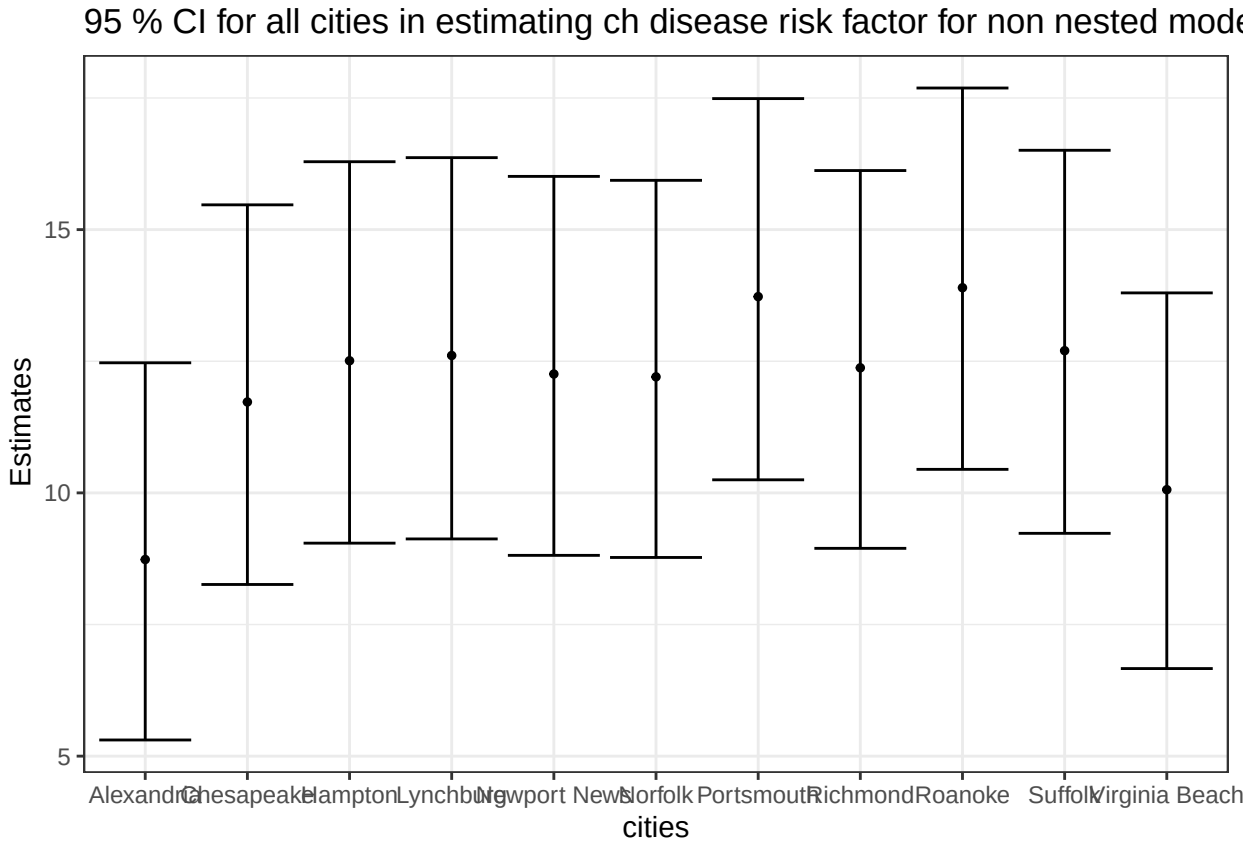
$$\beta_k^{ChronicMeasure} \sim \mathcal{N}(0, \sigma_{ChronicMeasure}^2)$$

Empirical Finding:

Here we will observe how well the multimodel predicts the accurate estimates with less variation in 95 confidence intervals. I fit 2 model here to predict the chronic disease risk estimates, one is varying intercept and varying slope(VIVS) and other is non-nested model. We fitted VIVS to see if their is any relationship in risk estimate with varying cities intercept and varying health outcome. All the parameters reached their convergence level with Rhat less than 1.1. However, the super population standard deviation,which represents the variation among the modeled probability distribution from which the group averages were drawn, for cities intercept is very high(28.73759992) and its standard deviation is also high indicating that it is highly uncertain to get the value of new city if it is not in the original set of cities. Here is the plot for SAE estimates to check city estimates for 95 %CI this model.



However, I tried fitting a non-nested as we have 2 categories that are not subset of other: 1660 Small area estimates are divided into 11 cities and 13 health outcomes. All the estimates in the bugs model are converged with Rhat value less than 1.1. More importantly, the super population standard deviation is very less compared to the VIVS model. So the estimates are less uncertain that the estimate drawn for a new city will be in the city group. The estimates standard deviation is a little more compared to the VIVS model which can be noted. Lets look at the SAE estimates to check city estimates CI for Non-nested model.



The city estimates are almost similar with both the models with sight higher standard deviation for non nested model. From the plot, I see that Alexandria has less chronic disease risk estimate(9.3) and it looks like Roanoke city has high chronic disease risk estimate (14.8). Here we are just analyzing small area risk factors that vary for different cities by fitting 2 types of model architecture with also considering the type of chronic disease health outcome in the model. This will just give the overview of the estimate with the Bayesian models, which can latter analyse the other factors that contribute to the risk estimate in the cities.

Future Other works :

We can also analyse just considering single health outcome like obesity, cancer etc across all the cities in the united states and see how it varies for different cities in united state.