Research Brief,

Short Paper

Vol. 2, No. 15

(2020, November 29)



The Illinois Institute for Rural Affairs (IIRA) works to improve the quality of life for rural residents by partnering with public and private agencies on local development and enhancement efforts.

Western Illinois University

# Segments of Metro and Nonmetro Population that are Prone to Severe Illness from Covid-19: Insights for Policy

ISSN 2687-8844

Adee Athiyaman[1]

## Introduction

As at November 28, 2020, US reported 13.4mil Covid-19 cases and a 2% case-fatality rate[2]. The virus induces severe illness[3] in persons with underlying medical conditions such as COPD, obesity, and type2 diabetes.[4] For policy purposes, it is essential that communities know about population segments that could suffer severe illness from Covid-19. While county health rankings provide information about obesity and smoking, information on other correlates of severe Covid-19 illness such as cancer, COPD, and kidney disease are not provided. Other issues with county health ranking data include outdated information - obesity and smoking data are for the 2016 and 2017 time periods, and inability for the user to perform multivariate analysis of data, for example, cross-classify persons who are both obese and diabetic with their demographics such as gender, age, and educational status.

This paper overcomes these deficiencies. Specifically, based on responses from the 2019 National Survey on Drug Use and Health, the metro and nonmetro population of the nation were classified based on their underlying medical conditions; a range of medical conditions that are correlated with the severity of Covid-19 illness is presented. Research questions that guide the data analysis include:

---

[1] Athiyaman is Professor at the Illinois Institute for Rural Affairs.
[2] https://coronavirus.jhu.edu/us-map
[3] The definition of severe illness includes events such as hospitalization, admission to the ICU, intubation or mechanical ventilation, or death (see: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html).
[4] Recent research from CDC suggests that severe illness is mostly associated with persons with the following medical conditions: cancer, chronic kidney disease, COPD, heart conditions, obesity, pregnancy, sickle cell disease, smoking, solid organ transplantation, and type2 diabetes (see: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html).

(i) Who is prone to severe illness from Covid-19 (for example, gender and race) and what proportion of them live in metro and nonmetro regions?

(ii) Is Covid-19 severity more for a person with preexisting medical condition? If "yes", what are the predictors of severe illness (for example, cancer, COPD, obesity, type2 diabetes)?

The information presented in this paper should be of interest to policymakers seeking to influence healthcare in communities.

## Methodology

### Theoretical Background

The economic theory of market segmentation is used as the basis for the classification[5]. The focus is on partitioning regional population into mutually exclusive categories; the basis of such a classification is medical condition, for example, COPD[6].

### Data

Data for the study on medical conditions are from the 2019 National Survey on Drug Use and Health (NSDUH). The survey data included responses on health topics from 67,625 individuals, 12 years of age and over[7].

Our interest is in the correlates of "severity of Covid-19 illness"; the NSDUH provides information about six predictors - cancer, chronic kidney disease, COPD, heart conditions, obesity, smoking, and type2 diabetes. We extracted the frequency or number of cases in each of the categories from NSDUH and weighted the frequency counts to be representative of the nation.

Demographic variables were used to profile the segments of the population with medical conditions. These include:

i. The respondent's place of residence; the variable had three levels: large metro (more than 250,000 inhabitants); small metro (counties and MSA with less than 250,000 people), and nonmetro (micropolitan counties in MSA and noncore counties not in MSA); see www.cdc.gov/nchs/data_access/urban_rural.html);

ii. Gender; two levels, female and male;

iii. Race; four levels: white, black, Hispanic, and other;

iv. Education; three levels: less than high school, high school graduate, and college education;

v. Family income; three levels: < $50,000, $50,000 to $74,999, and GTE$75,000

[5] The theory explains how a firm selling a homogenous product to a market characterized by heterogenous demand could maximize profits.

[6] Although we use the concept of market segmentation to describe the task of clustering the population into groups, the term 'product differentiation' is also applicable for the task; in lay terms, product differentiation is a supply-side definition of clustering population into groups and market segmentation is a demand-side definition.

[7] See https://www.datafiles.samhsa.gov/info/browse-series-nid3453 for information about the 2019 survey.

To relate severity of Covid-19 with medical conditions, a dataset that contained information about Covid-19 hospitalizations was used[8]. The clinical variables used in the analysis include: serum glucose, urea, creatinine, sodium, eosinophils, hematocrit, lymphocytes, leukocytes, urine ketone, urine protein and urine sugar; age was the only demographic variable available in the dataset.

## Statistical Analysis

Regional differences in prevalence of medical conditions were examined using $\chi^2$ tests of independence. The severity of Covid-19 among segments of the population with underlying medical conditions were assessed using a dummy, endogenous variable, econometric model.

## Results

*Differences among the Regional Population[9]*

The nonmetro population is unhealthy; the smaller the population, the higher is the prevalence of one or more medical conditions among the inhabitants of the region
(Table 1). Put differently, approximately one-in-four residents (24%) in large metro report no underlying medical conditions[10]; the same number for nonmetro is 8%[11].

**Table 1: Medical Condition by Geography[12]**

| Medical Condition | Large Metro | Small Metro | Nonmetro | All |
|---|---|---|---|---|
| Cancer | 6% | 6% | 7% | 6% |
| Smoking | 52% | 57% | 60% | 54% |
| COPD | 3% | 4% | 6% | 4% |
| Type2 Diabetes | 9% | 11% | 11% | 10% |
| Heart Disease | <1% | <1% | 1% | <1% |
| Kidney Disease | 2% | 2% | 2% | 2% |
| Obesity | 4% | 5% | 5% | 5% |
| None of the above[13] | 24% | 15% | 8% | 18% |
| Pop 12+ age (100%) | 153,671,414 | 83,421,567 | 38,128,267 | 275,221,248 |

Of the medical conditions listed in Table 1, type2 diabetes and obesity are common among minorities living in large metros; in all regions, white population suffers the most from heart and kidney diseases (Table 2).

---

[8] Hospital Israelita Albert Einstein in Brazil posted the data online for crowdsourcing predictive models of Covid-19. The dataset contained 5644 cases and 111 variables; see https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/discussion/139347

[9] Data for the regions were subjected to Chi-square tests of independence; only statistically significant differences are highlighted and discussed in this section.

[10] Inference is based on the seven medical conditions listed in column 1.

[11] One could **wrongly** conclude that the difference is due to concentration of elderly in the nonmetro; Chi-square test indicates null association between age of residents and geographical location.

[12] The obtained Chi-square value exceeds the critical value of 22 ($\alpha$=0.05); hence we reject the hypothesis of independence between the row and the column variables.

[13] Obtained by subtraction, 100% - Σ medical condition.

**Table 2: Impact of Race on Medical Conditions**

| Race | Geography | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Large Metro | | | | Small Metro | | | | Nonmetro | | | |
| | D | H | K | O | D | H | K | O | D | H | K | O |
| Black, non-Hispanic | 19% | 15% | 15% | 26% | 10% | 7% | 8% | 19% | 11% | 5% | 2% | 11% |
| Hispanic | 26% | 16% | 13% | 21% | 16% | 10% | 11% | 17% | 6% | 4% | 4% | 6% |
| White, non-Hispanic | 48% | 62% | 62% | 48% | 67% | 72% | 76% | 60% | 76% | 79% | 90% | 79% |
| Other | 11% | 7% | 10% | 4% | 8% | 10% | 6% | 5% | 6% | 11% | 4% | 3% |
| **N**umber in the Population | See Table 1 | | | | | | | | | | | |

**Note:** D = Type2 diabetes; H = Heart disease; K = Kidney disease, and O = obesity. All Chi-square tests for differences among the regions were significant at the p <.05 level.

Majority of people with medical conditions have no college education (Table 3). Furthermore, while COPD is associated with families earning less than $50,000 across the regions; cancer is widespread among families earning more than $75,000 in the metro region. Middle income families, those earning between $50,000 - $74,999, are least affected by cancer and COPD (Table 4).

**Table 3: Impact of Education on Medical Conditions[1]**

| Education | Geography | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Large Metro | | | Small Metro | | | Nonmetro | | |
| | C | H | K | C | H | K | C | H | K |
| Less than High School | 7% | 19% | 9% | 7% | 18% | 15% | 9% | 27% | 19% |
| High School | 45% | 46% | 62% | 55% | 63% | 66% | 66% | 58% | 68% |
| College | 48% | 35% | 30% | 37% | 19% | 20% | 25% | 16% | 13% |
| **N**umber in the Population | See Table 1 | | | | | | | | |

**Note:** C = Cancer; H = Heart disease; and K = Kidney disease. All Chi-square tests for differences among the regions were significant at the p <.05 level.

**Table 4: Impact of Family Income on Medical Conditions: Cancer and COPD**

| | Geography | | | | | |
|---|---|---|---|---|---|---|
| **Income** | **Large Metro** | | **Small Metro** | | **Nonmetro** | |
| | **C** | **COPD** | **C** | **COPD** | **C** | **COPD** |
| Less than $50, 000 | 33% | 54% | 42% | 63% | 54% | 72% |
| $50,000 - $74,999 | 16% | 19% | 18% | 11% | 20% | 15% |
| Greater than $75,000 | 51% | 27% | 41% | 26% | 26% | 13% |
| **N**umber in the Population | See Table 1 | | | | | |

**Note:** C = Cancer. All Chi-square tests for differences among the regions were significant at the p <.05 level.

In summary, diabetes and obesity are common among minorities living in large metros; white inhabitants suffer the most from heart and kidney diseases. Typically, noncollege educated people, with family income less than $50,000, suffer from medical conditions such as COPD. Appendix 1 provides multiway tables for medical conditions, for each geo-graphical region cross-classified with demographic variables.

***Severity of Covid-19***

Consider the following conceptualization:

$y^*1$ = expected severity of Covid-19 illness with preexisting medical condition;

$y^*2$ = expected severity of Covid-19 illness without medical condition.

The hypothesis is $y^*1 > y^*2$.

Our hypothesis is premised on the notion that an individual with one or more of the medical conditions listed in Table 1 will suffer severe illness if the individual contracts Covid-19.

Econometrically,

$y^*1 = \alpha 1 + u1$;
$y^*2 = \alpha 2 + u2$[14]

The difference $\alpha 1 - \alpha 2$ measures the average increase in severity of illness due to preexisting medical condition(s). Probability of severe illness is:

$$P\,(y^*1 > y^*2) = P\,(u1 - u2 > \alpha 2 - \alpha 1) = 1 - \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t2/2} dt \quad \text{where } z = \frac{\alpha 2 - \alpha 1}{\sigma}$$

[14] It is assumed that u1 and u2 are normally distributed.

An estimate of this quantity is the number of people admitted into a hospital for Covid-19 with preexisting medical conditions (n1) over all patients admitted to the hospital with Covid-19 (n1 + n2). Given this number, we can estimate z. From the hospital database, we estimate z at 1.28. In other words, statistically there is higher chance for an individual with preexisting medical condition to suffer severe illness from Covid-19. Further statistical analysis suggests that severity of illness is higher for individuals with cancer, COPD, and kidney ailments (Table 5).

**Table 5: Predictors of Sever Illness from Covid-19**

Probit Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | severe | No. Observations: | 558 |
| Model: | Probit | Df Residuals: | 549 |
| Method: | MLE | Df Model: | 8 |
| Date: | Sun, 29 Nov 2020 | Pseudo R-squ.: | 0.4677 |
| Time: | 16:08:30 | Log-Likelihood: | -92.041 |
| converged: | True | LL-Null: | -172.90 |
| Covariance Type: | nonrobust | LLR p-value: | 6.993e-31 |

| | coef | std err | z | P>\|z\| |
|---|---|---|---|---|
| const | -2.6148 | 0.298 | -8.781 | 0.000 |
| age | 0.0610 | 0.022 | 2.725 | 0.006 |
| Serum_Glu | -0.2403 | 0.332 | -0.723 | 0.469 |
| Urea | -0.5197 | 0.314 | -1.654 | 0.098 |
| Creatinine | 0.6763 | 0.296 | 2.283 | 0.022 |
| Sodium | 0.0099 | 0.231 | 0.043 | 0.966 |

**Note:** Clinically, cancer could be indicated by Lymphocytosis and Leukocytosis; COPD by the clinical Eosinophilia; Creatinine could indicate kidney disease.

## Summary and Conclusion

This research empirically demonstrates the positive relationship between severe illness from Covid-19 and preexisting medical conditions. Two sets of data are used to gain insights into population that are vulnerable to severe Covid-19 infection: 2019 National Survey on Drug Use and Health and patient data from Hospital Israelita Albert Einstein, Brazil. Patient data suggest that cancer, COPD, and kidney disease predict severity of Covid-19 illness. White population in both the metro and the nonmetro regions suffer from kidney ailments. COPD is typical of low-income families, income below $50,000. Across both the metro and the nonmetro, it is people with no college education that suffer from cancer, heart, and kidney disease. It is hoped that the data presented here will be a useful guide for healthcare policy and decision making at the metro / nonmetro level; see Appendix 1 for multi-way tables.