Middlebury

Emily Zhou
Advisor: Joseph Holler

Social Inequities under Public Health Crisis

A Reproduction Study of Chakraborty 2021

Co-author: Drew An-Pham, Joseph Holler, Derrick Burt
Middlebury College Department of Geography

INTRODUCTION

In this study, we reproduced a national-scale study on COVID-19 incidence rate and socio-demographic characteristics of people with disability by county.

Throughout this process, we:

- Implemented Chakraborty's study to the greatest extent possible in R/RStudio.
- Validated Chakraborty's observation that **COVID-19 incidence is higher in counties with more socio-demographically disadvantaged people with disabilities**.
- Published a reproduction research compendium available to the public on GitHub and registered the analysis plans and reports with OSF.io.

The goal of our reproduction study is to **expand the accessibility, transparency, and potential impact of Chakraborty's research** to facilitate public health decision-making, resource allocation, and the ability of students and researchers of spatial epidemiology to review, extend, modify the study.

ORIGINAL STUDY

Chakraborty obtained the county level **sociodemographic subcategories of people with disability data** from the 2018 American Community Survey (with race, ethnicity, poverty, age, and biological sex as the subcategories within disability) and the **COVID-19 incidence data** from the Johns Hopkins Coronavirus Resource Center.

01 Map COVID-19 incidence rate and computed summary statistics for sociodemographic variables.

02 Test for **county-level bivariate correlations** between COVID-19 incidence against percentage of disability and socio-demographic category.

03 Implement the **Kulldorff spatial scan statistic** to determine the relative risk score of counties falling within significant COVID incidence clusters.

04 Use **generalized estimating equation** models to estimate the relationship between COVID incidence and disability subgroups while controlling for the state and relative COVID risk of each county.

DEVIATIONS AND ENHANCEMENTS

Conceptualization and Design

Before: COVID-19 spatial clusters are used as a control in models predicting COVID-19 rates, but the original study operationalized COVID-19 risk as the local relative risk of the county at the center of the cluster.

After: Since doing so excluded all but the center county of each cluster and assigned the other counties to the lowest risk category (Figure 1), *we changed our conceptualization of COVID-19 clusters to include all counties within any clusters* (Figure 2).

Measurement and Data Processing

Before: The original study implemented the Kulldorff spatial scan statistics in the SaTScan software and the generalized estimating equation model in SPSS, while using ArcGIS for mapping.

After: *We implemented the study with a fully open-source processing environment.* We used the SpatialEpi package in R for the spatial scan statistics, the geepack package for the GEE model, and the tmap package for data visualization.

Communication

Before: The original study does not communicate the geographic distribution of disability population, COVID-19 cluster conceptualization, and COVID-19 relative risk classifications.

After: We *visualized disability rates by county*. We also ran the spatial scan statistics in SaTScan and visualized COVID-19 clusters based on SaTScan's output and for *both the local and cluster-based relative risk* score classifications.

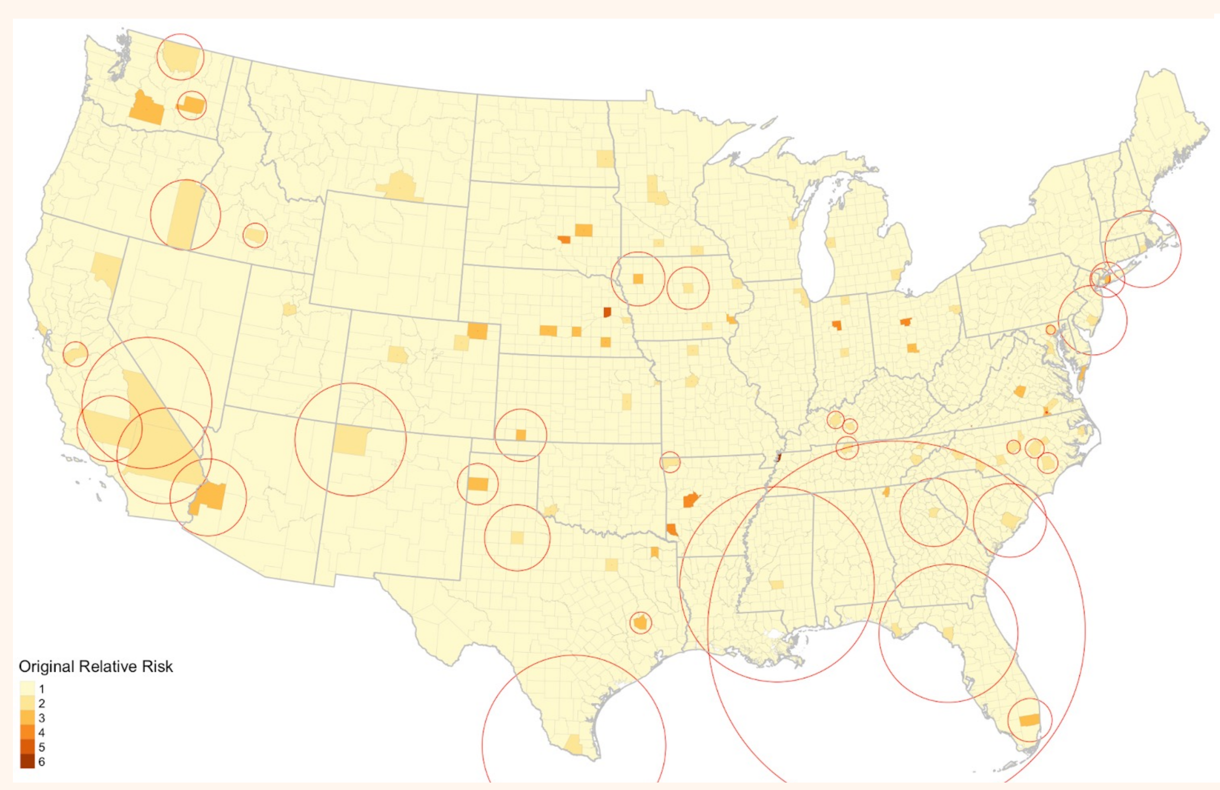

Figure 1: Original Relative Risk Classification

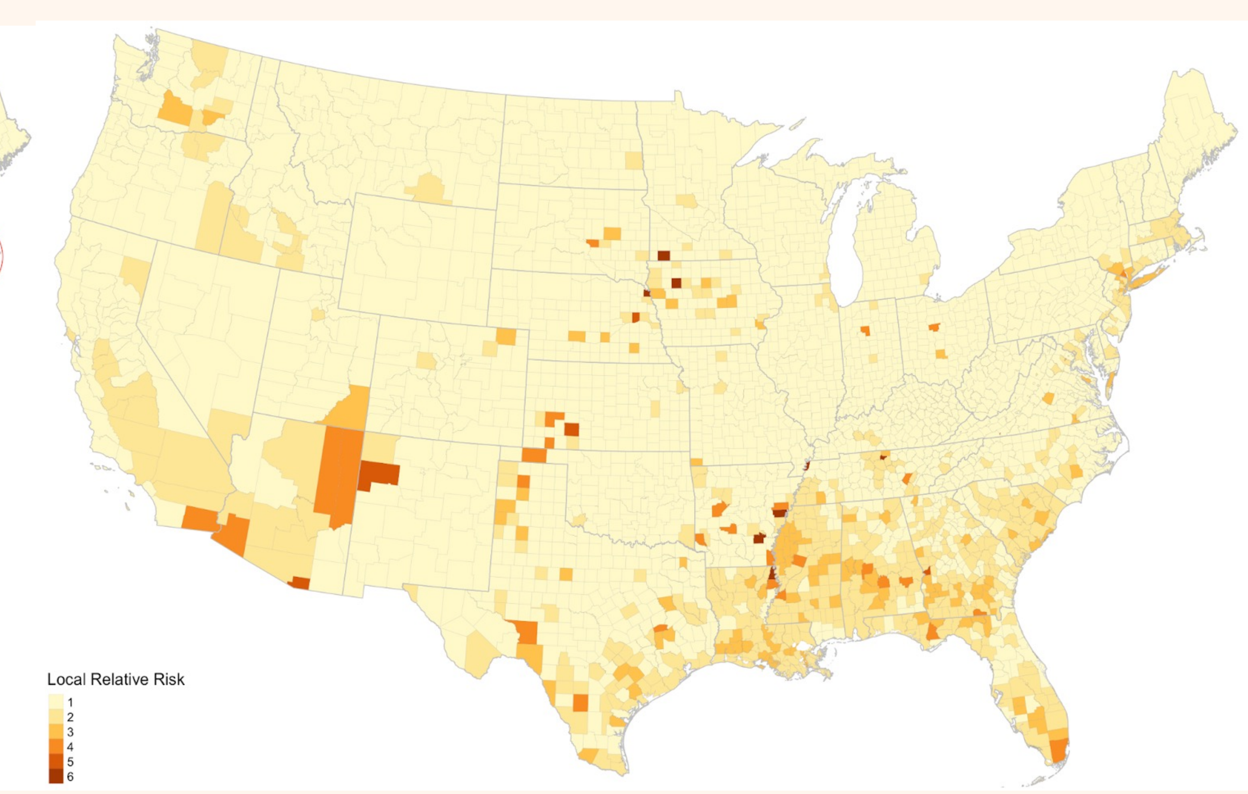

Figure 2: Local Relative Risk Classification

Analysis and Inference

Before: We found that most of the independent variables had significantly non-normal distributions.

After: We used the *nonparametric Spearman's rank correlation* coefficient for bivariate tests of correlation between the independent variables and COVID-19 incidence rate.

We further reconceptualized the classification of COVID risk *using cluster-based relative risk* (Figure 3) to account for spatial dependence within state and COVID-19 hotspots.

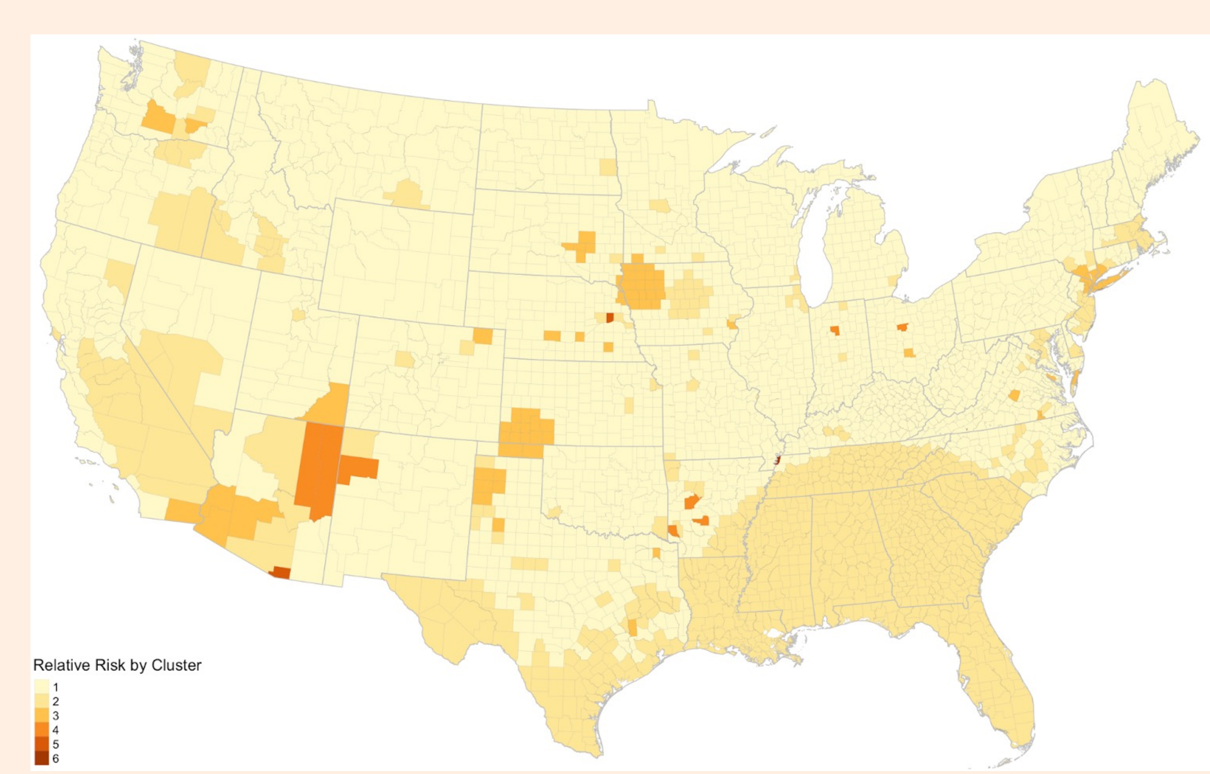

Figure 3: Cluster Based Relative Risk Classification

RESULTS

Our reproduction results are inexact but mostly consistent with that of the original study. There are some changes in direction and magnitude of the correlations as we compare the Pearson's and Spearman's results. We implemented three versions of the GEE models. 1) the original data in R & geepack computational environment, 2). reproduced data & a local relative risk score, 3) reproduced data & a cluster relative risk score.

Table 2: GEE Equations

	Our Coef	Orig Coef	Coef Diff	Our sig	Orig sig*	Our QIC	Jay QIC	QIC Diff
race						2616.4	2582.5	33.9
(Intercept)	7.72	7.11	0.62	< 0.001	< 0.01			
white_pct	-0.13	-0.20	-0.07	< 0.001	< 0.01			
black_pct	0.02	0.11	-0.09	< 0.05	< 0.01			
native_pct	0.02	0.05	-0.03	< 0.001	< 0.01			
asian_pct	0.02	0.08	-0.06	< 0.001	< 0.01			
other_pct	0.02	0.08	-0.06	< 0.001	< 0.01			
ethnicity						2616.3	2586.6	29.8
(Intercept)	7.72	7.19	0.53	< 0.001	< 0.01			
non_hisp_white_pct	-0.15	-0.24	-0.09	< 0.001	< 0.01			
hisp_pct	0.01	0.12	-0.11	0.198	< 0.01			
non_hisp_non_white_pct	0.02	0.12	-0.10	< 0.01	< 0.01			
poverty status						2562.7	2801.5	-238.8
(Intercept)	7.77	7.18	0.59	< 0.001	< 0.01			
bpov_pct	0.02	0.15	-0.13	< 0.01	< 0.01			
apov_pct	-0.11	-0.27	-0.16	< 0.001	< 0.01			
age						3577.1	2978.7	598.3
(Intercept)	7.78	7.24	0.54		< 0.01			
pct_5_17	0.02	0.05	-0.03	< 0.001	< 0.01			
pct_18_34	0.01	0.04	-0.02	< 0.001				
pct_35_64	-0.02	-0.03	0.00	< 0.01				
pct_65_74	-0.06	-0.09	-0.03	< 0.001	< 0.01			
pct_75	-0.05	-0.11	-0.06	< 0.001	< 0.01			
biological sex						2012.3	2892.4	-880.1
(Intercept)	7.78	7.22	0.56	< 0.001	< 0.01			
male_pct	-0.14	-0.30	-0.16	< 0.001	< 0.01			
female_pct	0.04	0.15	-0.11	< 0.001	< 0.01			

Figure 4: Comparing GEE Models

Comparing across the parameter estimates,

- *The race and ethnicity variables were most sensitive to changes in the computational environments and different operationalizations of COVID clusters.*

Comparing the average differences in estimated independent variable coefficients,

- *The GEE models can be sensitive to different computational environment alone.*

CONCLUSIONS

- Our study supports Chakraborty's observation that people with disabilities are likely to experience multiple jeopardies during the pandemic.
- We adjusted the bivariate correlation calculation and reconceptualized COVID-19 clusters.
- We identified several recurrent challenges in geospatial research, including sensitivity of the GEE model to varying operationalizations of geographic clusters.
- Further studies should build upon this initial exploratory analysis to test the hypothesis of higher burden of COVID-19 on disability population with new data.

REFERENCES: Chakraborty, J. 2021. Social inequities in the distribution of COVID-19: An intra-categorical analysis of people with disabilities in the U.S. *Disability and Health Journal* **14**:1-5. DOI:10.1016/j.dhjo.2020.101007