Social Vulnerability Index and COVID-19

BY
HAEWON OH
B.S., Dankook University, 2014
M.A., Seoul National University, 2016

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:

Sage Kim, Chair and Advisor
Caryn E Peterson, Epidemiology and Biostatistics
Jiehuan Sun, Epidemiology and Biostatistics
Sreenivas Konda, Epidemiology and Biostatistics

Acknowledgments

I would like to thank my thesis committee--Professors Sage Kim, Caryn E Peterson, Jiehuan Sun, and Sreenivas Konda--for their support and advice. They have provided me with guidance in all areas of my thesis project that helped me to achieve my research goals in the process. In addition, I thank Ms. Shelley Hoover for working with me on this study.

Contribution of Authors

Dr. Kim provided overall guidance and advice for this study as the corresponding author. Ms. Hoover analyzed the relationship of the Social Vulnerability Index and COVID-19 Community Vulnerability Index to the COVID-19 infection and fatality rates for selected counties in Illinois (Table I and Appendices A-C and F-K). I was responsible for performing the literature review, determining the relationships between the indexes and the infection and fatality rates for all counties in Illinois, identifying the index themes that most contributed to those rates, assessing the relative importance of the themes to the rates and index composite scores, and writing this manuscript from the Introduction through the Discussion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF FIGURES (CONTINUED)

FIGURE PAGE

LIST OF ABBREVIATIONS

CCVI	COVID-19 Community Vulnerability Index
SVI	Social Vulnerability Index
PCA	Principal Component Analysis

SUMMARY

Although the COVID-19 crisis has had global impacts, COVID-19 has disproportionately affected poor, segregated racial/ethnic populations on the local level across the United States. To understand the effects of COVID-19 on such populations, various indexes have been developed to help identify the communities most likely to be impacted by the disease. These indexes include the Social Vulnerability Index (SVI) and the COVID-19 Community Vulnerability Index (CCVI). However, it is not clear whether or how well the SVI or CCVI explains COVID-19 infection and death. To determine the efficacy of the SVI and CCVI, this study (1) uses Poisson regression analysis to compare the ability of SVI and CCVI themes to explain actual Illinois county infection rates and case fatality rates, (2) examines the relative contributions of the four SVI themes and six CCVI themes to the composite score using the Principal Component Analysis, Random Forest model, and relative importance metrics in a linear regression model, and (3) applies the three methods to explain case fatality rates. The study revealed that the SVI better explains COVID-19 infection and case fatality rates compared to the CCVI, with the Minority Status and Language theme of both indexes appearing to be related to infection rates. The Housing Type and Transportation theme of the SVI was also significantly related to the infection rate. In addition, using Principal Component Analysis, Random Forest, and relative importance metrics in linear regression analysis to rank the importance of SVI and CCVI themes, the Minority Status and Language theme was found to make the least contribution to the composite score at the county level. The Minority Status and Language theme was the most important CCVI factor in explaining the infection rate, but it appeared to be the least important to the composite score. This would explain why the CCVI composite score showed no

SUMMARY (CONTINUED)

significant relationship to the infection rate. The SVI's Minority Status and Language theme was an important factor in explaining the infection rate, and this theme was the least important to the SVI composite score. However, unlike the CCVI, this theme contributed to a significant positive relationship between the SVI composite score and the infection rate. Given these findings, the CCVI would not be appropriate for explaining overall COVID-19 infection rates. When the three analysis methods were applied to explain the case fatality rate, the Random Forest model and Principal Component Analysis revealed that the Socioeconomic Status theme of the SVI made the greatest contribution to that rate. Thus, I conclude that the case fatality rate is strongly related to Socioeconomic Status factors such as income, as the severity of COVID-19 is dependent on the ability to access medical treatment, and that the infection rate is related to the Minority Status and Language theme. For the CCVI, all three methods produced very different results. On the whole, I conclude that the SVI is a more appropriate index than the CCVI for explaining overall COVID-19 infection and case fatality rates. One issue with the CCVI is that the variables used for the Epidemiological Factors and Healthcare System Factors themes are measured at the state or county level. Thus, I argue that indexes such as the CCVI need to be constructed at the community and census tract levels.

I. INTRODUCTION

Since COVID-19 spread around the world, the United States has had the greatest number of confirmed COVID-19 cases--36,943,389--at the end of 2020 (World Health Organization, 2020). Of the top five states in regard to total cases, California, Texas, Florida, New York, and Illinois, Illinois is known to have the most segregated communities (The New York Times, 2020). As of January 5, 2021, there had been 991,719 confirmed COVID-19 cases and 16,959 deaths in the state (Illinois Department of Public Health, 2021). COVID-19 is known to disproportionately affect African American and Hispanic communities, in Illinois. In Chicago, a highly segregated city, there were 144,410 confirmed COVID-19 cases as of November 17, 2020. Of these cases, 37% involved Hispanic Americans, 18.2% involved African Americans, and 17.3% involved White Americans. COVID-19-related deaths show even greater racial/ethnic discrepancies, as African Americans showed 41.1% of the deaths, Hispanic Americans showed 33.3%, and White Americans showed 20.2% (Chicago Department of Public Health, 2020). The health inequities associated with these values are clear given that Chicago's population is approximately 30% non-Hispanic African American, 29% Hispanic, and 33% White (US Census Bureau, 2020).

To determine the reasons for differences in the outcomes of hazardous events between communities as well as to effectively respond to such events, an accurate vulnerability index is required (Centers for Disease Control and Prevention [CDC], 2018). As COVID-19 has spread globally, many studies have shown relationships between various vulnerability indexes and virus-related rates of infection, morbidity, and mortality. Internationally, some researchers have constructed specific countries' vulnerability index and investigated the relationship between the

vulnerability indexes and COVID-19 cases and deaths (Daras et al., 2020; Judson et al., 2020; Macharia, 2020; Martines et al., 2020; Mishra et al., 2020; Nishant et al., 2020; Sarkar et al., 2020; Wong et al., 2020). For example, Sarkar et al. (2020) showed that COVID-19 outcomes in India were correlated with socioeconomic conditions identified using a vulnerability index and that COVID-19 impacts were more severe in less advantaged communities. Similarly, in applying their vulnerability index in some areas of England, Daras et al. (2020) found that COVID-19 was likely to disproportionately impact communities designated as highly vulnerable.

In the United States, CDC created the Social Vulnerability Index (SVI) to help identify the U.S. communities most likely to need assistance when a hazardous event occurs (CDC, 2018). This index, which can be applied to events involving natural disasters and hazards such as COVID-19, consists of four themes: (1) Socioeconomic Status, (2) Household Composition and Disability, (3) Minority Status and Language, and (4) Housing Type and Transportation (CDC, 2018). As has occurred in other countries, American researchers have attempted to interpret the relationship between the SVI and COVID-19 outcomes (Dasgupta et al., 2020; Garcia et al., 2020; Gaynor et al., 2020; Hathaway, 2020; Karaye et al., 2020; Khazanchi et al., 2020; LeRose et al., 2020; Nayak et al., 2020; Snyder et al., 2020; Wang et al.,2020). However, only a few studies have used statistical data to examine this relationship. Some of these studies showed that social vulnerability was associated with higher COVID-19 incidence and mortality across the United States (Dasgupta et al., 2020; Karaye et al.,2020; Khazanchi et al. 2020; Nayak et al., 2020; Wang et al.,2020). For example, Khazanchi et al. (2020) found that disproportionate COVID-19 impacts in both urban and rural areas were driven by the SVI's Minority Status and Language theme. In addition, some studies found that social vulnerability factors impacted COVID-19 death rates in specific cities and showed that racial inequity was associated with

COVID-19 incidence (Bilal et al., 2020; Kim et al.,2020). In those studies, social vulnerability was significantly associated with the higher COVID-19-related death rates in the cities.

In an effort to adjust the existing SVI to meet the particular demands of the COVID-19, many studies developed new vulnerability indexes (Amram et al., 2020; Anuj et al., 2020; DeCaprio et al., 2020; Marvel et al., 2020; Wiemers et al., 2020). For example, Anuj et al. (2020) developed index of county level vulnerabilities and compare it with the COVID-19 Community Vulnerability Index (CCVI) constructed by the Surgo Foundation (2020) using Random Forest approach. Surgo Foundation (2020) developed CCVI based on SVI and incorporated two new themes--Epidemiological Factors and Healthcare System Factors--in addition to the four existing SVI themes. The SVI's four themes and the CCVI's six themes are summarized in Table I (CDC 2018; Surgo Foundation 2020). In my study, the CCVI was selected for comparison to the CDC's SVI because it expands on the SVI to include additional COVID-19-specific themes.

Geographically, data for the first four themes of both the SVI and CCVI have been aggregated at the census tract level. Of the 15 variables that comprise the CCVI's additional two themes, only data for population density have been collected at the census tract level. Data for the other variables have been collected at broader geographic levels such as the Hospital Referral Region, county, and state-wide levels and then equally applied to the census tract and county levels.

TABLE I. COMPOSITION OF SVI AND CCVI THEMES

Theme No.	Index	Theme	Variables	Geo Precision
1	SVI & CCVI	Socioeconomic Status	Below Poverty, Unemployed, Income, No High School Diploma	All Census Tract
2	SVI & CCVI	Household Composition and Disability	Aged 65 or Older, Aged 17 or Younger, Civilian with a Disability, Single-Parent Household	All Census Tract
3	SVI & CCVI	Minority Status and Language	Minority, Speaks English "Less than Well"	All Census Tract
4	SVI & CCVI	Housing Type and Transportation	Multi-Unit Structures, Mobile Homes, Crowding, No Vehicle, Group Quarters	All Census Tract
5	CCVI Only	Epidemiological Factors	Cardiovascular Conditions, Respiratory Conditions, Immunocompromised, Obesity, Diabetes, Population Density, Influenza and Pneumonia Death Rates	Census Tract, County
6	CCVI Only	Healthcare System Factors	Hospital Beds per 100,000 Intensive Care Unit (ICU) Beds per 100,000 Epidemiologist Jobs per 100,000 Health Spending per Capita Agency for Healthcare Research and Quality-Prevention Quality Indicator Overall Composite Total Public Health Emergency Preparedness (PHEP) Funding per Capita Health Labs per 100,000 Emergency Services per 100,000	Hospital Referral Regions (see Appendix A), State, County

The studies conducted to date have produced unclear results as to which SVI themes show significant relationships with COVID-19 cases and deaths in particular communities. Therefore, to help determine how well vulnerability indexes explain COVID-19 infection and case fatality rates, this analysis aimed to compare the ability of the SVI and CCVI to explain these rates and to examine whether the two additional themes of the CCVI improved its ability to explain these rates. I first performed a descriptive analysis of whether SVI and CCVI explained COVID-19 fatality and infection rates in Illinois counties as expected. My next step was to perform a more detailed analysis of each theme with respect to COVID-19 fatality and infection rates for Illinois counties. Finally, I conducted an importance assessment of the SVI and CCVI themes with respect to their contribution to composite scores and case fatality rates using three different approaches: Principal Component Analysis (PCA), feature importance in machine learning approach-- Random Forest--and relative importance metrics in a linear regression model for Illinois. On the whole, my analysis (1) compared the CDC's SVI to the CCVI and (2) evaluated the importance of the four SVI themes and six CCVI themes to the indexes' composite scores and to the COVID-19 infection and case fatality rates.

II. METHODS

For this study, I used the Centers for Disease Control and Prevention Social Vulnerability Index 2018 and the COVID-19 Community Vulnerability Index updated in November 13, 2020. CCVI is an effort to adjust the existing SVI in order to meet the specific needs posed by the COVID-19 pandemic. Thus, CCVI incorporates the four SVI themes and adds two new themes for a total of six: Socioeconomic Status, Household Composition, Minority Status, Housing and Transit Type, Epidemiologic Risks, and Healthcare Systems. With both SVI and CCVI, the composite score can be used to designate a level of vulnerability. For both indexes, the composite scores range from 0 to 1, with higher values indicating greater vulnerability. Variables for each theme are represented by its percentile rank, and the percentiles for the variables are summed to represent a theme. To create a composite score, percentiles are aggregated across all six themes with equal weight (Surgo Foundation, 2020). This information is available on the Surgo Foundation's CCVI website. For this study, I employed the total number of COVID-19 infection cases at the Illinois county level and the total number of COVID-19 deaths published by the Illinois Department of Public Health as of November 17, 2020. Employing the total population value for each county from the U.S. Census, infection rate was calculated as "total number of infection cases/total population," and case fatality rate, a measure of the severity of a disease, was calculated as "total number of deaths/total number of infection cases." (World Health Organization, 2020) The first step was to perform a descriptive analysis of the SVI and CCVI themes. In order to visualize the ability of the two indexes, I created a series of grids comparing the infection rate and case fatality rate of each Illinois county to the index scores. I analyzed 4x4 and 10x10 grids comparing the quartile and decile rankings of index scores to infection rates and case fatality rates. If the indexes are strongly associated with the rates, I

would expect most counties with higher vulnerability to have higher infection and case fatality rates compared to a county with a lower score (Figure 1).

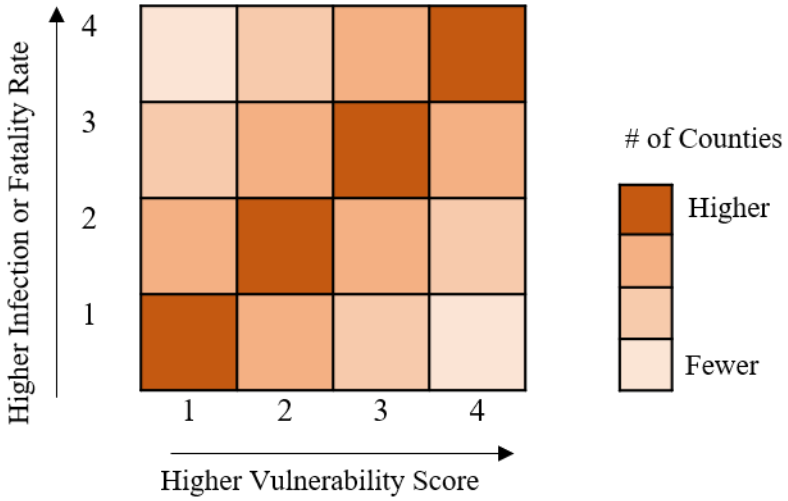

Figure 1. Expected grid outcome

As the next step, I analyzed the ability of SVI and CCVI themes with respect to explain Illinois county infection rates and case fatality rates using a Poisson regression model. Finally, I analyzed the relative size of the themes' contribution to the SVI or CCVI composite score using relative importance in linear regression analysis, Principal Component Analysis (PCA), and machine learning method-- Random Forest analysis. In addition, I applied these three methods to analyze the themes' contribution to case fatality rate. I decided to use these three approaches because they are the most intuitive ways of ranking the importance of each theme to the index composite score.

Examining which predictor is more important to explain the outcome variable often becomes important interest of research. In the regression framework, the standardized regression coefficients would be compared but the predictors might not be directly compared. It is because correlated predictors' coefficients would not identify which predictor is more important. Therefore, there are several ways to estimate the relative importance of predictors.

First, a parameter importance assessment using linear regression analysis exploits the increased R^2 by adding predictors to the regression model sequentially. By averaging over orderings of a predictor, the increased R^2 can be considered the contribution (Grömping, 2009). Second, Random Forest analysis creates shuffled copies of all features called shadow features to add randomness to a dataset. At this point, the shadow data frame is attached to the original data frame in order to obtain a new data frame. Then a Random Forest classifier is trained on the extended dataset, and a feature importance measure is applied to examine the importance of features; a higher z score indicates higher importance. The Random Forest model checks whether a real feature has a higher importance z score than the maximum z-score of the shadow features and then eliminates features considered to be significantly unimportant at every iteration (Breiman, 1999). Finally, PCA is a good method for extracting important information from highly correlated variables and for reducing the dimensionality of multivariate data to fewer principal components that form a set of a few new variables, which correspond to a linear combination of the original variables (Holland, 2019; Gniazdowski, 2017). The themes of SVI and CCVI data are expected to be correlated to each other. Consequently, we conducted PCA in addition to Random Forest and linear regression analysis.

III. RESULTS

A. Relationships Between the Two Indexes and the COVID-19 Infection and Case Fatality Rate

This study employed quartile and decile grids to evaluate how well SVI and CCVI explained COVID-19 infection and death rates as of November 17, 2020. In the grids, as the number of counties becomes larger, a cell has a darker color (Figure 1). I hypothesized that cells on the diagonal would have the largest number of counties if the SVI or CCVI score had a strong relationship with infection and mortality rates.

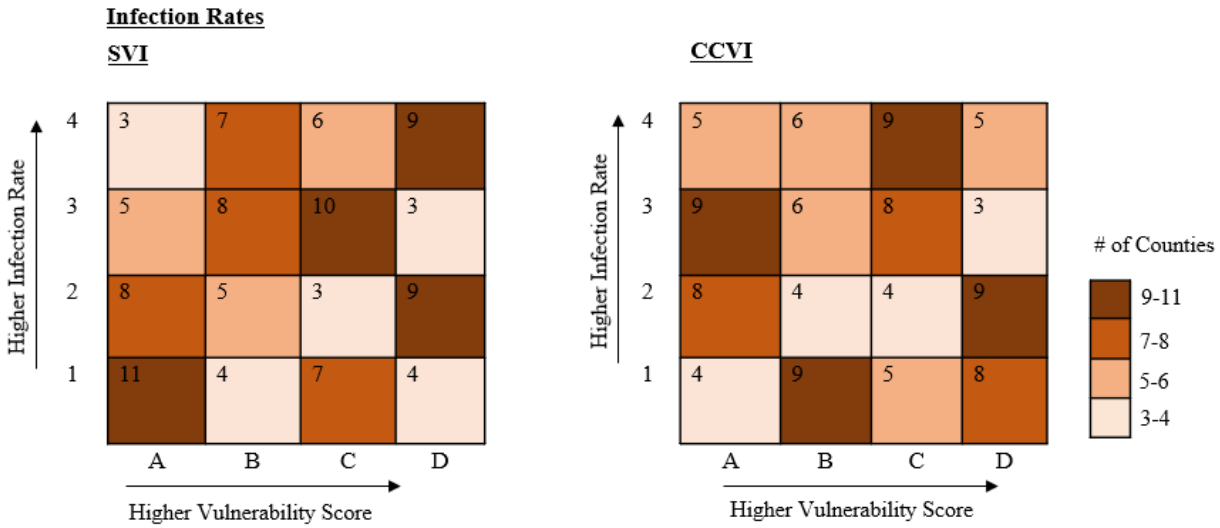

Figure 2. 4x4 Grid: Association of infection rates with SVI and CCVI

As shown in the quartile grid in Figure 2, a large number of counties do not appear on the diagonal in the CCVI grid, indicating that the CCVI is probably not related to the infection rate. In contrast, Figure 2 indicates that the SVI is more likely to be related to the infection rate

because a large number of counties appear on the diagonal. The quartile grid indicates that the SVI is more likely to explain the infection rate accurately, while the CCVI is probably not related to the infection rate.

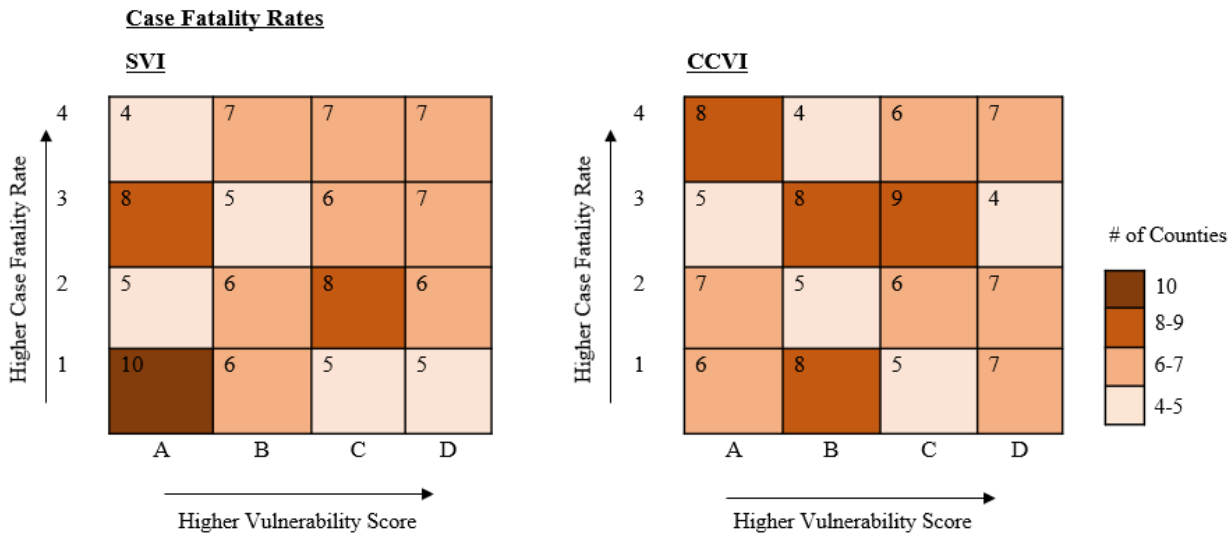

Figure 3. 4x4 Grid: Association of case fatality rates with SVI and CCVI

Figure 3 indicates that the SVI is likely to have a positive relationship with the case fatality rate. However, the CCVI was less likely to be related to that rate. In Figures 2 and 3, the number of counties is more dispersed throughout the CCVI grids than is the case for the SVI grids. Both Figures 2 and 3 indicate that the CCVI is unlikely to have any relationship with the infection rate or case fatality rate. As shown in Appendixes D and E, the decile grids also indicate that the SVI explains both the infection rate and case fatality rate better than the CCVI. In the SVI grid for the infection rate, a larger number of counties appear around the diagonal.

The regression results confirm the significant positive relationship between the SVI and infection and case fatality rates (Table II). The coefficients from the regression of SVI on infection rate and case fatality rate were 0.1410 and 0.3605, respectively, and both values were significant at the 10% significance level. However, the CCVI did not have any significant relationship with either the infection or case fatality rate. Figure 4 shows the relationship of the infection rate to the SVI and CCVI, and Figure 5 shows the relationship of the case fatality rate to the SVI and CCVI. Both figures show a fitted line corresponding to the regression results.

TABLE II. POISSON REGRESSION OF INFECTION/CASE FATALITY RATE ON SVI/CCVI

	Infection Rate	Infection Rate	Case Fatality Rate	Case Fatality Rate
SVI	0.1410 (0.0781)[a]			
CCVI		0.0113 (0.1953)		
SVI			0.3605 (0.1100)	
CCVI				0.0542 (0.2730)

a. Robust standard errors in parentheses.

The Poisson regression of infection rate on each SVI theme showed that only themes 3 and 4 (Minority Status and Language and Housing Type and Transportation) showed significant positive relationships with infection rate (Table III), whereas Socioeconomic Status and Household Composition and Disability (themes 1 and 2) showed no significant relationship with infection rate. In contrast, the Poisson regression of case fatality rate on SVI themes showed no

significant relationship with any SVI themes (Table III). Nonetheless, the coefficients for the Socioeconomic Status and Household Composition and Disability themes were larger than those for Minority Status and Language and Housing Type and Transportation. In Illinois, which contains one of the most segregated cities in the United States as a population center, minority ethnic populations would be expected to be the most susceptible to infection by contagious disease, but fatalities would be more directly related to socioeconomic status factors such as income and to disability. In addition, people who live in multiple-unit housing structures or who cannot work from home and have to commute to work via public transportation would be particularly at risk of COVID-19 exposure.

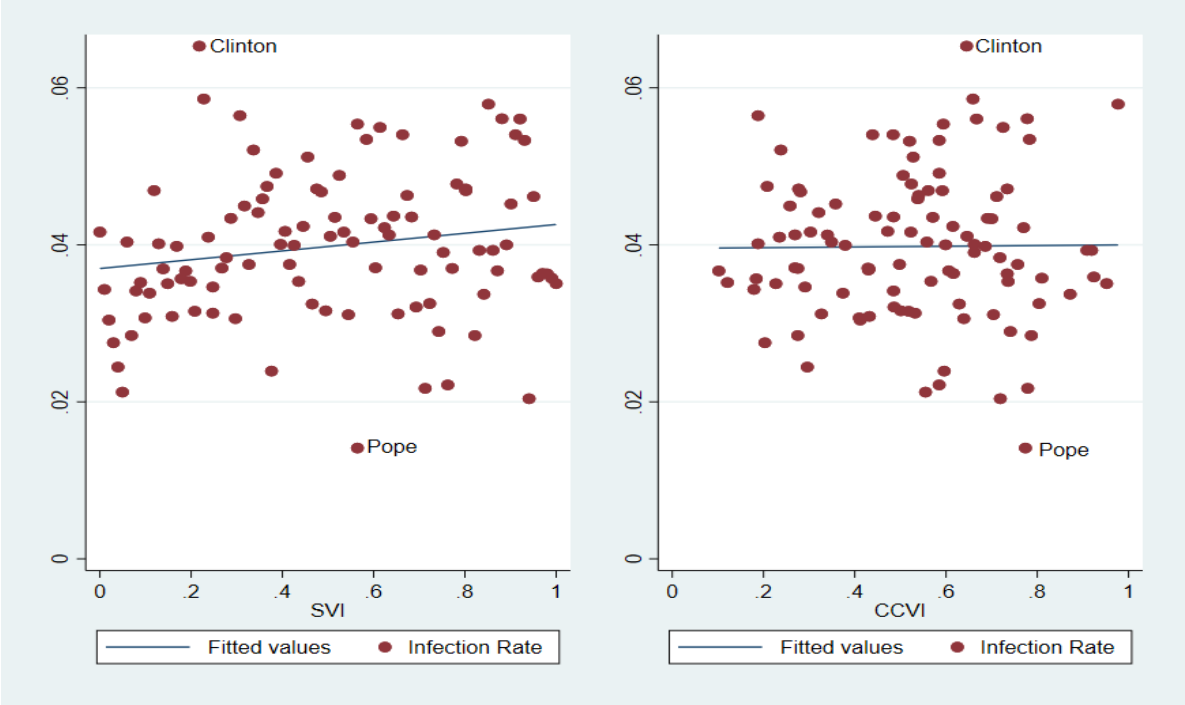

Figure 4. Scatter plot of relationships of infection rates to SVI and CCVI

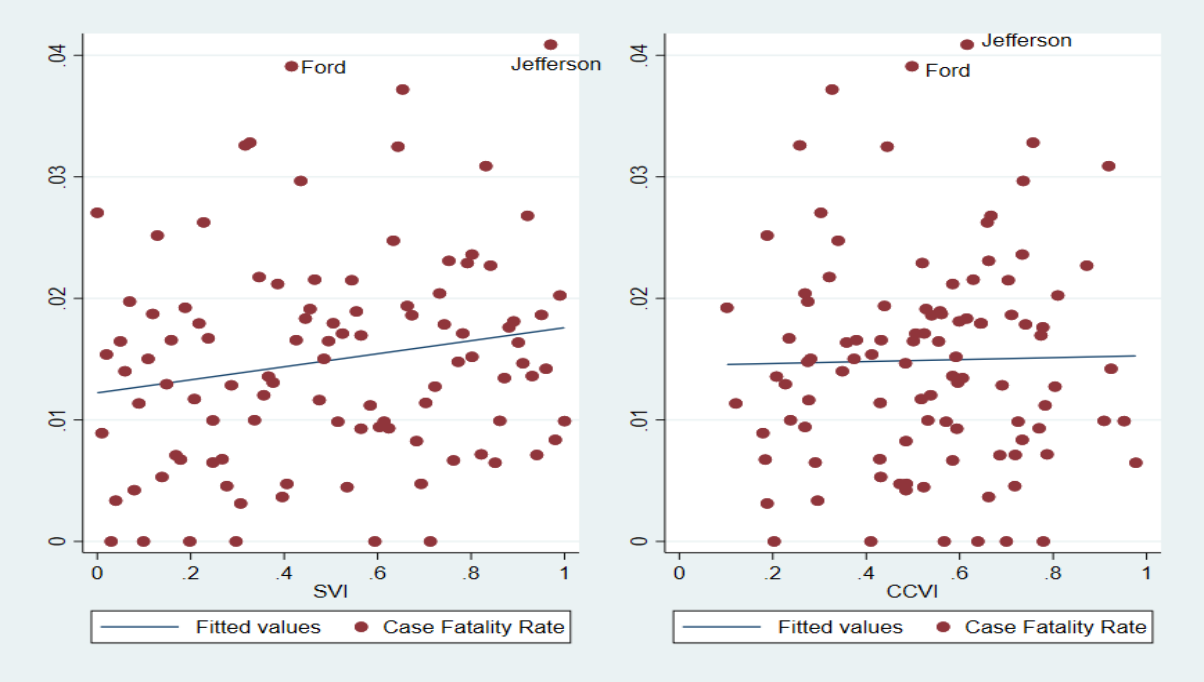

Figure 5. Scatter plot of relationships of case fatality rates to SVI and CCVI

TABLE III. POISSON REGRESSION OF INFECTION/CASE FATALITY RATE ON SVI THEMES

	Infection Rate	Case Fatality Rate
Socioeconomic Status	-0.1330 (0.1593)[a]	0.3115 (0.3095)
Household Composition and Disability	0.0437 (0.0969)	0.3180 (0.2470)
Minority Status and Language	0.1725 (0.0828)	-0.0802 (0.2010)
Housing Type and Transportation	0.2188 (0.1320)	-0.0594 (0.2842)

a. Robust standard errors in parentheses

Figures 6 and 7 show the relationship of each SVI theme to the infection rate and case fatality rate, respectively, and the fitted line in the figures corresponds to the simple regression results for each theme. Because of confounding effects in the multiple regression models, the regression coefficients in Table III do not directly correspond to the fitted line in the figures. In Figure 6, the Minority Status and Language theme and Housing Type and Transportation theme show a strong positive relationship with the infection rate, whereas the Socioeconomic Status theme and Household Composition and Disability theme show little association with this rate. In contrast, Figure 7 shows that the Socioeconomic Status theme and Household Composition and Disability theme have a positive relationship with the case fatality rate, whereas the Minority Status and Language theme and Housing Type and Transportation theme show little association with this rate.

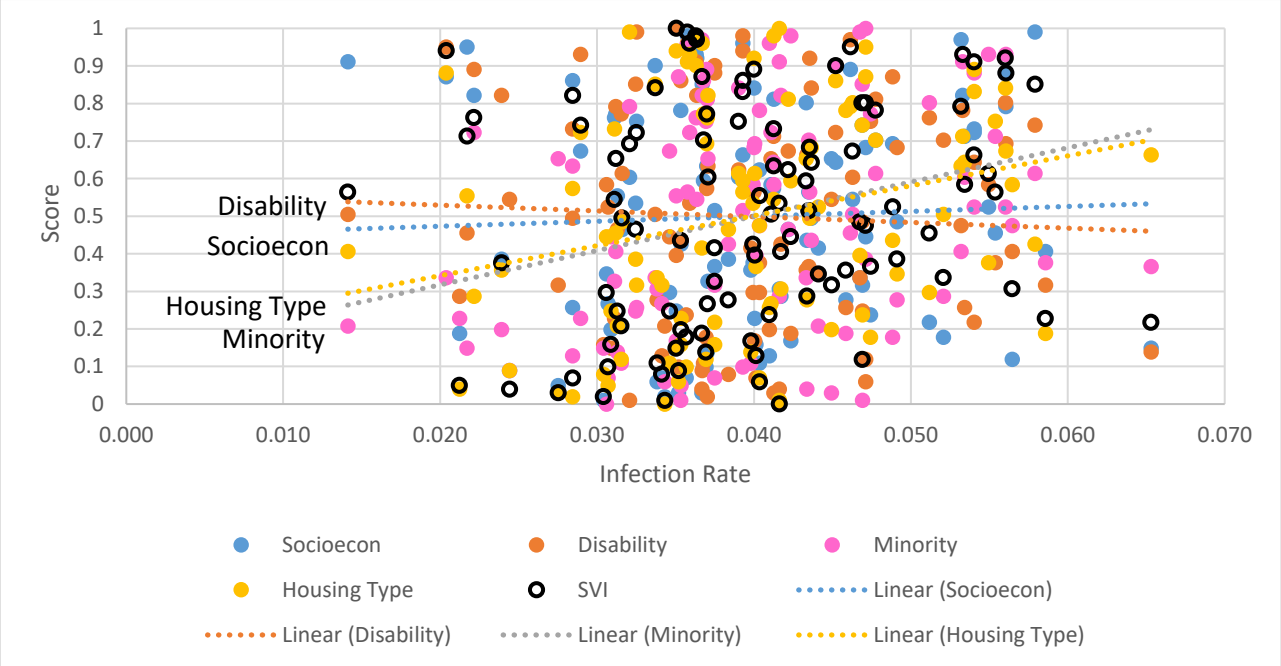

Figure 6. SVI themes and infection rates

Figure 7. SVI themes and case fatality rates

Table IV shows that only Minority Status and Language had a significant positive association with the infection rate (Coefficient = 0.2887) and that Healthcare System Factors was negatively associated with the case fatality rate (Coefficient = -0.7255). One potential reason for this result is that people who become infected with COVID-19 would naturally prefer to obtain treatment in a county having a good healthcare system, which would lead to higher infection and case fatality rates in counties having such a system. Figures 8 and 9 show the associations of each CCVI theme to the infection rate and case fatality rate, respectively, and the fitted line in the figures corresponds to the simple regression results for each theme. In Figure 8, the Minority Status and Language theme and Housing Type and Transportation theme show a positive relationship with the infection rate, whereas the Healthcare System theme shows a negative

relationship with this rate. The Socioeconomic Status, Household Composition and Disability, and Epidemiological Factors themes show little association with the infection rate. Figure 9 shows that the Socioeconomic Status theme and Household Composition and Disability theme have a positive relationship with the case fatality rate and that the Healthcare System theme has a negative relationship with this rate. The Minority Status and Language, Housing Type and Transportation, and Epidemiological Factors themes show little association with the case fatality rate.

TABLE IV. POISSON REGRESSION OF INFECTION/CASE FATALITY RATE ON CCVI THEMES

	Infection Rate	Case Fatality Rate
Socioeconomic Status	-0.1582 (0.2262)	0.2572 (0.4465)
Household Composition and Disability	0.0640 (0.1440)	0.3621 (0.3037)
Minority Status and Language	0.2887[a] (0.1151)	-0.1668 (0.2842)
Housing Type and Transportation	0.1600 (0.1258)	-0.0813 (0.2660)
Epidemiological Factors	0.0104 (0.0773)	-0.0558 (0.2161)
Healthcare System Factors	-0.1524 (0.1039)	-0.7255 (0.2790)

a. Robust standard errors in parentheses

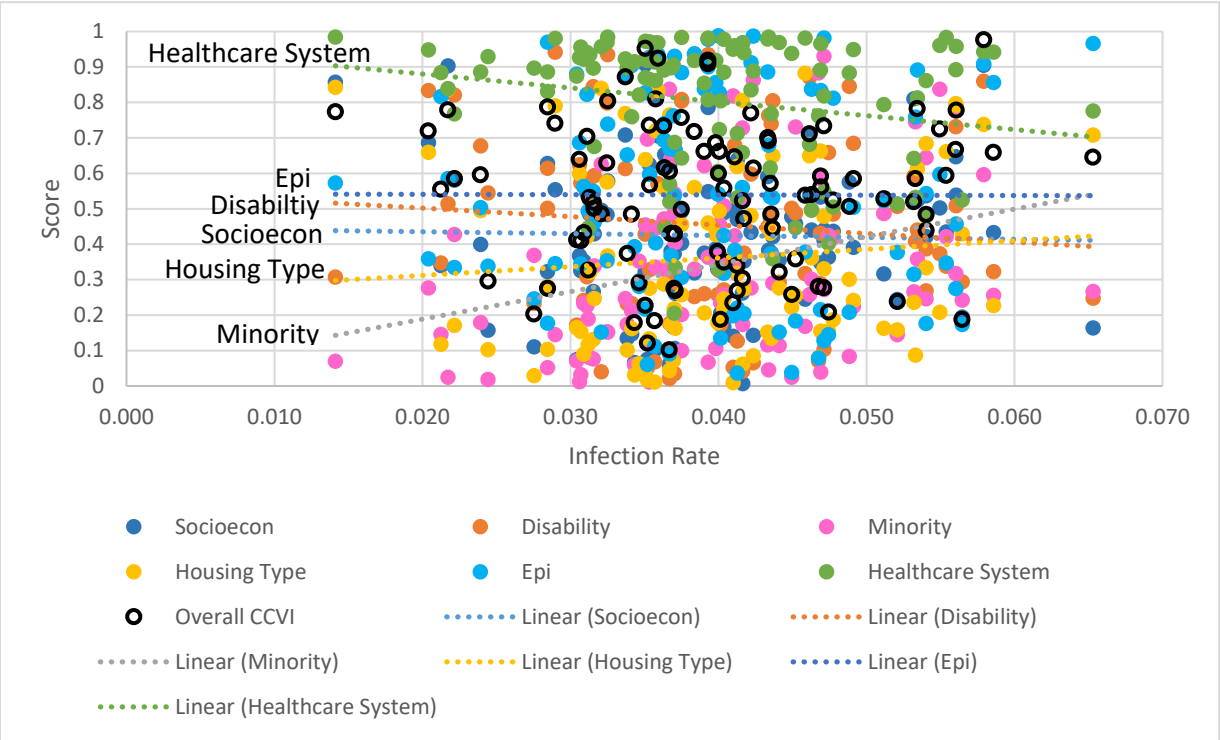

Figure 8. CCVI themes and infection rates

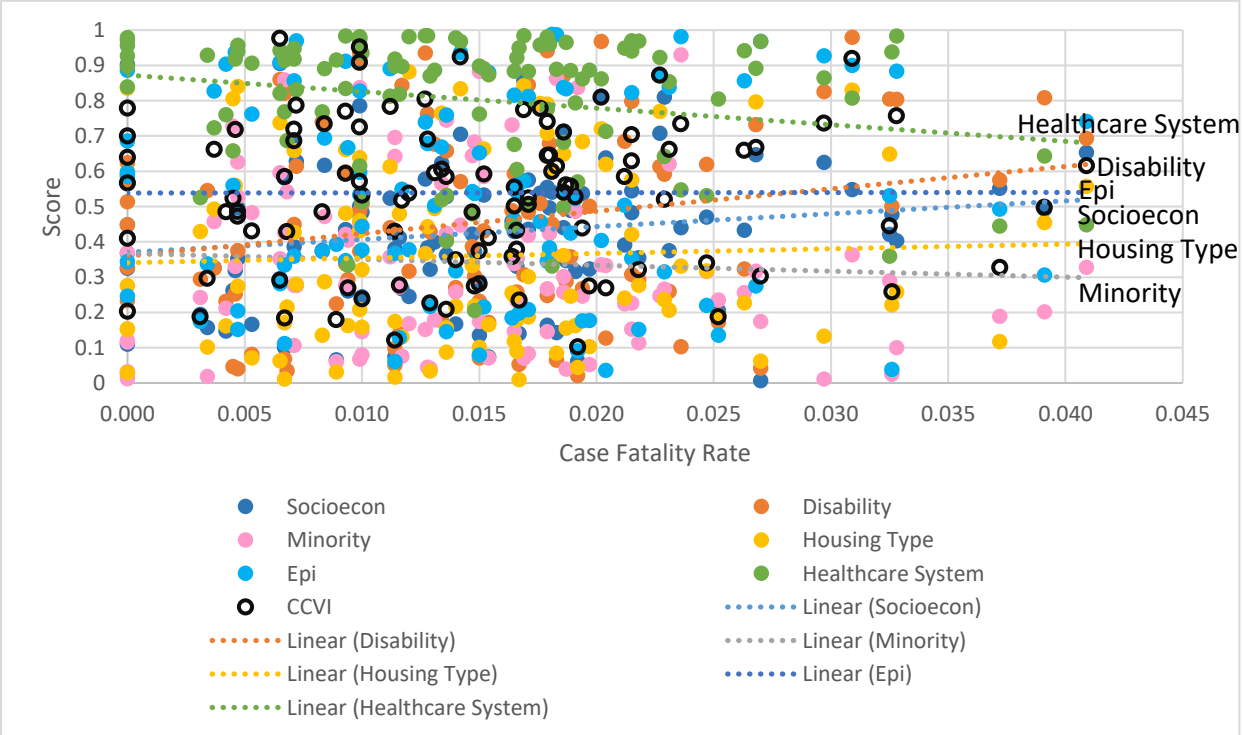

Figure 9. CCVI themes and case fatality rates

B. Relative Contributions of the Four Themes of SVI to the Composite Score

To measure the contribution of each theme to the composite score of the SVI, I used PCA, linear regression techniques, and Random Forest modeling. In the regression framework, the predictors might not be directly comparable because correlated predictors' coefficients would not identify which predictor is more important. Thus, I checked the correlation of themes and this Table V shows that SVI themes were statistically correlated with each other. Table V shows the correlations among SVI themes at the county level and that some themes are highly correlated with each other.

TABLE V. THE CORRELATIONS OF SVI THEMES

Themes	Socioeconomic Status	Household Composition and Disability	Minority Status and Language	Housing Type and Transportation	SVI
Socioeconomic Status	1				
Household Composition and Disability	0.5193	1			
Minority Status and Language	0.0751	-0.178	1		
Housing Type and Transportation	0.6059	0.0815	0.3881	1	
SVI	0.8881	0.5137	0.3893	0.7978	1

At the county level for Illinois, the relative importance for linear regression model showed that among the four SVI themes, Socioeconomic Status had the largest R^2 (0.417), followed by Housing Type and Transportation (0.32), Household Composition and Disability (0.148), and Minority Status and Language (0.097). However, the confidence interval using the bootstrap method revealed that there was no significance difference between the Household Composition and Disability and Minority Status and Language themes in terms of relative importance.

The Random Forest model showed that at the county level, Socioeconomic Status had the highest mean Z score (31.2), followed by Housing Type and Transportation (23.86), Household Composition and Disability (16.26), and Minority Status and Language (12.56). Figure 13 displays the Z score for each SVI theme and shows that the Z score for each of the four themes exceeds the maximum Z score of a shadow feature.

For the SVI, PCA showed that the first principal component explained 47.1% of the variation. Thus, the first principal component was the most important in this study because the associated samples showed the largest variation. Within the first principal component, the highest coefficient was for Socioeconomic Status (0.660), followed by Housing Type and Transportation (0.594), Household Composition and Disability (0.395), and Minority Status and Language (0.237).

Table IX compares the county-level SVI results obtained using the three different approaches. At the county level, PCA, relative importance metrics for linear regression, and Random Forest all produced the same relative importance ranking for explaining the composite SVI score for Illinois.

TABLE VI. RELATIVE IMPORTANCE OF SVI THEMES TO COMPOSITE SCORE USING LINEAR REGRESSION

Factor	Accuracy
Socioecon	0.417
Disability	0.148
Minority	0.097
HousingType	0.320

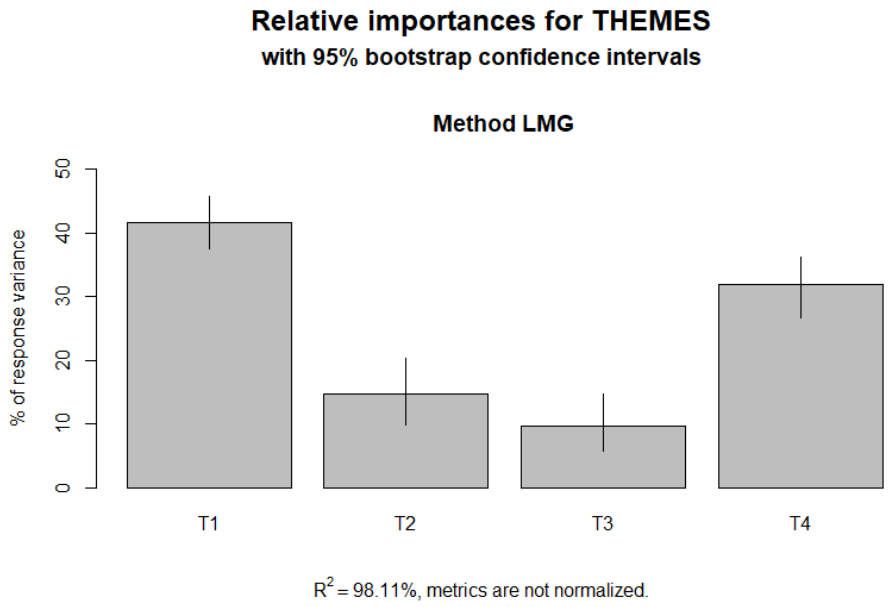

Figure 10. Relative importance of SVI themes to composite score using linear regression

TABLE VII. RELATIVE IMPORTANCE OF SVI THEMES TO COMPOSITE SCORE USING RANDOM FOREST

Factor	Mean	Decision
Socioecon	31.201	Confirmed
Disability	16.259	Confirmed
Minority	12.559	Confirmed
HousingType	23.863	Confirmed

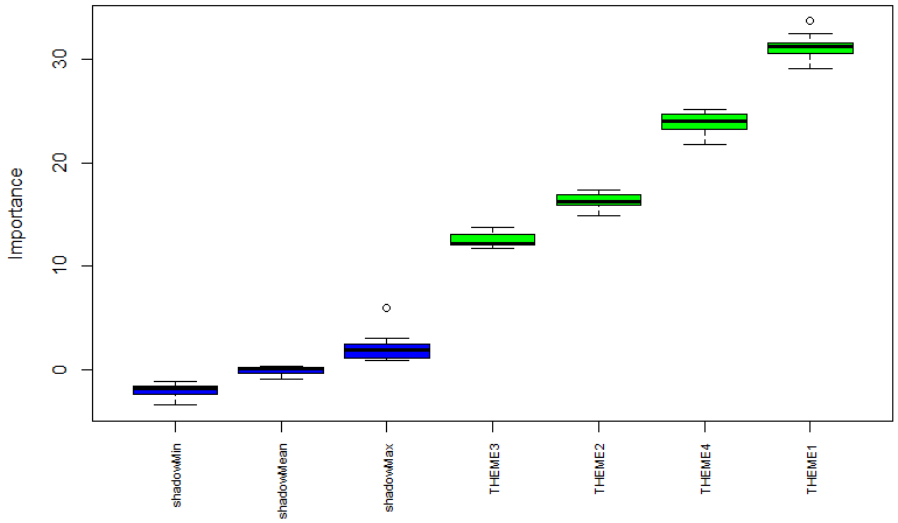

Figure 11. Relative importance of SVI themes to composite score using Random Forest

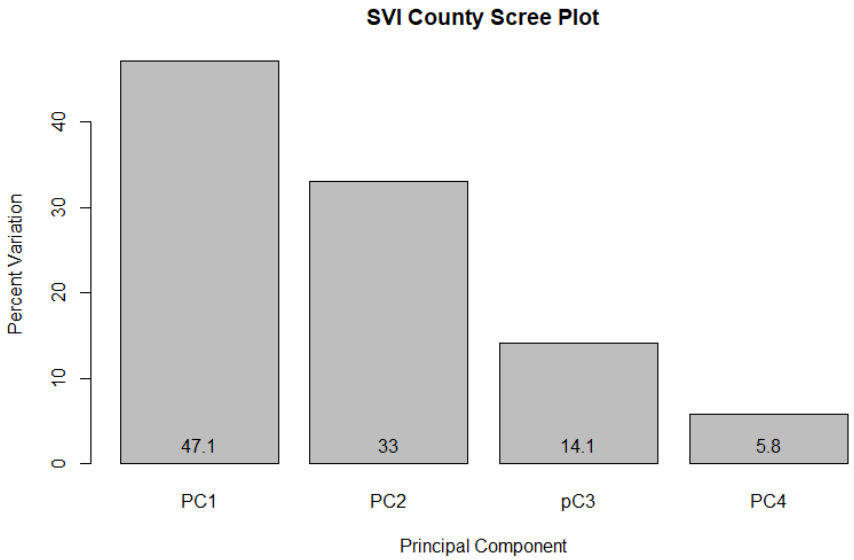

Figure 12. Proportion of variance for principal components: SVI

TABLE VIII. RELATIVE IMPORTANCE OF SVI THEMES TO COMPOSITE SCORE USING PCA

Factor	Coefficient
Socioecon	0.660
Disability	0.395
Minority	0.237
HousingType	0.594

However, relative importance in linear regression analysis did not show a significant difference between Household Composition and Disability theme and Minority Status and Language theme. In Table IX, Socioeconomic Status contributed most to the composite score, and Minority Status and Language contributed least. Based on these results, I concluded that the Minority Status and Language theme was strongly related with the infection rate. Even though the Minority Status and Language theme was the least important to the composite score, it had a significant relationship with the infection rate and also appeared to contribute to a significant positive relationship between the SVI composite score and the infection rate.

TABLE IX. COMPARISON OF RELATIVE IMPORTANCE OF SVI THEMES TO COMPOSITE SCORE AMONG THREE APPROACHES

Methods	Rank			
PCA	T1	T4	T2	T3
Linear regression	T1	T4	T2	T3
Random Forest	T1	T4	T2	T3

C. Relative Contributions of the Six Themes of CCVI to the Composite Score

First, Table X shows the correlation of themes of CCVI at the county level and shows that those six themes are not highly correlated to each other.

TABLE X. THE CORRELATIONS OF CCVI THEMES

Themes	Socioeconomic Status	Household Composition and Disability	Minority Status and Language	Housing Type and Transportation	Epidemiological Factors	Healthcare System Factors	CCVI
Socioeconomic Status	1						
Household Composition and Disability	0.6161	1					
Minority Status and Language	-0.1092	-0.4011	1				
Housing Type and Transportation	0.5894	0.3101	-0.0541	1			
Epidemiological Factors	0.1513	0.1132	-0.0868	0.1316	1		
Healthcare System Factors	0.0279	0.0592	-0.2873	0.0729	0.0085	1	
CCVI	0.6969	0.5482	-0.0709	0.6064	0.6524	0.3051	1

At the county level, the regression model confirmed the relative importance of the six themes to the overall CCVI score. The regression model showed that among the six CCVI themes for the state of Illinois, the largest contribution to CCVI was the Epidemiological Factors (R^2= 0.34), followed by Socioeconomic Status (0.221), Housing Type and Transportation (0.165), Household Composition and Disability (0.137), Healthcare System Factors (0.087), and

Minority Status and Language (0.02). The Figure 15 shows the R^2 for each theme. However, based on the confidence interval, only themes Epidemiological Factors and Minority Status and Language showed statistically significant difference with the other themes.

The Random Forest model confirmed the relative importance of the five themes to the overall CCVI score. The Minority Status and Language theme was dropped, as its mean z score was below the maximum z score of the shadow feature. At the county level, Epidemiological Factors showed the highest mean Z score (29.20), followed by Socioeconomic Status (20.76), Housing Type and Transportation (18.53), Household Composition and Disability (12.76), and Healthcare System Factors (11.30). The Figure 16 displays the mean Z scores for six CCVI themes, and it shows that the Z scores for five themes exceed the maximum Z score of a shadow feature.

Principal Component Analysis showed that the first principal component explained 36.5% of the variation. PCA assumes that the components with the largest variances are the most important. Thus, the first principal component was the most important in our study because the associated samples showed the largest variation. Within the first principal component, the highest coefficient was Socioeconomic Status (0.571) followed by Household Composition and Disability (0.547), Housing Type and Transportation (0.469), Epidemiological Factors (0.2), Healthcare System Factors (0.144), and Minority Status and Language (0.307).

Table XIV compares the county level CCVI results obtained using the three different approaches. Relative importance in linear regression and Random Forest produced the same relative importance ranking for explaining the overall CCVI score for Illinois at the county level, even though relative importance in linear regression analysis revealed that only theme 5 and 3 show significant difference than the other themes.

TABLE XI. RELATIVE IMPORTANCE OF CCVI THEMES TO COMPOSITE SCORE USING LINEAR REGRESSION

Factor	Accuracy
Socioecon	0.221
Disability	0.137
Minority	0.020
HousingType	0.165
Epi	0.340
Healthcare	0.087

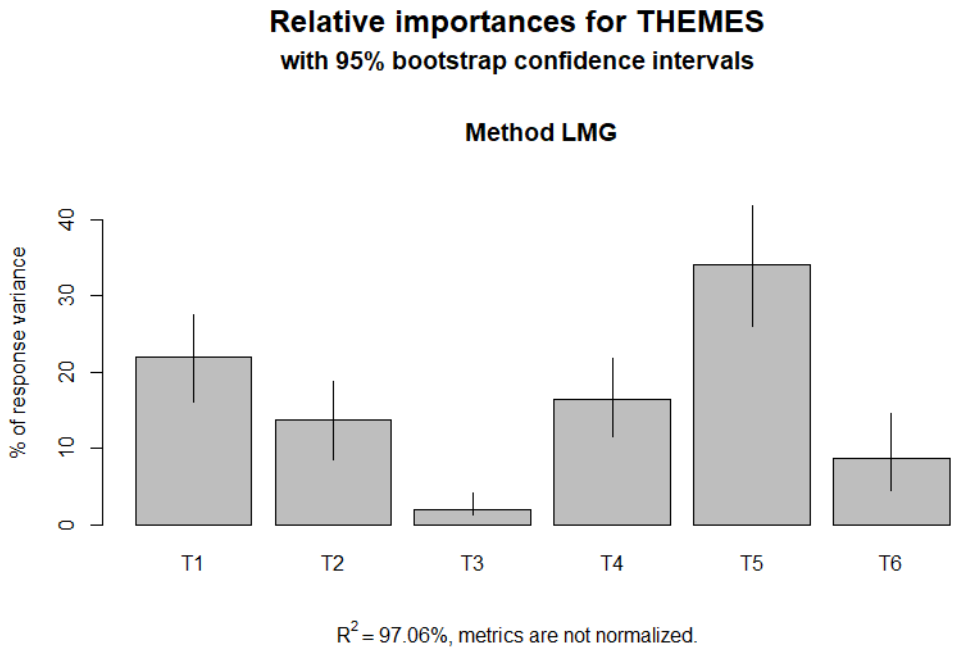

Figure 13. Relative importance of CCVI themes to composite score using linear regression

TABLE XII. RELATIVE IMPORTANCE OF CCVI THEMES TO COMPOSITE SCORE USING RANDOM FOREST

Factor	Mean	Decision
Socioecon	20.756	Confirmed
Disability	12.755	Confirmed
Minority	-0.409	Rejected
HousingType	18.530	Confirmed
Epi	29.202	Confirmed
Healthcare	11.290	Confirmed

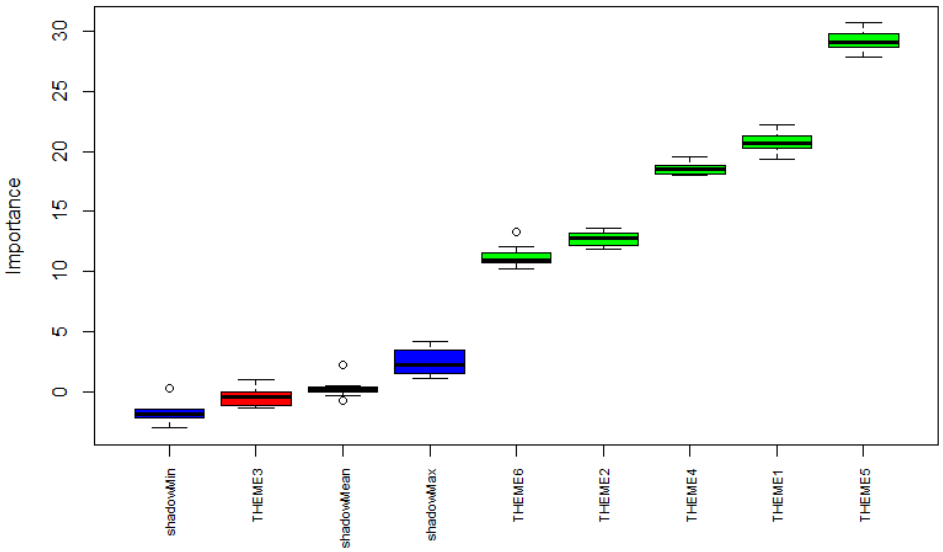

Figure 14. Relative importance of CCVI themes to composite score using Random Forest

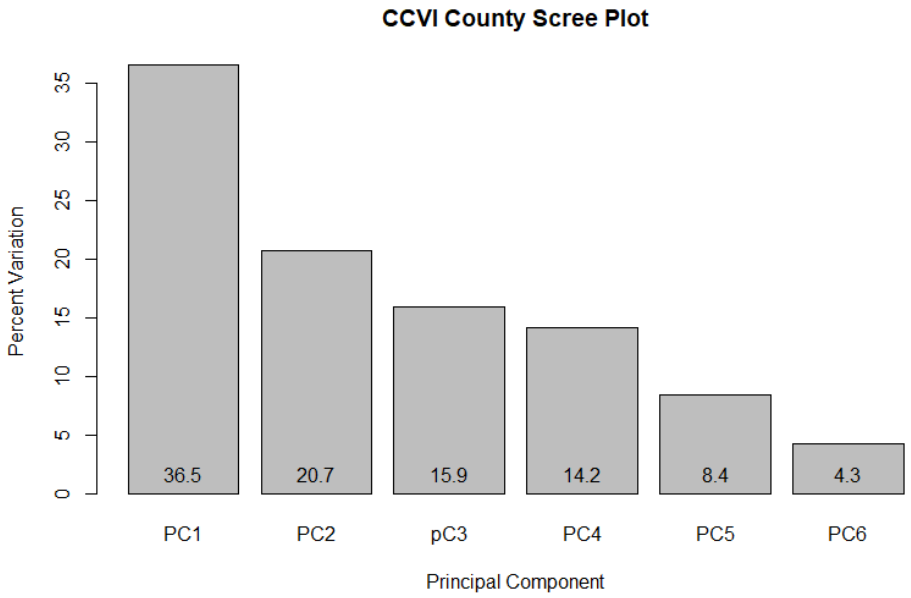

Figure 15. Proportion of variance for principal components: CCVI

TABLE XIII. RELATIVE IMPORTANCE OF CCVI THEMES TO COMPOSITE SCORE USING PCA

Factor	Coefficient
Socioecon	0.571
Disability	0.547
Minority	-0.307
HousingType	0.469
Epi	0.200
Healthcare	0.144

In Table XIV, these two methods reveal that Epidemiological Factors theme contributes the most and Minority Status and Language theme contributes the least to the composite score, whereas PCA generated a very different ranking than linear regression and Random Forest, showing that Socioeconomic Status contributes the greatest. However, PCA showed that Minority Status and Language theme contributes the least, consistent with the results of the relative importance in linear regression and random forest models. From these results, it was found that Minority Status and Language theme of CCVI is also very strongly related with the infection rate. While Minority Status and Language theme is the least important to the composite score, only Minority Status and Language theme shows the significant relationship with the infection rate and also seem to lead a significant positive relationship between CCVI and the rate. The PCA considers only the variation within each theme but does not consider how each theme is related with the outcome variable. Therefore, PCA would generate very different results. In addition, the correlations of each theme were not strong, thus PCA is the least attractive method used.

TABLE XIV. COMPARISON OF RELATIVE IMPORTANCE OF CCVI THEMES TO COMPOSITE SCORE AMONG THREE APPROACHES

Method	Rank					
PCA	T1	T2	T4	T5	T6	T3
Linear regression	T5	T1	T4	T2	T6	T3
Random Forest	T5	T1	T4	T2	T6	T3

D. Relative Contributions of the Four Themes of SVI to the Case Fatality Rate

Three different methods were applied to measure the contribution of each theme to the case fatality rate. The regression results in Table III and IV did not reveal much about which themes would contribute the fatality rate the most, thus I decided to applied feature importance method in Section A and B to the case fatality rate. I also examined the ability of SVI and CCVI to explain the infection rate using the three approaches and the results show that the Minority Status and Language theme of both indexes was the most important factor in explaining the infection rate (Appendix J and K).

In Table XV, the relative importance in linear regression model showed that among the four SVI themes for Illinois, Household Composition and Disability had the largest R^2 (0.036), followed by Socioeconomic Status (0.028), Minority Status and Language (0.003), and Housing Type and Transportation (0.003). However, all themes did not show any statistically significant importance than the other themes, which is similar results with the regression results in Table III.

TABLE XV. RELATIVE IMPORTANCE OF SVI THEMES TO CASE FATALITY RATES USING LINEAR REGRESSION

Factor	Accuracy
Socioecon	0.028
Disability	0.036
Minority	0.003
HousingType	0.003

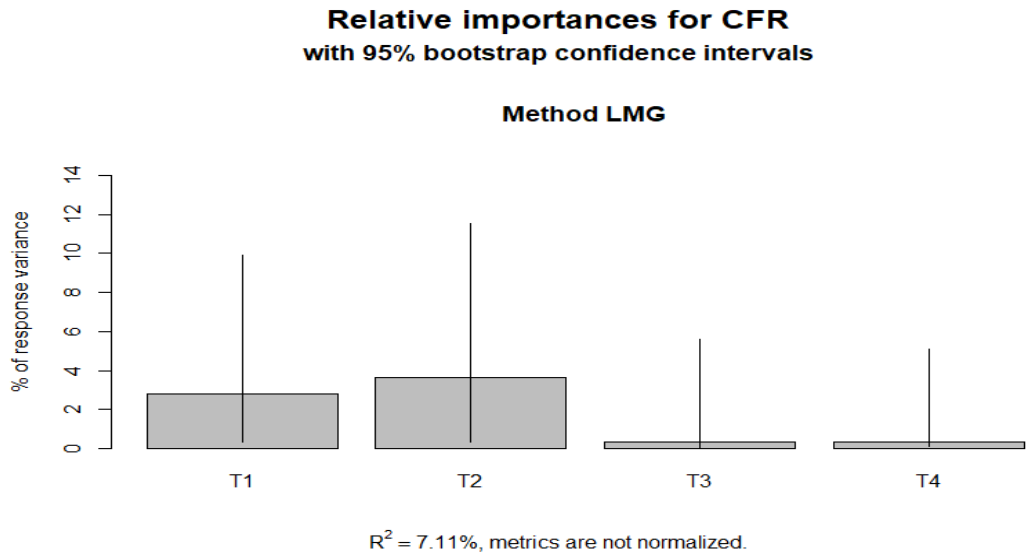

Figure 16. Relative importance of SVI themes to case fatality rates using linear regression

The Random Forest model showed that at the county level, Socioeconomic Status had the highest mean Z score (7.106), followed by Household Composition and Disability (5.988), and Housing Type and Transportation (2.207). Minority Status and Language was rejected (0.712). The Figure 19 displays the Z score for each SVI theme and shows that the Z score for two themes exceeds the maximum Z score of a shadow feature.

Principal Component Analysis produced the same results as the Table VIII and figure 12. The Table XVII compares the county level SVI results obtained using the three different approaches. At the county level, PCA, relative importance metrics in linear regression, and Random Forest produced the different relative importance ranking for explaining the case fatality rate for Illinois. The use of relative importance metrics in linear regression showed the similar pattern shown in the regression results above (Table II).

TABLE XVI. RELATIVE IMPORTANCE OF SVI THEMES TO CASE FATALITY RATES USING RANDOM FOREST

Factor	Mean	Decision
Socioecon	7.106	Confirmed
Disability	5.988	Confirmed
Minority	0.712	Rejected
HousingType	2.207	Confirmed

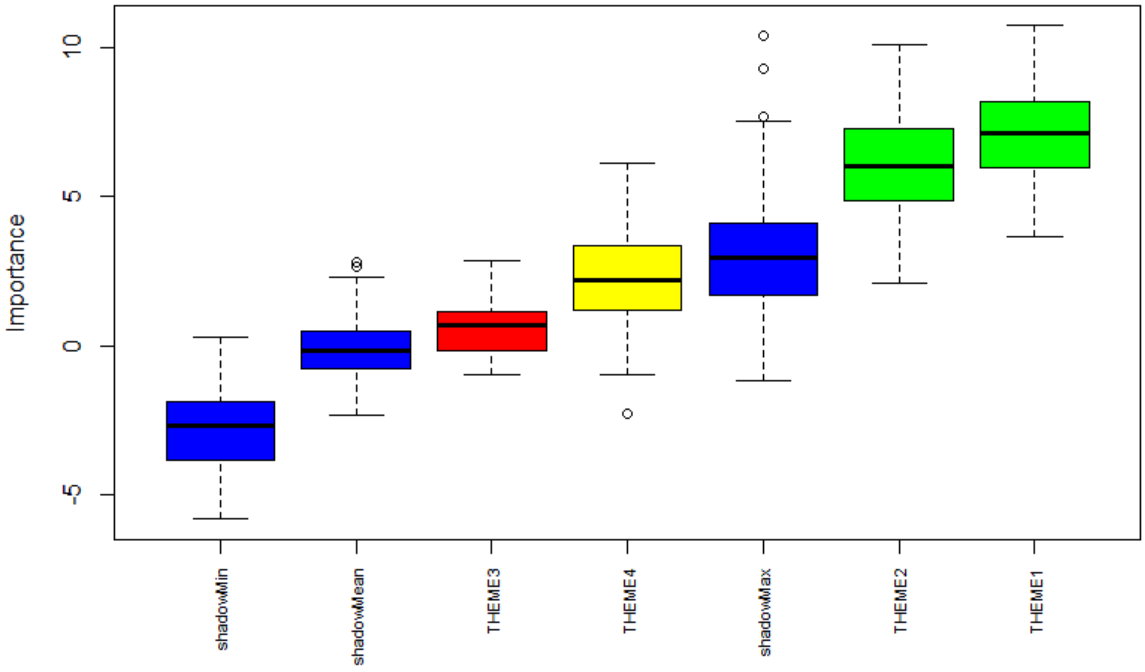

Figure 17. Relative importance of SVI themes to case fatality rates using Random Forest

The Household Composition and Disability theme contributes the most case fatality rates and the Socioeconomic Status theme the second most case fatality rates. However, all themes show no statistically significant difference than the other themes. The Random Forest model also showed that the Socioeconomic Status theme and the Household Composition and Disability theme are relatively important to the rates. The PCA showed that Socioeconomic Status is the most important factor to the rate but showed different relative importance of themes as a whole. In summary, comparing three methods of analysis, Socioeconomic Status theme is found to be statistically the most important to the infection fatality rates.

TABLE XVII. COMPARISON OF RELATIVE IMPORTANCE OF SVI THEMES TO CASE FATALITY RATES AMONG THREE APPROACHES

Methods	Rank			
PCA	T1	T4	T2	T3
Linear regression	T2	T1	T3	T4
Random Forest	T1	T2	T4	T3

E. Relative Contributions of the Six Themes of CCVI to Case Fatality Rate

At the county level, the regression model confirmed the relative importance of the six themes to the case fatality score. The regression model showed that of the six CCVI themes for Illinois, the largest contribution to CCVI was the Housing Type and Transportation theme (R^2= 0.048), followed Household Composition and Disability by (0.018), Healthcare System Factors

(0.018), Socioeconomic Status (0.005), Epidemiological Factors (0.005), and Minority Status and Language (0.002). The Figure 20 shows the R^2 for each theme. However, all themes show no statistically significant importance than the other themes, which is similar results with the regression results in Table IV.

Similarly, the Random Forest model confirmed the relative importance of the six themes to the case fatality rate. This model showed that at the county level, Epidemiological Factors has the highest mean Z score (6.508), followed by Housing Type and Transportation (2.890), Healthcare System Factors (2.553). The Socioeconomic Status, Household Composition and Disability, and Minority Status and Language were rejected. The Figure 21 displays the Z score for each CCVI theme and shows that the Z score of two themes exceeds the maximum Z score of a shadow feature.

Principal Component Analysis produce the same results shown in the Table XIII and figure 15. The Table XX compares the CCVI results obtained using the three different approaches. At the county level, PCA, relative importance in linear regression, and Random Forest produced very different relative importance ranking for explaining the case fatality rate for Illinois. The use of relative importance metrics in linear regression also showed the very different results shown in the regression results above (Table III) even though all themes show no significant difference between themes. In other words, three different methods produced non-consistent results when I use CCVI.

TABLE XVIII. RELATIVE IMPORTANCE OF CCVI THEMES TO CASE FATALITY RATES USING LINEAR REGRESSION

Factor	Accuracy
Socioecon	0.005
Disability	0.018
Minority	0.002
HousingType	0.048
Epi	0.005
Healthcare	0.018

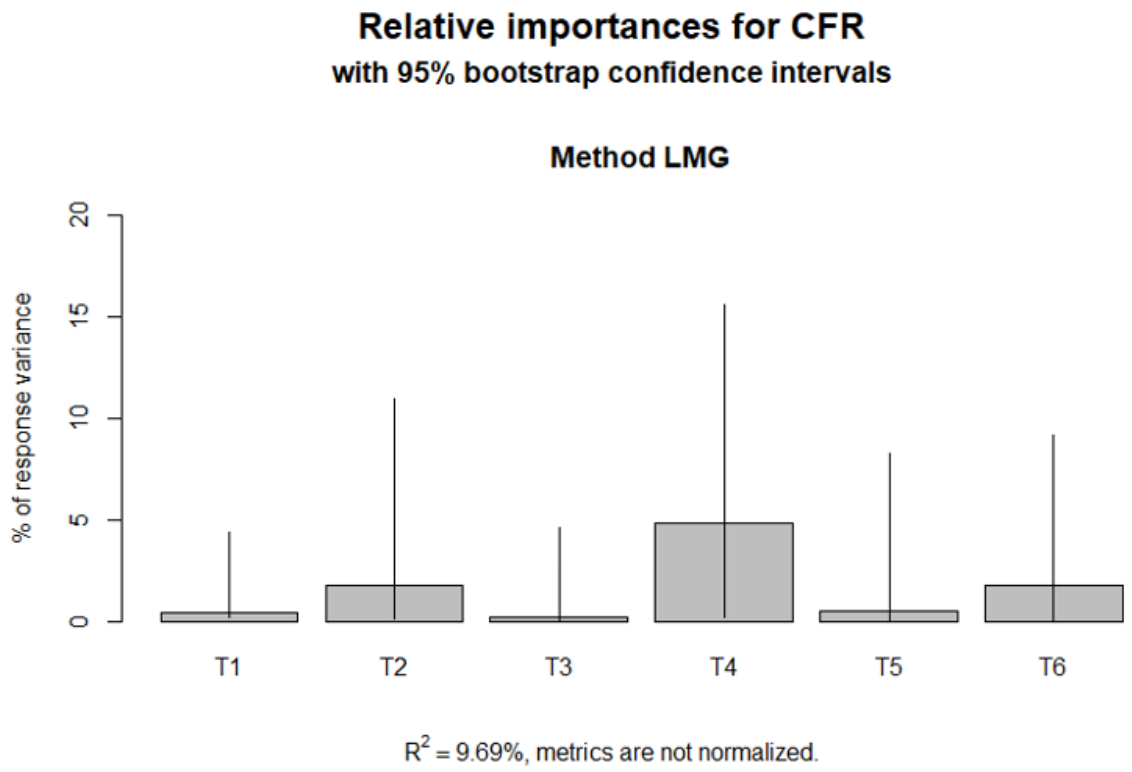

Figure 18. Relative importance of CCVI themes to case fatality rates using linear regression

TABLE XIX. RELATIVE IMPORTANCE OF CCVI THEMES TO CASE FATALITY RATES USING RANDOM FOREST

Factor	Mean	Decision
Socioecon	1.521	Rejected
Disability	-0.861	Rejected
Minority	1.213	Rejected
HousingType	2.890	Confirmed
Epi	6.508	Confirmed
Healthcare	2.553	Confirmed

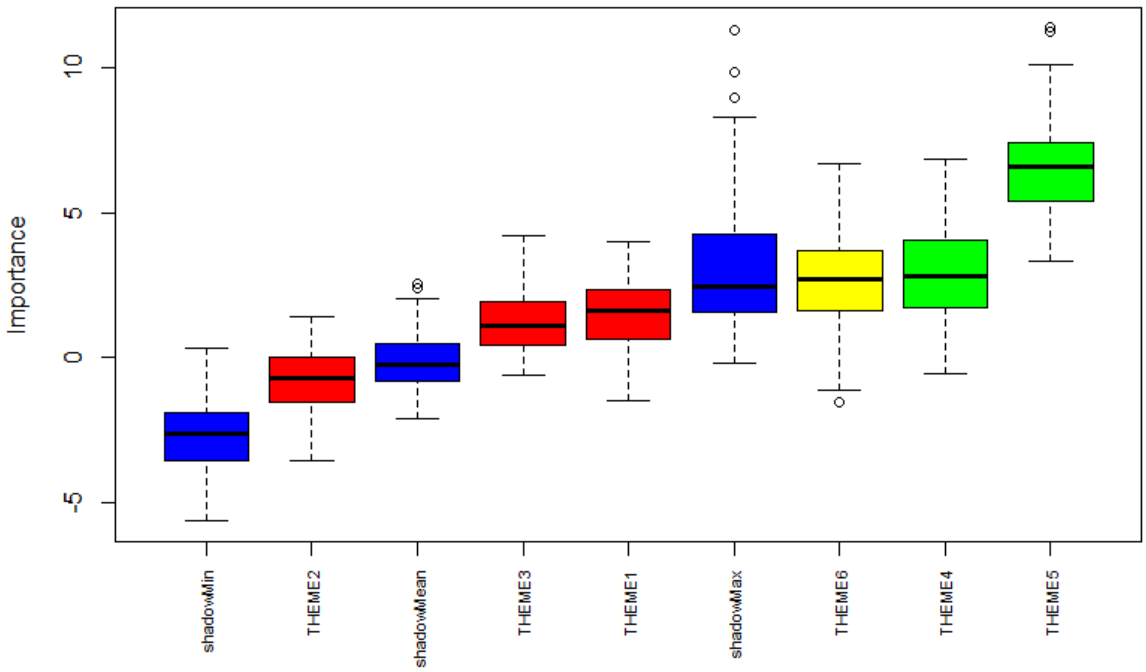

Figure 19. Relative importance of CCVI themes to case fatality rates using Random Forest

TABLE XX. COMPARISON OF RELATIVE IMPORTANCE OF CCVI THEMES TO CASE FATALITY RATES AMONG THREE APPROACHES

Method	Rank					
PCA	T1	T2	T4	T5	T6	T3
Linear regression	T4	T2	T6	T1	T5	T3
Random Forest	T5	T4	T6	T1	T3	T2

IV. DISCUSSION

Illinois' most populous city, Chicago, is one of the most segregated in the United States. Therefore, COVID-19 would be expected to have an unequal impact on infection and fatality rates in minority neighborhoods based on their social conditions and risk exposure. This is the case because social conditions such as poverty, congested housing, and potentially limited language ability would affect such communities' risk of exposure to infectious disease and their ability to minimize this risk.

The SVI is a valuable tool for understanding disproportionate outcomes resulting from a natural disaster such as the pandemic among communities. In addition, the CCVI was specifically developed by Surgo Foundation in response to the current COVID-19 outbreak. Therefore, this study examined how well the SVI and CCVI explain COVID-19 infection and case fatality rates and compared their themes to determine which most contributed to the pandemic-related outcomes.

This study found that the SVI explains COVID-19 infection and fatality rates better than the CCVI using quartile and decile grids. Poisson regression results revealed that only the SVI was correlated with infection and fatality rates, whereas the CCVI was not. The Minority Status and Language theme of both the SVI and CCVI appeared to be related to the infection rate, and the Housing Type and Transportation theme of the SVI was related to the infection rate. In addition, the Healthcare System Factors theme of the CCVI was negatively related to the case fatality rate. Minority ethnic groups and people who live in congested housing or commute to work using public transportation would be at greater risk of exposure to COVID-19. Furthermore, the Healthcare System Factors theme of the CCVI showed a negative relationship

with the case fatality rate. In other words, people infected with COVID-19 and living in counties with poor healthcare would tend to travel to counties with good hospitals to obtain better quality of care, and thus the latter would have a higher case fatality rate.

In addition, using the PCA, Random Forest and relative importance metrics in linear regression analysis to rank the relative importance of themes, the study found that the Socioeconomic Status theme of the SVI made the greatest contribution and the Minority Status and Language theme made the least contribution to the composite score at the county level. For the CCVI, relative importance in linear regression and Random Forest analyses indicated that the Epidemiological Factors theme most contributed to the composite score at the county level, whereas PCA produced different rankings of themes. This is understandable, as PCA does not consider how each theme is related to the outcome variable; therefore, PCA would be expected to generate very different results. Furthermore, the correlations between CCVI themes were weak. Thus, PCA was found to be the least suitable method in this study.

In the SVI regression model (Table III), this was one of the themes that showed a significant relationship with the infection rate, and it was the only statistically significant theme in the CCVI regression model (Table IV). However, all three methods indicated that the Minority Status and Language theme was the least important to the composite score of the CCVI. Thus, the Minority Status and Language theme of both the SVI and CCVI contributed least to their composite scores. Therefore, I conclude that this would explain why the CCVI composite score had no significant relationship with the infection rates. The SVI's Minority Status and Language theme was an important factor in explaining the infection rate, and this theme was the least important to the SVI composite score. Unlike the CCVI, this theme contributed to a significant

positive relationship between the composite score of the SVI and the infection rate. Thus, CCVI would not be appropriate for explaining overall COVID-19 infection rates.

I used the three methods to measure the relative importance of each SVI and CCVI theme to the case fatality rate. The regression results (Tables III and IV) did not show any significant relationship between the case fatality rate and each theme, so it was not clear which theme had the greatest relative importance to this rate. For this reason, I examined ability of SVI and CCVI to explain the case fatality rate using the three approaches. For the SVI, the Random Forest model and PCA indicated that the Socioeconomic Status theme was most important to the case fatality rate. However, use of the relative importance method for the linear regression model indicated that the Household Composition and Disability theme contributed most to the rate, followed by the Socioeconomic Status theme. This was the same pattern as was found in the regression results in Table III and use of the relative importance method showed no statistically significant difference among the themes. The case fatality rate is more directly related to Socioeconomic Status factors such as income, while the infection rate is more directly related to the Minority Status and Language theme. For the CCVI, the three methods produced very different results that were also inconsistent with the regression results for the CCVI themes (Table III). Thus, I conclude that the CCVI is not appropriate for explaining overall COVID-19 infection and fatality rates or to estimate the relative importance of each theme to these rates.

With respect to COVID-19, one of the issues with the CCVI is that the variables used for the Epidemiology and Healthcare System Factors themes are measured at the state or county level, which limits the utilization of the CCVI at the census tract level. I argue that Indexes such as the CCVI need to be constructed at the community and census tract levels. The study findings suggest a need for better data collection at lower levels and for smaller units of analysis to

support preparation for pandemics such as COVID-19. The Chicago Department of Public Health has developed the Chicago COVID-19 Community Vulnerability Index using local-level data. This new index is a modification of the CCVI and SVI (Chicago Department of Public Health, 2021). Measurement at the local level is expected to improve the ability to identify communities that have been disproportionately impacted by COVID-19.

CITED LITERATURE

Amram, O., Amiri, S., Lutz, R.B., Rajan, B., and Monsivais, P.: Development of a vulnerability index for diagnosis with the novel coronavirus, COVID-19, in Washington State, USA. Health Place. 64,102377, 2020.

Bilal, U., Barber, S., Tabb, L., and Diez-Roux, A. V.: Spatial Inequities in COVID-19 Testing, Positivity, Incidence and Mortality in 3 US Cities: a Longitudinal Ecological Study. 2020. medRxiv. url: https://www.medrxiv.org/content/10.1101/2020.05.01.20087833v3

Breiman, L.: Random Forest. a technical report for the University of California Berkeley Statistics Department. 1999. url: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Chicago Department of Public Health.: Latest Data. 2020. url: https://www.chicago.gov/city/en/sites/covid-19/home/latest-data.html

Chicago Department of Public Health.: Chicago COVID-19 Community Vulnerability Index. 2021. url: https://www.chicago.gov/content/dam/city/sites/covid/reports/012521/Community_Vulnerability_Index_012521.pdf

Daras, K., Alexiou, A., Rose, T., Buchan, I., Robinson, D., and Bar, B.: How does Vulnerability to COVID-19 Vary between Communities in England? Developing a Small Area Vulnerability Index (SAVI). J. Epidemiol. Community. Health. Published Online First: 04 February 2021.

Dasgupta, S., Bowen, V.B., Leidner, A., Fletcher, K., Musial, T., … and Oster, A. M.: Association Between Social Vulnerability and a County's Risk for Becoming a COVID-19 Hotspot - United States, June 1-July 25. MMWR. Morb. Mortal. Wkly. Rep. 69(42):1535-1541, 2020.

Gaynor, T.S. and Wilson, M.E.: Social Vulnerability and Equity: The Disproportionate Impact of COVID-19. Public. Admin. Rev. 80(5): 832-838, 2020.

Grömping, U.: Variable Importance Assessment in Regression: Linear Regression versus Random Forest. Am. Stat. 63(4):308-319, 2009.

Hathaway, E.D.: American Indian and Alaska Native People: Social Vulnerability and COVID-19. J. Rural. Health. 37(1):256-259, 2021.

Holland, S.M.: Principal Components Analysis (PCA). University of Georgia. 2020

Illinois Department of Public Health.: Coronavirus Disease 2019 (COVID-19). Illinois Department of Public Health. 2020. url: https://www.dph.illinois.gov/covid19

Judson, S. D., Njabo, K. Y., and Torimiro, J. N.: Regional vulnerability for COVID-19 in Cameroon. Pan. Afr. Med. J. 37(Suppl 1):16, 2020.

Karaye, I. M., and Horney, J. A.: The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships. Am. J. Prev. Med. 59(3):317-325, 2020.

Khazanchi, R., Beiter, E.R., Gondi, S., Beckman, A. L., Bilinski, A., and Ganguli, I.: County-Level Association of Social Vulnerability with COVID-19 Cases and Deaths in the USA. J. Gen. Intern. Med. 35:2784–2787, 2020.

Kim, S., and Bostwick, W.: Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago. Health. Educ. Behav. 47(4):509-513, 2020.

LeRose, J., Merlo, C., Duong, P., Harden, K., Rush, R., Artzberger, A., . . . and Chopra, T.: The role of the social vulnerability index in personal protective equipment shortages, number of cases, and associated mortality during the coronavirus disease 2019 (COVID-19) pandemic in Michigan skilled nursing facilities. Infect. Control. Hosp. Epidemiol. 1-4, 2020.

Macharia, P.M., Joseph, N.K., and Okiro, E.A.A.: Vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya. BMJ Global Health. 5(8), 2020

Marvel, S., House, J., Wheeler, M., Song, K., Zhou, Y., … and Reif, D.: The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: monitoring county level vulnerability. Environ. Health. Perspect. 129(1), 2020

Mishra, S. V., Gayen, A., and Haque. S. M.: COVID-19 and urban vulnerability in India. Habitat. Int. 103, 2020.

Nayak, A., Islam, S. J., Mehta, A., Ko, Y.A., Patel, S. A., Goyal, A., Sullivan, S., Lewis, T. T., Vaccarino, V., Morris, A. A., and Quyyumi, A. A.: Impact of Social Vulnerability on COVID-19 Incidence and Outcomes in the United States. 2020. medRxiv. url: https://www.medrxiv.org/content/10.1101/2020.04.10.20060962v2

New York Times.: Coronavirus in the U.S.: Latest Map and Case Count. The New York Times. 2020. url: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

Sarkar, A., and Chouhan, P.: COVID-19: District level vulnerability assessment in India. Clin. Epidemiol. Glob. Health. 9: 204-215, 2021.

Snyder, B. F., and Parks, V.: Spatial variation in socio-ecological vulnerability to Covid-19 in the contiguous United States. Health. Place. 66, 2020.

Surgo Foundation.: Community Vulnerability Index. Surgo Foundation. 2020. url: https://precisionforcovid.org/ccvi

Tiwari, A., Dadhania, A. V., Ragunathrao, V.A.B., and Oliveira, E. R. A.: Using Machine Learning to Develop a Novel COVID-19 Vulnerability Index (C19VI). 2020. arxiv.org. url: https://arxiv.org/abs/2009.10808

Wang, C., Li, Z., Mathews, M. C., Praharaj, S., Karna, B., and Solís, P.: The spatial association of social vulnerability with COVID-19 prevalence in the contiguous United States. Int. J. Environ. Health. Res. 2020.

Wong, M. C., Teoh, J. Y., Huang, J., & Wong, S. H.: The potential impact of vulnerability and coping capacity on the pandemic control of COVID-19. J. Infect. 81(5): 816–846, 2020.

World Health Organization.: Estimating mortality from COVID-19. 2020. url: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci-Brief-Mortality-2020.1

Wiemers, E. E., Abrahams, S., AlFakhri, M., Hotz, V. J., Schoeni, R. F., Seltzer, J.A.: Disparities in vulnerability to complications from COVID-19 arising from disparities in preexisting conditions in the United States. Res. Soc. Stratif. Mobil. 69, 2020.

Zenon, G.: New Interpretation of Principal Components Analysis. Zeszyty Naukowe WWSI. 11(16): 43-65, 2017.

US Census Bureau.: Quick Facts: United States. US Census Bureau. 2020. url: https://www.census.gov/quickfacts/fact/table/US/PST045219

APPENDICES

APPENDIX A

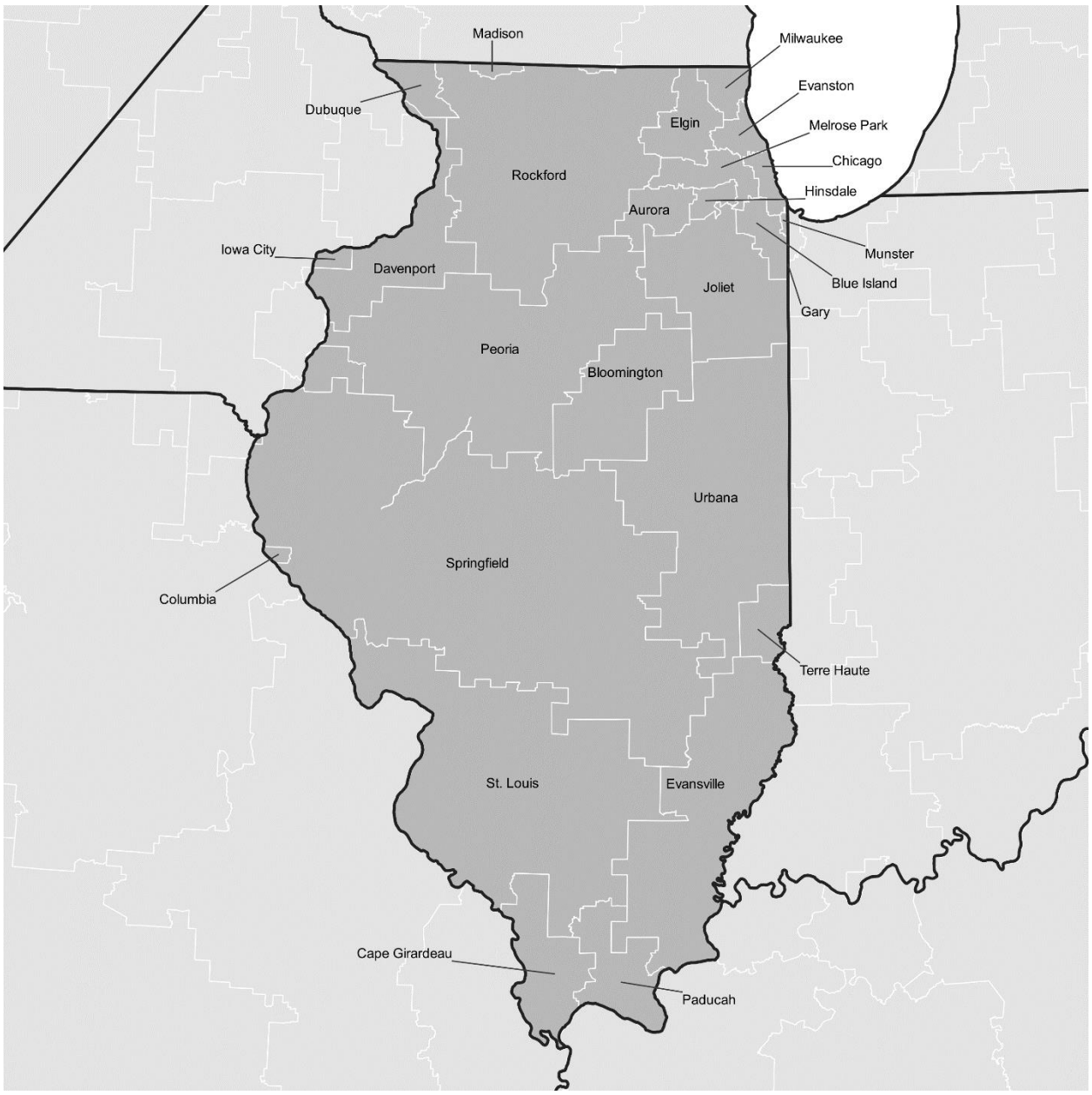

Figure 20. Hospital Referral Regions in Illinois

APPENDIX B

TABLE XXI. COUNTIES WHERE SVI/CCVI INDEXES BETTER-PREDICTED INFECTION RATES

County	**SVI** **N = 38**	**CCVI** **N = 25**	**IR**	**CCVI**	**SVI**
Boone		CCVI	5.49%	0.725	0.6139
Calhoun	SVI		3.06%	0.639	0.297
Cass	SVI	CCVI	5.61%	0.778	0.8812
Champaign	SVI		4.16%	0.523	0.5347
Clark	SVI		3.75%	0.757	0.3267
Cook	SVI	CCVI	4.71%	0.734	0.802
Crawford	SVI	CCVI	4.35%	0.571	0.5149
Cumberland		CCVI	4.34%	0.691	0.2871
De Witt	SVI		3.09%	0.432	0.1584
DeKalb		CCVI	3.68%	0.43	0.703
Fayette	SVI		5.32%	0.52	0.7921
Ford	SVI	CCVI	3.75%	0.498	0.4158
Grundy		CCVI	3.69%	0.431	0.1386
Hamilton	SVI		3.15%	0.518	0.2079
Henderson	SVI	CCVI	3.07%	0.41	0.099
Henry	SVI	CCVI	3.46%	0.291	0.2475
Iroquois	SVI	CCVI	4.11%	0.646	0.505
Johnson	SVI	CCVI	4.33%	0.7	0.5941
Kankakee	SVI		5.40%	0.484	0.9109
Knox		CCVI	4.00%	0.599	0.8911
Lake		CCVI	4.24%	0.615	0.4455
LaSalle	SVI	CCVI	4.04%	0.558	0.5545
Livingston	SVI		4.22%	0.77	0.6238
Macon	SVI		4.78%	0.524	0.7822
Macoupin	SVI		3.13%	0.532	0.2475
Madison	SVI		3.99%	0.379	0.4257
Marion		CCVI	4.61%	0.711	0.9505
Marshall	SVI		2.12%	0.555	0.0495
Mason	SVI		3.53%	0.736	0.4356
Menard	SVI	CCVI	2.44%	0.296	0.0396
Mercer	SVI		3.51%	0.227	0.1485
Morgan	SVI		4.63%	0.54	0.6733
Pulaski	SVI	CCVI	5.79%	0.977	0.8515
Putnam	SVI	CCVI	2.75%	0.203	0.0297
Randolph		CCVI	5.34%	0.783	0.5842

APPENDIX B (CONTINUED)

TABLE XXI. COUNTIES WHERE SVI/CCVI INDEXES BETTER-PREDICTED INFECTION RATES (CONTINUED)

County	**SVI** **N = 38**	**CCVI** **N = 25**	**IR**	**CCVI**	**SVI**
Richland		CCVI	3.12%	0.327	0.6535
Rock Island	SVI		4.52%	0.358	0.901
Sangamon	SVI		3.71%	0.269	0.604
Stark	SVI	CCVI	2.84%	0.275	0.0693
Stephenson	SVI		4.36%	0.485	0.6832
Tazewell	SVI	CCVI	3.38%	0.374	0.1089
Union	SVI		5.60%	0.667	0.9208
Warren	SVI		4.69%	0.592	0.802
Washington		CCVI	3.41%	0.485	0.0792
Wayne	SVI		4.37%	0.445	0.6436
Williamson	SVI		4.13%	0.34	0.6337
Winnebago	SVI		5.33%	0.585	0.9307
Woodford	SVI	CCVI	3.04%	0.412	0.0198
No. Counties where SVI and CCVI performed well: 15					

APPENDIX C

TABLE XXII. COUNTIES WHERE SVI/CCVI INDEXES BETTER-PREDICTED CASE FATALITY RATES

County	**SVI** **N=36**	**CCVI** **N=30**	**CFR**	**CCVI**	**SVI**
Adams	SVI		1.00%	0.238	0.3366
Bond	SVI		1.16%	0.277	0.4752
Bureau	SVI		1.36%	0.208	0.3663
Clark		CCVI	3.28%	0.757	0.9208
Clay	SVI	CCVI	3.09%	0.919	0.8317
Clinton		CCVI	1.79%	0.645	0.0495
Coles	SVI		1.94%	0.439	0.6634
Cook	SVI	CCVI	2.36%	0.734	0.802
DeKalb		CCVI	1.14%	0.43	0.2079
Edwards	SVI		1.31%	0.596	0.3762
Effingham		CCVI	0.31%	0.188	0.3366
Fayette	SVI		2.29%	0.52	0.7921
Grundy	SVI		0.53%	0.431	0.1386
Hamilton	SVI	CCVI	1.17%	0.518	0.2079
Hancock	SVI		0.71%	0.686	0.1683
Henderson	SVI		0.00%	0.41	0.099
Henry	SVI	CCVI	0.65%	0.291	0.2475
Iroquois	SVI	CCVI	1.80%	0.646	0.505
Jefferson	SVI		4.09%	0.616	0.9703
Kane	SVI		1.50%	0.281	0.4851
Kankakee		CCVI	1.47%	0.484	0.3861
Kendall	SVI	CCVI	0.67%	0.184	0.1782
Knox		CCVI	1.81%	0.599	0.9406
Lake		CCVI	1.83%	0.615	0.7327
Logan	SVI		0.46%	0.718	0.2772
Macon	SVI		1.71%	0.524	0.7822
Macoupin	SVI	CCVI	1.00%	0.532	0.2475
Madison	SVI		1.66%	0.379	0.4257
Marion		CCVI	1.86%	0.711	0.0099
Marshall		CCVI	1.65%	0.555	0.2772
Mason		CCVI	2.97%	0.736	0.9802
McDonough	SVI		2.04%	0.269	0.7327
McLean	SVI	CCVI	0.68%	0.429	0.2673
Menard	SVI	CCVI	0.34%	0.296	0.0396
Montgomery		CCVI	2.15%	0.704	0.901
Morgan	SVI	CCVI	1.86%	0.54	0.6733

APPENDIX C (CONTINUED)

TABLE XXII. COUNTIES WHERE SVI/CCVI INDEXES BETTER-PREDICTED CASE FATALITY RATES (CONTINUED)

County	**SVI** **N=36**	**CCVI** **N=30**	**CFR**	**CCVI**	**SVI**
Moultrie	SVI	CCVI	1.20%	0.538	0.3564
Perry	SVI	CCVI	2.27%	0.872	0.8416
Pike	SVI		1.71%	0.506	0.5248
Pope	SVI		1.69%	0.774	0.5644
Putnam	SVI	CCVI	0.00%	0.203	0.0297
Sangamon		CCVI	0.94%	0.269	0.1881
Scott	SVI		0.00%	0.567	0.198
St. Clair	SVI	CCVI	2.31%	0.662	0.7525
Stephenson		CCVI	0.83%	0.485	0.4653
Union	SVI		2.68%	0.667	0.9208
Wabash	SVI	CCVI	1.65%	0.501	0.495
Warren		CCVI	1.52%	0.592	0.7822
Washington	SVI		0.42%	0.485	0.0792
Winnebago		CCVI	1.36%	0.585	0.9703
White	SVI		1.79%	0.741	0.7426
No. Counties where SVI and CCVI performed well: 15					

APPENDIX D

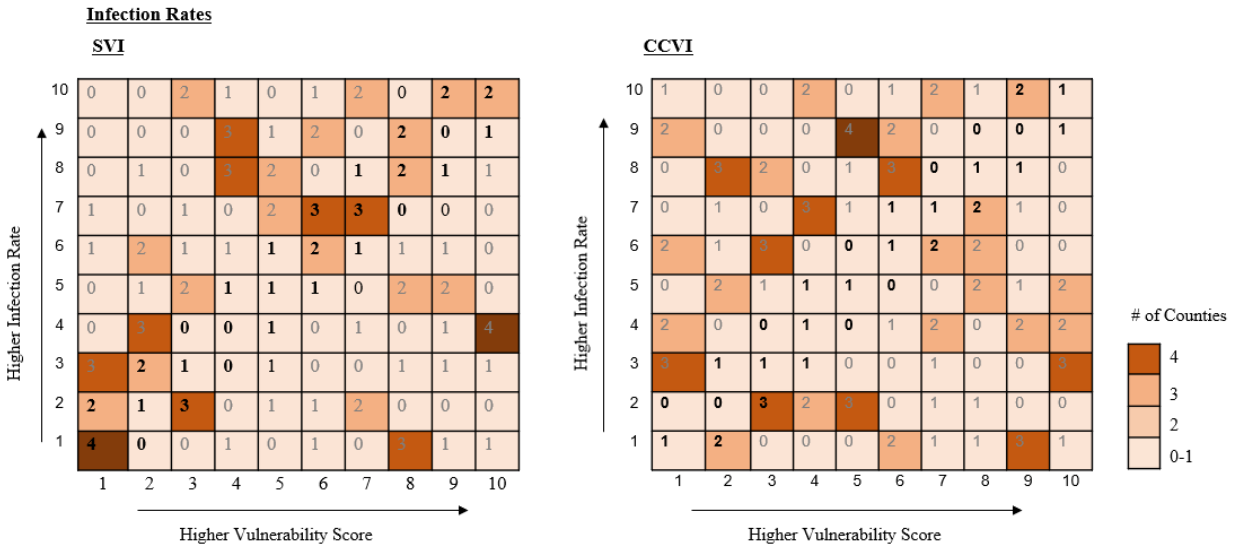

Figure 21. 10x10 Grid: Association of infection rates with SVI and CCVI

APPENDIX E

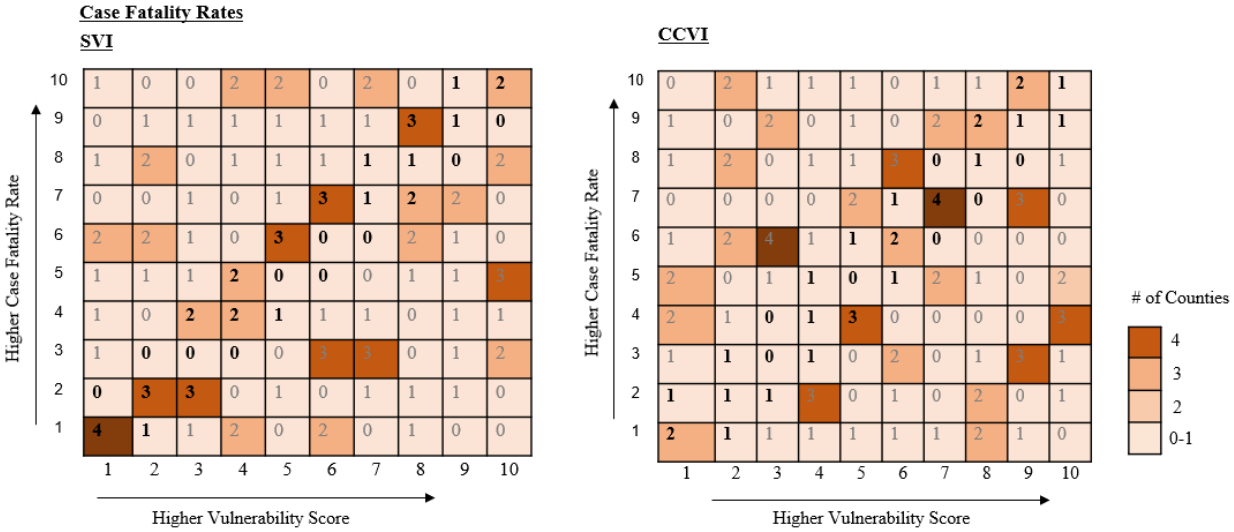

Figure 22. 10x10 Grid: Association of case fatality rates with SVI and CCVI

APPENDIX F

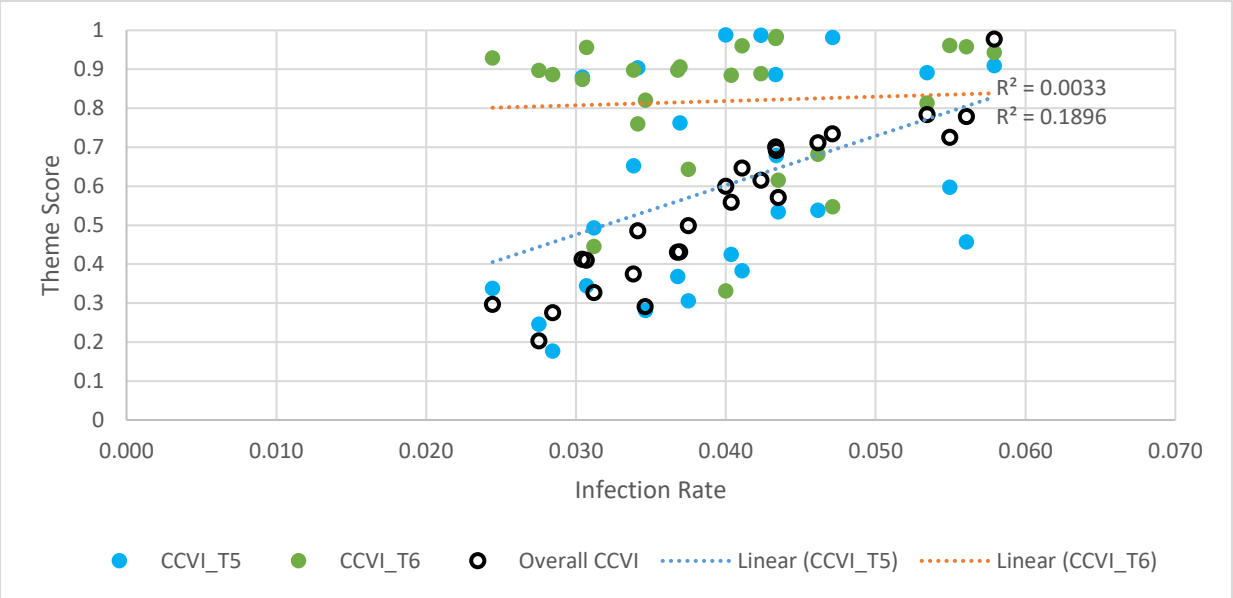

Figure 23. CCVI themes 5 and 6 and infection rates - selected counties

APPENDIX G

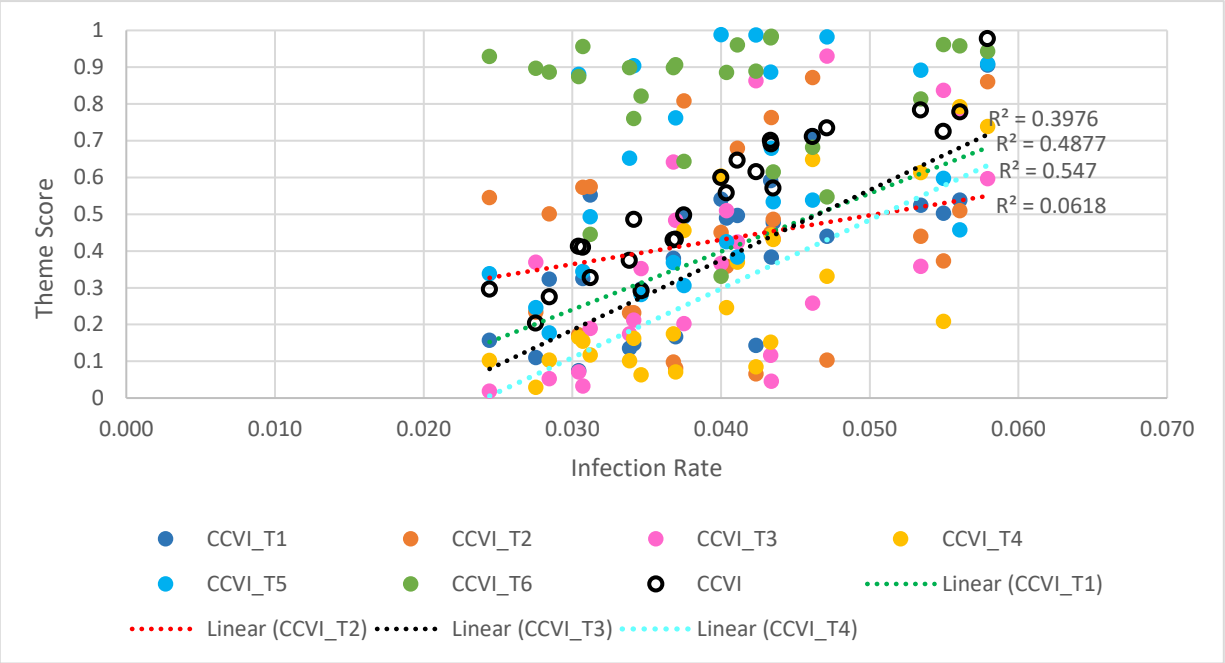

Figure 24. CCVI themes 1-4 and infection rates - selected counties

APPENDIX H

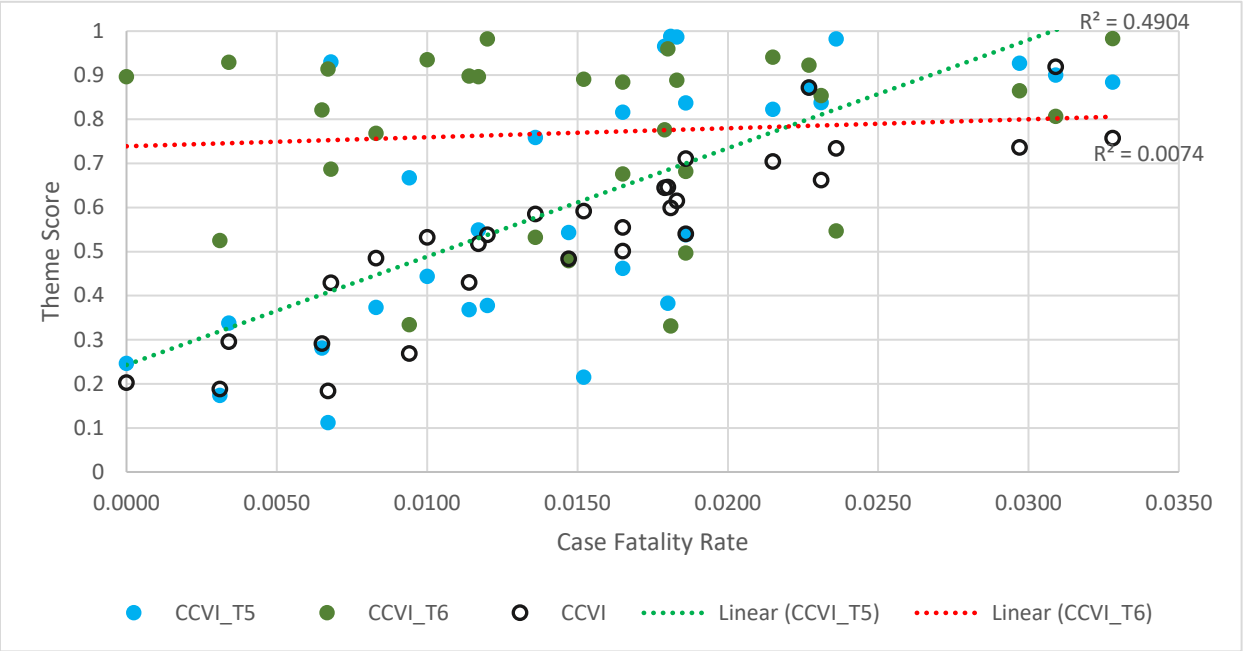

Figure 25. CCVI themes 5 and 6 and case fatality rates - selected counties

APPENDIX I

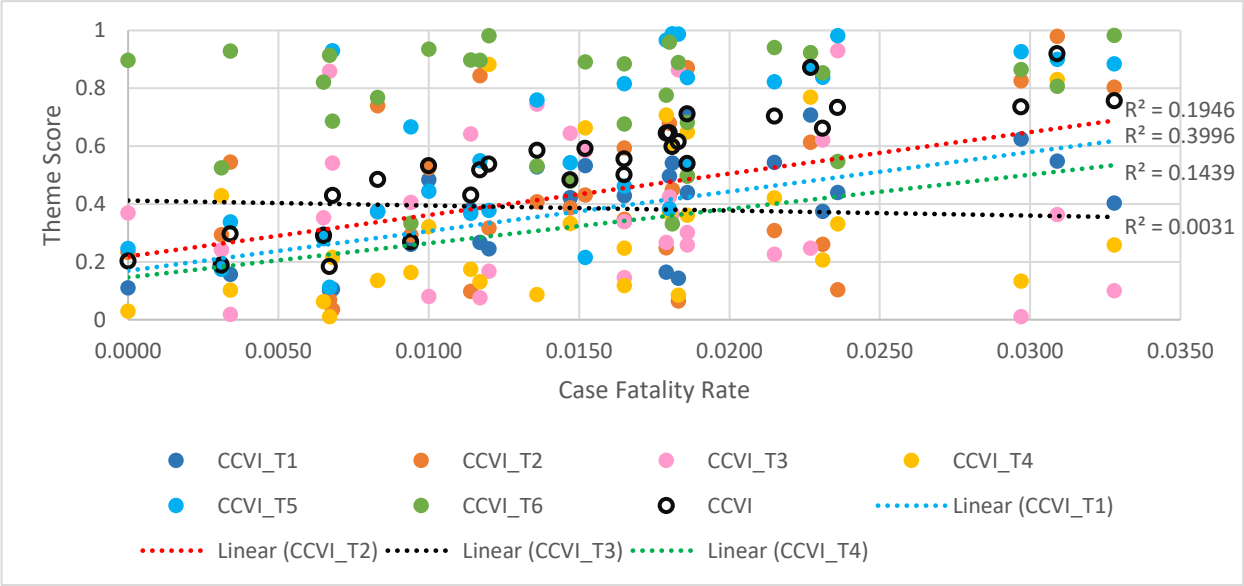

Figure 26. CCVI themes 1-4 and case fatality rates - selected counties

APPENDIX J

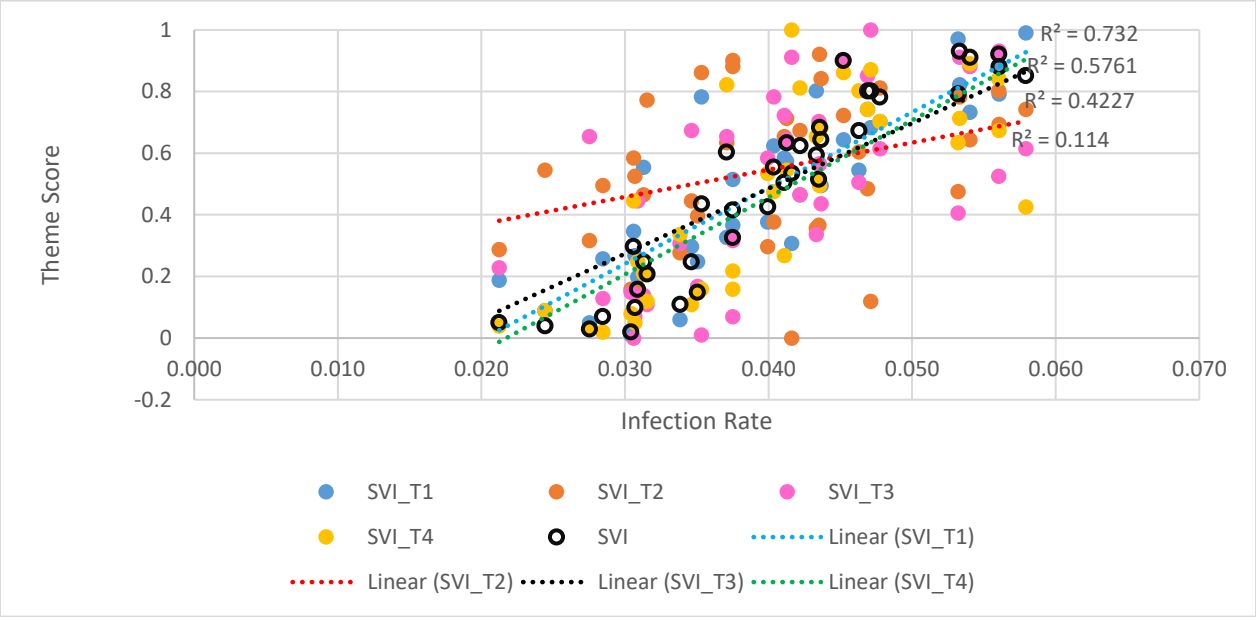

Figure 27. SVI themes and infection rates - selected counties

APPENDIX K

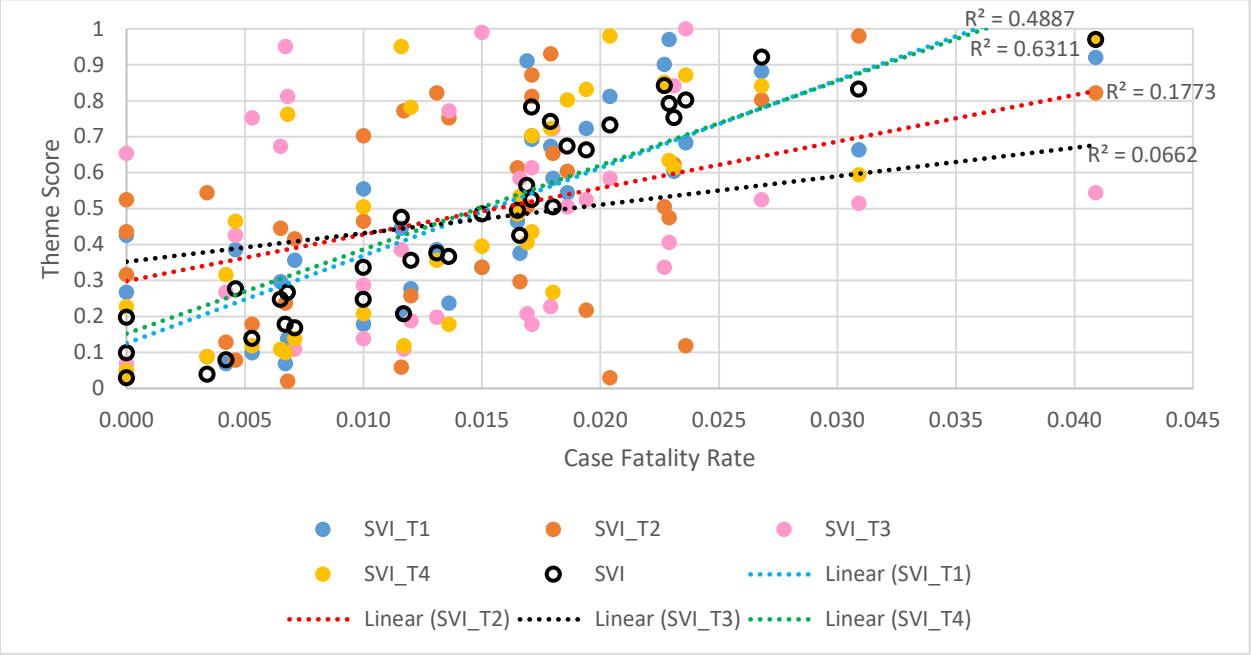

Figure 28. SVI themes and case fatality rates- selected counties

APPENDIX L

TABLE XXIII. COMPARISON OF RELATIVE IMPORTANCE OF SVI THEMES TO INFECTION RATES AMONG THREE APPROACHES

Methods	Rank			
PCA	T1	T4	T2	T3
Linear regression	T3	T4	T1	T2
Random Forest	T3	T4	T1	T2

APPENDIX M

TABLE XXIV. COMPARISON OF RELATIVE IMPORTANCE OF CCVI THEMES TO INFECTION RATES AMONG THREE APPROACHES

Method	Rank					
PCA	T1	T2	T4	T5	T6	T3
Linear regression	T3	T6	T4	T2	T1	T5
Random Forest	T3	T2	T4	T1	T6	T5

VITA

NAME: Haewon Oh

EDUCATION:

B.S., Economics, Dankook University, South of Korea, 2014

M.A., Economics, Seoul National University, South of Korea, 2016

HONORS:

Brain Korea 21+ Scholarship, Seoul National University, South of Korea, 2016-2018.

Superior Academic Performance Scholarship, Dankook University, South of Korea, 2011 and 2013.

Merit-Based Scholarship, Dankook University, South of Korea, 2010-2013.

PUBLICATIONS:

Kim, S.J., Oh, H.W., Chang, J.W., and Kim, S.J.: Recovery of Tendon Characteristics by Inhibition of Aberrant Differentiation of Tendon-Derived Stem Cells from Degenerative Tendinopathy. Int. J. Mol. Sci. 21. 2020.

Kim, S.J., Oh, H.W., and Choi, S.J. "Evaluating Information Campaigns to Reduce Food Waste: A Field Experiment." Seoul National University Journal (In Korean). 2017