COVID-19 Community Vulnerability Index (CCVI) Methodology

Updated December 2020

Overview

Surgo Ventures constructed a COVID-19 Community Vulnerability Index (CCVI) to assess which US communities may be less resilient to the impacts of the COVID pandemic. The CCVI is a modular index that was built to capture the multi-dimensionality construct of vulnerability spanning health, economic, and social disadvantages at the neighborhood level (i.e. census tract). This current release constitutes a newer version from previous CCVIs to incorporate insights from a plethora of outcome data that has been collected throughout the evolution of the pandemic.

Mapped nationally at the state, county, and census tract levels, the CCVI can aid in COVID-19 planning and mitigation at a granular level. The CCVI builds on the Centers for Disease Control and Prevention's (CDC) Social Vulnerability Index (SVI), a validated metric intended to help policy makers and public health officials respond to emergencies. The current COVID-19 outbreak poses new challenges contingent on a host of health, economic, and structural factors, not all of which are captured in the SVI. To understand vulnerability within the context of the coronavirus pandemic, COVID-specific epidemiological risk factors, public health system capacity, and variables capturing specific high risk environments known to facilitate the spread of COVID-19 were combined with SVI sociodemographic variables. The 40 variables cover 7 core themes: Social Vulnerability themes, as well as COVID-specific themes to account for additional factors that make a community or individual susceptible to the COVID-19 pandemic (Table 1). The composite CCVI metric ranks each geography (state, county, or census tract) relative to one another on a 0 - 1 scale (0 = least vulnerable, 1 = most vulnerable). Data was sourced from the CDC, Centers for Medicare, & Medicaid Services (CMS), the Harvard Global Health Institute, PolicyMap, the U.S Bureau of Labor Statistics (BLS), the U.S. Census Bureau (USCB), the National Cancer Institute (NCI), the National Center for HIV, STD and TB Prevention (NCHSTP), Kaiser Health News, Centers for Medicare & Medicaid Services (CMS), Census Economic Annual Surveys, the Vera institute for Justice, and the Association of Public Health Laboratories.

Main differences between previous version of CCVI

An in depth analysis as to the statistical variability between the newer and older versions of the CCVI will be released in the coming weeks after the initial release of the new CCVI. It is within the scope of this document to discuss the conceptual differences between the 2 indexes.

1. The baseline SVI themes have been rearranged into themes that address similar types of vulnerability. Several indicators have been shown to be correlative among the sociodemographic themes, particularly between the previous *Household Composition & Disability* theme and the *Housing Type & Composition* theme. They have been combined to address vulnerability according to housing type and dependents.
    a. An indicator around *access to indoor plumbing* was added to help address housing type vulnerabilities.
2. The *age over 65* indicator has proven to be an epidemiological risk instead of a socioeconomic one. It has been removed from the *Household Composition* theme and embedded within the *Epidemiological* theme.
3. Most of the *Epidemiological* risk indicators have been replaced with census tract level granularity (previously county level).
4. A sub-theme of *Healthcare accessibility* (i.e. percent of population with a primary care physician) was added to the *Healthcare System* theme. We also added *aggregate cost of medical* care (both from insurance and out-of-pocket) to bolster the *Healthcare strength* sub-theme.
5. A *High Risk Environment* theme has been added to address living and working conditions that lead to adverse and disproportionate COVID-related outcomes (e.g. nursing home residents).
6. Population density has been made into its own theme, implicitly giving it much more weight compared to other indicators in the index.
7. Construction of the county and state level index has been updated to a more direct approach. The county-level CCVI is now constructed using all county-level (or state where applicable) raw indicators as done for the census tract-level index. The raw indicators are the exact same as what is used to construct the census tract CCVI, but their base geographic unit is the county. This creates a more direct measure of the index at the county (and state) level instead of generating population-weighted averages of census tract-level ranks.

surgo VENTURES

COVID-19 Variable Selection

The CDC SVI links socioeconomic status, household composition and disability, minority status and language, and housing type and transportation as a composite metric representative of populations disproportionately affected by and less resilient to disasters. Social issues can be impacted across the disaster cycle, such as economic and infrastructure loss, and demographic characteristics, such as age, race, and economic status, can underlie the differential impacts of hazardous events.[1] Though the SVI applies to a variety of emergencies, including natural events (e.g. hurricanes) and disease outbreaks, COVID-19 has brought unprecedented challenges globally.[2] To adequately account for COVID-19 vulnerability, we modified the SVI themes and compiled 4 additional themes to address epidemiological, healthcare systems and high risk environment factors influenced by the COVID-19 pandemic.

Themes 1-3: Modifications of the CDC's SVI

Themes 1-3 are modified from the CDC SVI themes 1-4. This constitutes a departure from the previous versions of the CCVI where themes 1-4 were taken directly from the SVI and used as a baseline of sociodemographic vulnerability. The uninsured population was added to Theme 1 to account for increased vulnerability to severe COVID-19 outcomes due to delayed or complete avoidance of health care seeking behavior. Minority Status (CCVI Theme 2) was integrated from SVI theme 3 and was not modified. SVI Themes 2 and 4 were merged to create CCVI theme 3 - Housing type, Transportation, Household Composition and Disability. Access to indoor plumbing was added to theme 3 to account for characteristics that might limit the ability of household members to avoid exposure to COVID-19. This theme also comprises households that have dependents, and would therefore be at a greater socioeconomic or mobility disadvantage.

Theme 4: Epidemiological Risk Factors

Theme 4 epidemiological factors were selected according to CDC guidelines, which identify high risk populations as elderly adults and individuals with underlying conditions including respiratory conditions, heart conditions, obesity, diabetes, and conditions related to

[1] Flanagan, Barry & Gregory, Edward & Hallisey, Elaine & Heitgerd, Janet & Lewis, Brian. (2011). A Social Vulnerability Index for Disaster Management. Journal of Homeland Security and Emergency Management. 8. 10.2202/1547-7355.1792.
[2] World Health Organization. "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020." 11 Mar. 2020.

immunodeficiency.[3] These indicators have been enhanced from previous versions of the CCVI because they are now at a census tract geographic resolution, which offers much more variability than the previous county level variables.

Theme 5: Healthcare System Factors

Theme 5 variables were selected as a measure of the capacity, strength, accessibility and preparedness of the healthcare system in regards to COVID-19. Hospital beds are needed to accommodate the influx of coronavirus patients, who stay an average of 11-12 days in care.[4,5] Healthcare capacity has been an issue of growing importance as we head into the winter months. The density of epidemiologists was included as a proxy measure for state capability of COVID-19 surveillance and contact tracing, an effective method to detect cases and slow COVID-19 spread.[6,7,8] Accessibility to healthcare information and treatment from trusted sources is captured by the percentage of the population with a primary care physician. Health system strength, as measured by total health expenditure and quality of care, can reflect potential effectiveness of an outbreak response.The Agency for Healthcare Research and Quality's (AHRQ) prevention quality indicator (PQI) composite was selected as a metric of poor acute care in the context of patients with chronic conditions.[9] Additional factors on health system preparedness to address disease outbreaks include public health emergency funding available from the CDC[10] and the density of emergency services for rapid response. Public health laboratory density was also included as a proxy of readiness to test, which is essential to understanding and mitigating the COVID-19 spread in respective communities.[11]

---

[3] Centers for Disease Control and Prevention. People who are at higher risk for severe illness. Accessed March 31, 2020.
[4] W.-j. Guan *et al.*, "Clinical Characteristics of Coronavirus Disease 2019 in China," *New England Journal of Medicine,* 2020.
[5] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *The Lancet,* vol. 395, no. 10229, pp. 1054-1062, 2020.
[6] Ng Y, Li Z, Chua YX, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore — January 2–February 29, 2020. MMWR Morb Mortal Wkly Rep 2020;69:307-311
[7] Q. Bi *et al.*, "Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts," *medRxiv,* p. 2020.03.03.20028423, 2020
[8] J. Hellewell *et al.*, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *The Lancet Global Health,* vol. 8, no. 4, pp. e488-e496, 2020.
[9] Agency for Healthcare Research and Quality (2019). *Quality Indicator User Guide: Prevention Quality Indicators (PQI) Composite Measures*.
[10] CDC, Center for Preparedness and Response (2020). *Emergency Preparedness Funding*.
[11] M. Fisher and C. Sang-Hun, "How South Korea Flattened the Curve," in *The New York Times*, ed, 23 March 2020.

Theme 6: High Risk Environment

Differences in living and workplace conditions can put sub-groups of the population at increased risk of contracting COVID-19. These "high risk" conditions - proximity to and interaction with other people, and exposure to diseases - and their geographical distribution can make counties more or less vulnerable to COVID-19. Also, current estimates show that nursing home residents account for ~38% of all COVID-related deaths while only accounting for 5% of the cases[12]. This disproportionate ratio in mortality to case load prompts a drastic response and inclusion in a COVID-specific vulnerability index.

Workplace crowding and high exposure environments can also make communities more vulnerable to the effects of COVID-19. For instance, rural counties were particularly affected by outbreaks due to their employment dependence on industries with high-risk work conditions like the meatpacking industry[13]. Around 50 counties in the US employ over 20% of their workforce in animal slaughtering and processing. The CCVI captures this community reliance on high risk employment by considering industries where workers are in enclosed spaces for extended time, in proximity of or interacting with others, and/or exposed to diseases. The selection of "high risk" industries was informed by O*Net's database on workplace conditions by occupation[14].

Theme 6 - high risk environments and jobs - includes indicators for sub-groups of the population whose living or work environments puts them at high risk of contracting the virus:

1. Nursing home and assisted living residents
2. Prison population
3. Workers in high-risk industries. Indicator: Percentage of the county's population employed in a high-risk industry, where high risk industries were selected based on the frequency and duration of contacts with other people in the workplace.These include meat and poultry processing, manufacturing and passenger ground transportation (see table below).
    a. Filtered by % of the population of the county that a specific industry concentrates.

[12] More Than 100,000 U.S. Coronavirus Deaths Are Linked to Nursing Homes," in *The New York Times.* https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html.

[13] USDA. "The Meatpacking Industry in Rural America During the COVID-19 Pandemic," Retrieved December 14, 2020 from https://www.ers.usda.gov/covid-19/rural-america/meatpacking-industry/

[14] National Center for O*NET Development. *O*NET OnLine*. Retrieved November 6, 2020, from https://www.onetonline.org/

i. A community (i.e. county) that is dominated by one type of employer (manufacturing/production/etc.)

High-risk industries included

Industry NAICS Code	Short name
311	Food_mfg
312	Beverage_tobacco_mfg
313	Textile_mills
314	Textile_product_mills
315	Apparel_mfg
316	Leather_and_product_mfg
321	Wood_product_mfg
322	Paper_mfg
323	Printing_and_support_activities
324	Petroleum_coal_mfg
325	Chemical_mfg
326	Plastics_rubber_mfg
327	Nonmetallic_mineral_product_mfg
331	Primary_metal_mfg
332	Fabricated_metal_product_mfg
333	Machinery_mfg
334	Computer_and_electronic_product_mfg
335	Electrical_equipment_and_appliance_mfg
336	Transportation_equipment_mfg
337	Furniture_and_related_product_mfg
339	Miscellaneous_mfg
485	Transit_ground_passenger_transport

Theme 7: Population Density

Population density has been shown to be a good predictor of dynamic health outcomes influenced by COVID-19 throughout the evolution of the pandemic specifically in the US[15]. Population density not only increases the total number of people in a defined area, but it also presents the opportunity for more interactions among the population because there is more connectivity among the population. The US has had much more lenient restrictions regarding mobility throughout the pandemic compared to other countries around the world, which could explain why population density is a bigger factor in health outcomes in the US specifically. Other countries with more stringent measures regarding decreased mobility among its residents may be able to mitigate the impact population density could have on the transmission of the virus. However, in the US, these measures were not as strict, thereby creating continued opportunities for the transmission of the virus in denser communities.

There have been studies suggesting that population density is not the sole factor contributing to the adverse COVID-related impacts in cities[16]. These studies suggest that other factors such as socioeconomic status, crowding, and total metropolitan size are confounding factors that should be considered. The CCVI also accounts for these and more confounding factors, and together with population density can create a holistic view into vulnerability throughout the pandemic.

We also ran regression analyses with individual indicators against COVID-specific outcomes, and population density consistently was one of the top predictors of adverse health-outcomes justifying its inclusion in the index. While population density may not be the "exact" or "only" indicator leading to adverse COVID outcomes, it certainly warrants inclusion in the CCVI index.

Construction of V3

Social vulnerability indexes have historically been constructed according to 4 different methods:

1. **Inductive index**

---

15 Neiloy Medhi. COVID-19 Cases in the US by Population Density. Retrieved December 14, 2020 from https://www.endcoronavirus.org/projects-1/covid-19-cases-in-the-us-by-population-density

16 Jeremy Hsu. "Population Density Does Not Doom Cities to Pandemic Dangers", Retrieved December 14, 2020 from https://www.scientificamerican.com/article/population-density-does-not-doom-cities-to-pandemic-dangers/

    a. Applies factor analysis based on principal component analysis (PCA) to reduce initial indicator set into a smaller number of latent (unobserved) factors.
2. **Hierarchical configurations with weighting** → conceptually the CCVI
    a. Aggregate indicators into themes that share an underlying dimension of COVID vulnerability (e.g. socioeconomic status, health, etc.)
    b. The themes are then aggregated to create the index
    c. Such indexes require a greater level of theoretical organization than a more data-driven inductive-model
    d. Can create weights using SMEs (subject matter experts)
3. **Deductive design without weighting** → functionally the CCVI
    a. Themes are mathematically ignored in the aggregation of individual indicators to create the index (e.g. no weighting/implicit weighting)
4. **Spatially varying typologies** (strongest statistically - but not the best programmatically)
    a. Construct a set of varying vulnerability profiles using the same dataset (depending on the geography)
    b. Correlation analyses are used to reduce the number of variables which limits collinearity, prevents implicit weighting, strengthens statistical power, and preserves a balance between the different dimensions of vulnerability
    c. Remove variables that have strong correlations with each other
        i. E.g. Share of female-headed households used to represent all of the gender variables due to the strong correlation among them
        ii. E.g. share of non-white as a surrogate for all popualtion identifying with ethnic and race minorites which reduces selection of variables considerably
    d. Caveat: While statistically the best, it presents problems when breaking the index down to explain specific vulnerability in specific geographies, which has become a main feature of the CCVI

The CCVI is a combination of hierarchical and deductive configurations. While the CCVI is conceptually a *hierarchical design* since it groups individual variables into themes of similar vulnerability types, it is functionally a *deductive configuration* since it ignores mathematical aggregation (i.e. weighting) of individual indicators and relies on *implicit* weighting to provide the statistical variance.

Spatially varying typologies (i.e. profiles) offer the strongest statistical approach to conducting an index since it decreases collinearity, prevents implicit weighting, and increases statistical power. However, it reduces interpretability of specific indicators of vulnerability by limiting indicators to surrogates or proxies of vulnerability. This makes it difficult programmatically to suggest targeted interventions based upon specific "vulnerable" populations.

Differences in construction from the previous version

During construction of the previous version of the CCVI back in March/April of 2020, there were no COVID-specific outcomes to inform the inclusion of specific indicators into the CCVI. For this reason, the CDC's SVI variables were left untouched and the additional indicators were sourced based on preliminary evidence of COVID-specific complications surrounding pre-existing conditions and the burden on the healthcare system.

At the time of generating updates for this newest version of the CCVI in November of 2020, a plethora of COVID-influenced outcome data had been collected throughout the pandemic. The data captured spanned the range of different dimensions of negative impacts including health, economic, and social outcomes that the CCVI was built to address. Therefore, elements of an *inductive index* were utilized to test the validity of the inclusion of the additional indicators apart from the CDC's SVI. Regression analyses for each indicator within the Epidemiological, Healthcare, and High Risk Environment themes were run against the negative outcome data[17] compiled by the Opportunity Insights Tracker for every day since the beginning of the pandemic in which data was available. Each variable held some statically significant predictive power and explained some of the variance in the outcome data justifying their inclusion within the index with the exception of hospital beds and icu unit beds per capita. Historically throughout the pandemic in the US, these indicators did not hold any predictive power to COVID-related mortality, which would be the reason for their inclusion. However, news stories[18] regarding the strain on healthcare system capacity as we enter the winter months suggest that these indicators may be needed now more than ever.

How to deal with missing data (i.e. -999) values in SVI raw data

Dealing with the SVI data from the CDC presents its own challenges, and we briefly introduce and record our suggestions for dealing with the data munging process. There are cases where no estimate is provided (i.e. -999) for census tracts across the US. The reason for the missingness varies. Therefore, the recommendation for how to deal with the missingness depends on the reason for suppression or omission.

---

[17] Chetty et al. 2020. Opportunity Insights Tracker. https://opportunityinsights.org/tracker-resources/

[18] NBC News. "As Covid hospitalizations soar, states struggle to find enough beds and staff," Retrieved December 14, 2020 from https://www.nbcnews.com/news/us-news/covid-hospitalizations-soar-states-struggle-find-enough-beds-staff-n1247432

Tribal census tracts versus county census tracts

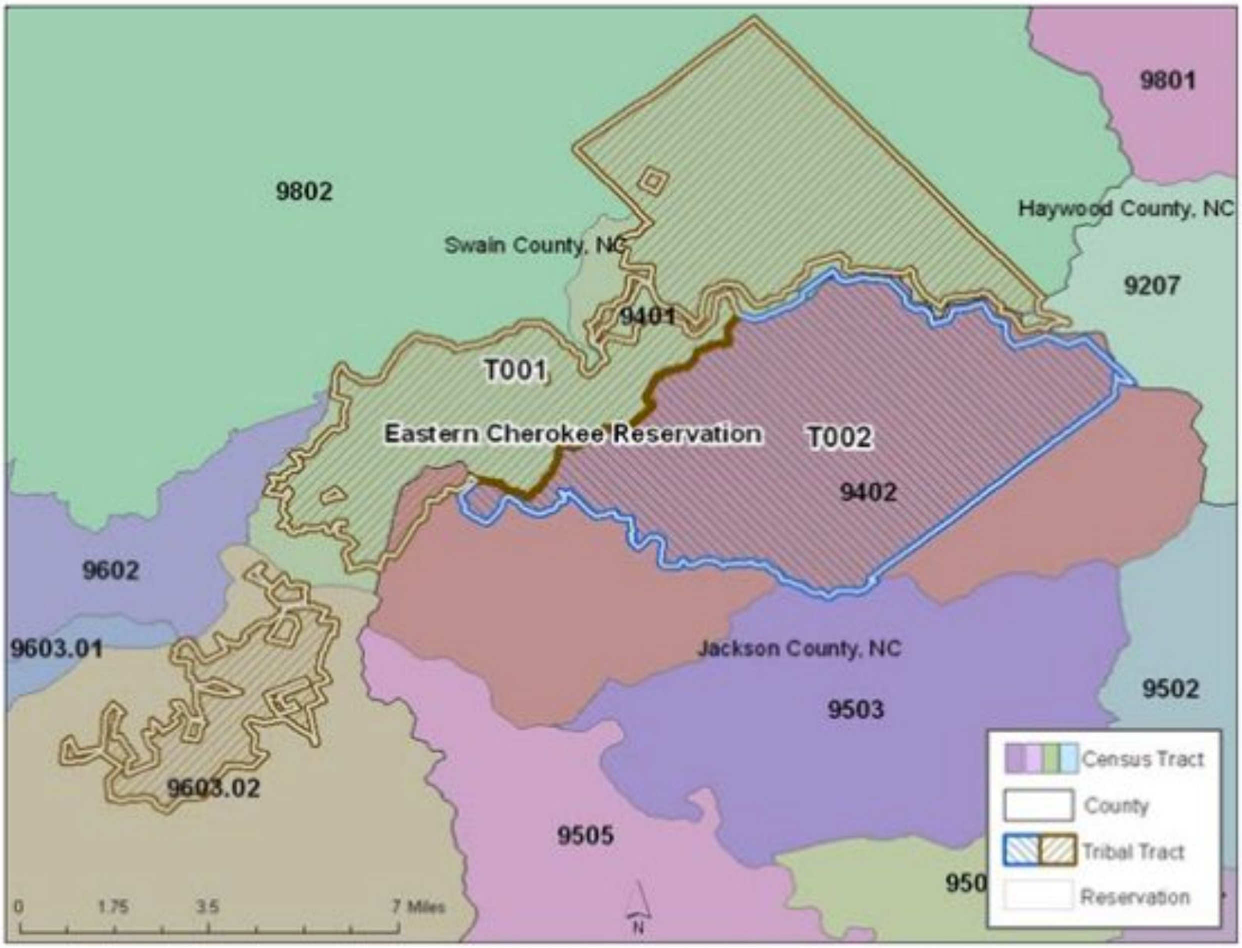
Starting with the SVI in 2014, the CDC added a database of Tribal Census Tracts. Tribal tracts are defined independently of, and in addition to, standard county-based tracts. **Because of geographic separation and cultural diversity, tribal tracts are not ranked against each other nor against standard census tracts.**

For information about decoding the differences between the 2 types of census tracts, we refer you to this Census Bureau explanation: Decoding State-County Census Tracts versus Tribal Census[19].

Standard county-based census tracts often cannot be fully utilized by those living on American Indian reservations because these tracts often do not conform to their reservation boundaries. Therefore, tribal census tract boundaries were created to only incorporate land within the federally recognized reservation boundaries.

State-County Census Tracts	Tribal Census Tracts
Unique to and nested within the county	Unique to and nested within the Reservation
	Can cross state and county lines
	Can be discontinuous
Numeric 6 digit name (e.g. 3042, 1801.02)	Name starts with a 'T' (e.g. T001, T002)
Cover the entire U.S.	Cover only Federally-recognized American Indian Areas with land

[19] US Census Bureau. "Decoding State-County Census Tracts versus Tribal Census Tracts," Retrieved December 14, 2020 from https://www.census.gov/newsroom/blogs/random-samplings/2012/07/decoding-state-county-census-tracts-versus-tribal-census-tracts.html#:~:text=A%20standard%20census%20tract%20must,and%20off%2Dreservation%20trust%20land



These tracts have also been given a total population (i.e. E_TOTPOP) of 0. **Tracts with zero estimates for total population (N = 645 for the U.S. or <1%) were removed during the ranking process.** These tracts were added back to the SVI databases after ranking. The E_TOTPOP field value is 0, but the percentile ranking fields (RPL_THEME1, RPL_THEME2, RPL_THEME3, RPL_THEME4, and RPL_THEMES) were set to -999.

Recommendation: Provided the discontinuity in the geographies as well as the cultural diversity, it is recommended to leave these estimates as NAs (i.e. not imputed) and to not include them in the overall ranking as done by the CDC as well as the procedure followed in construction of the CCVI. Also, leave the E_TOTPOP field as 0 so the population of these census tracts do not influence the weights when aggregating to higher geographic levels.

Omission due to unavailability of census data

For tracts with > 0 TOTPOP, a value of -999 in any field either means the value was unavailable from the original census data or we could not calculate a derived value because of unavailable census data. Any cells with a -999 were not used for further calculations. For example, total flags do not include fields with a -999 value.

- Only 238 census tracts that have E_TOTPOP > 0 are missing RPL_THEMES (overall SVI) scores which constitutes <1% of all census tracts

Recommendation: Since the missingness is relatively small (< 1%) and there is no data available for these census tracts, it is recommended to keep these variables as NAs and ignore them in the ranking calculations. We do not know the range of *socioeconomic status* for these census tracts, so taking the mean or median would be erroneous.

Rio Arriba County

The U.S. Census Bureau reports that data collection errors prohibited the inclusion of income and poverty data from Rio Arriba County, New Mexico. Please see a more detailed explanation provided by the Census Bureau here: Data Collection Error in Rio Arriba County, NM[20].

However, data from the Federal Reserve Bank of St. Louis (FRED)[21], indicates that the *per capita income (PCI)* of Rio Arriba county (i.e. $34,582) in 2018 is very similar to the average PCI for the entire US as reported by the CDC SVI (i.e. $32,280) in 2018.

Recommendation: Given the major similarity in the average *socioeconomic status* (i.e. per capita income) of Rio Arriba to the rest of the US and the fact that data is suppressed in CDC data simply for collection errors, it is recommended to impute the *mean* of each indicator for all of Rio Arriba county.

Creating the CCVI Composite

All construction of the CCVI was conducted using the statistical coding software, R. Variables per CCVI theme were represented by percentiles (using the `percent_rank`

[20] US Census Bureau. "Data Collection Error in Rio Arriba County, NM," Retrieved December 14, 2020 fromhttps://www.census.gov/programs-surveys/acs/technical-documentation/errata/125.html

[21] Federal Reserve Bank of St. Louis. "Per Capita Personal Income in Rio Arriba County, NM", Retrieved December 14, 2020 from https://fred.stlouisfed.org/series/PCPI35039.

function), a statistical measure ranking each data point in relation to the full dataset (e.g. the 20th percentile represents the value below which 20% of the data falls). To create a composite CCVI measure, percentiles of each variable were aggregated (i.e. summed) per CCVI theme. All sub-theme buckets were removed in this version of the CCVI construction meaning equal weight is provided to each indicator within a given theme. Each variable percentile was then aggregated per CCVI theme and subsequently across all 7 themes to create one metric. All themes were weighted equally. This method of aggregation was based on the CDC construction of the SVI. Vulnerability was classified into CCVI quintiles illustrating very low (<20%), low (20-40%), moderate (40-60%), high (60-80%), and very high vulnerability (>80%).

Through this method, the CCVI makes each geographic unit a comparison metric with every other similar unit across the entire US (i.e. it allows for a census tract in Miami-Dade, FL to be compared to a census tract in rural Montana). The CCVI was constructed at the county and state levels using the same method as for the census tract index except using county or state geographic representations of the raw indicators in the index.

We explain the process for generating the CCVI at the census tract level using bullets below:

1. For each individual variable (i.e. percent of population living below the poverty line) at the census tract level (or high geographic unit) → Percentile ranks were applied to generate a range of 0 - 1 for each census tract (or county or state) across the US
2. For each percentile ranked variable within a specific theme (i.e. Theme 1 → SES Status) → These ranks were summed across each census tract (or county or state)
3. Percentile ranking was once again applied to these aggregated ranks → Giving the theme ranks for each of the modular Themes (Ranks between 0 - 1).
4. The theme ranks were again summed to get the aggregated ranking for the entire CCVI for each census tract
5. Percentile rankings were applied to this final aggregation to get the overall CCVI score at the census tract-level (or county or state)

Mock-up of CCVI indicators and themes

Table 1: COVID-19 Community Vulnerability Index (CCVI)

	Theme	Variable	Indicator(s)	Geo Precision	Source
1	**Socioeconomic Status**	Below Poverty	Persons below poverty estimate	Census Tract	CDC, American Community Survey (ACS), 2014-2018 5-Year Estimates
		Unemployed	Civilian (age 16+) unemployed estimate		
		Income	Per capita income estimate		
		No High School Diploma	Persons with no high school diploma (age 25+) estimate		
		Uninsured	Percent of population uninsured		
2	**Minority Status & Language**	Minority	Minority (all persons except white, non-Hispanic) estimate		
		Speaks English “Less than Well”	Persons (age 5+) who speak English "less than well" estimate		
3	**Housing type, Transportation, Household Composition & Disability**	Multi-Unit Structures	Housing with structures with 10 or more units estimate		
		Crowding	Households with more people than rooms estimate		
		No Vehicle	Households with no vehicle available estimate		
		Group Quarters	Persons in institutionalized group quarters estimate		
		Aged 17 or Younger	Persons aged 17 and younger estimate		
		Single-Parent Households	Single parent households with children under 18 estimate		
		Access to Indoor Plumbing	Households without access to indoor plumbing		
		Mobile Homes	Mobile homes estimate		
		Older than Age 5 with a Disability	Civilian noninstitutionalized population with a disability estimate		

Table 1: COVID-19 Community Vulnerability Index (CCVI) (continued)

Theme		Variable	Indicator(s)	Geo Precision	Source
4	**Epidemiological Factors**	Cardiovascular Conditions	Estimated percent of adults diagnosed with high cholesterol	Census Tract (2017)	PolicyMap: CDC, Behavioral Risk Factor Surveillance System (BRFSS), 2017-2018
			Estimated percent of adults diagnosed with a stroke	Census Tract (2018)	
			Estimated percent of adults ever diagnosed with heart disease		
		Respiratory Conditions	Estimated percent of adults diagnosed with chronic obstructive pulmonary disease, emphysema, or chronic bronchitis	Census Tract (2018)	
			Estimated percent of adults reporting to smoke cigarettes		
		Immuno-compromised	Annual cancer incidence per 100,000 persons	County	PolicyMap: CDC, National Cancer Institute (NCI), 2011-2015
			Rate of persons living with an HIV diagnosis per 100,000 people		PolicyMap: CDC, National Center for HIV, STD and TB Prevention (NCHSTP), Division of STD/HIV Prevention, 2016
		Obesity	Estimated percent of adults reporting to be obese (a body mass index of 30 or greater)	Census Tract (2018)	PolicyMap: CDC, Behavioral Risk Factor Surveillance System (BRFSS), 2018
		Diabetes	Estimated percent of adults ever diagnosed with diabetes	Census Tract (2018)	
		Aged 65 or Older	Persons aged 65 and older estimate	Census Tract	CDC, American Community Survey (ACS), 2014-2018 5-Year Estimates

Table 1: COVID-19 Community Vulnerability Index (CCVI) (continued)

Theme		Variable	Indicator(s)	Geo Precision	Source
5	**Healthcare System Factors**	Health System Capacity	Intensive Care Unit (ICU) Beds per 100,000	County	Kaiser Health News, Centers for Medicare, & Medicaid Services (CMS), 2018-2019
			Hospital Beds per 100,000		Definitive Healthcare, 2020
			Epidemiologists per 100,000	State	U.S Bureau of Labor Statistics (BLS), Occupational Employment and Wages, May 2018
		Health System Strength	Agency for Healthcare Research and Quality - Prevention Quality Indicator Overall Composite (PQI): admission rates for preventable conditions (via good outpatient care) adjusted per population	County	Centers for Medicare, & Medicaid Services (CMS), Mapping Medicare Disparities (MMD) Tool, 2017
			Health Spending per Capita	State	Centers for Medicare, & Medicaid Services (CMS), Health Expenditures by State of Residence, 2014
			Aggregate cost of medical care	Census Tract	PolicyMap & Quantitative Innovations, 2017
		Healthcare Accessibility	Percent of population with a Primary Care Physician	Census Tract	PolicyMap & CDC BRFSS, 2018
		Health System Preparedness	Total Public Health Emergency Preparedness (PHEP) Funding Per Capita	State	CDC, Center for Preparedness and Response, 2019
			Health Labs per 100,000	County	Association of Public Health Laboratories

Table 1: COVID-19 Community Vulnerability Index (CCVI) (continued)

Theme		Variable	Indicator(s)	Geo Precisio n	Source
5	**Healthcare System Factors**	Health System Preparedness	Emergency Services per 100,000[22]	State	Census, Economic Annual Surveys, 2017
6	**High RIsk Environments**	Percentage of population working or living in environments with high infection risk	Long-term care (nursing homes, assisted living, and care homes) residents per 100,000	Census tract	ArcGIS/Dept of Homeland Security
			Prisons population per 100,000	County	Vera institute for Justice, 2016
			Percentage of population employed in high-risk industry[23]	County	BLS QCEW 2020
7	**Population Density**	Population Density	Estimated total number of people per unit area (sq. miles)	Census Tract	CDC Social Vulnerability Index

Mapping the CCVI

The CCVI was calculated nationally per census tract. For theme variables that were not at a census tract level (i.e. coarser geography such as county or state), the metric was mapped to the individual census tract units within that overarching geography (i.e. state-only metrics were reflected within the CCVI as the same value being attributed to every census tract within) to create the CCVI composite measure. Once calculated at the census tract level, CCVI data was aggregated to county and state levels by taking the average across the respective geographic unit and weighting the calculation by the population of each unit, and calculating the percentiles across geographic units again. ArcGIS, QGIS, Mapbox, and Tableau were used for geographic preprocessing and visualization to map the CCVI nationally per census tract, county, and state. Census tracts with missing SVI data were not included - shown as grayed map areas (n = 645, <1% of census tracts) - and ignored to compute the county- and state-level CCVI. The average for each variable was imputed for the small proportion of census tracts with available SVI data, but missing additional indicator data separate from the CDC's SVI variables.

[22] Includes emergency and relief services and freestanding ambulatory surgical and emergency centers.
[23] Includes employees in farming, manufacturing, printing and related support activities and textile NAICS subsectors.

Model Validation

True validation of the index can only occur deep into the outbreak, when accurate numbers regarding case fatality,economic, and social impact are available. It is also an iterative process to validate that the index has the recurring capacity to be predictive throughout a disaster-related event and not just at specific points in time. Model validity is the degree to which a model adequately represents its underlying construct. For the construct of COVID-vulnerability, to what extent does the index reflect multidimensionality, interactivity, and causal processes? In order to validate that the newest version of the CCVI will indeed capture vulnerability to COVID *better* than the previous CCVI version *and* the CDC's SVI, we needed a process to test the indexes against each other. Previous validation studies for social vulnerability indexes have focused on validation through *convergence* and *construct*.

Convergent validation

Convergent validation assesses the level of agreement among alternative models measuring the same construct (i.e. older version agreement with newer version).

- How well does the previous version correlate with the current version?
    - Do we want the two versions to be massively different?
- How much variance are we adding with the newest version?

Convergent validation is insufficient to demonstrate model validity, since two highly convergent (or divergent) models could both represent COVID vulnerability poorly.

Construct or empirical validation

Construct or empirical validation examines whether hypothesized processes underlying a measure are shown when tested against empirical data. Previous empirical studies have collected post-disaster data and assessed index performance using *statistical correlation analysis* and *ordinary least squares regression* (OLS). We have already done the work of collecting negative outcomes related to COVID and performed initial regressions using the previous version.

- What is the difference in predictability (e.g. coefficients of multivariate regression analyses, explained variance (R^2) of each model) of COVIDrelated health, economic, and social outcomes throughout the pandemic using both versions of the CCVI and the SVI?

surgo VENTURES

Validation practices implemented for the newest version of the CCVI

It is beyond the scope of this document to record all validation metrics using the newest version of the CCVI against negative COVID-related outcomes and in comparison to the previous version and the CDC's SVI. Validation analyses are ongoing and the work related to these will be published in the coming weeks and months after the initial release of the index. However, it will suffice to say that the newest version of the CCVI outperforms both the older version and the CDC's SVI when it comes to predicting negative COVID outcomes.

Convergent validation

Pearson correlation (r), which measures the linear relationship between two variables, was performed between the multiple versions of the CCVI and SVI in order to ascertain the level of agreement between the indexes. There is a high level of agreement between the latest version of the CCVI and the SVI ($r = 0.86$) as well as with the older version of the CCVI ($r = 0.87$). The high r-value indicates that we are not straying too far from the vulnerability that we captured in the previous version. However, we are adding a degree of variability that will be different from the prior release of the CCVI. This analysis does not inform us as to if the newer CCVI would perform better in terms of predicting negative COVID-related outcomes.

Construct or empirical validation

Empirical validation is the main form of validation analysis that we have chosen to focus on. Our main concern with the CCVI is that it is able to inform policy by predicting where the most negative impacts from COVID will occur across multiple dimensions of vulnerability. We compiled a plethora of negative outcome time-series data collected throughout the entire pandemic (since February 2020) from the Opportunity Insights Tracker that span health, economic, and social outcomes. We then ran regression analyses with the multiple versions of the CCVI and the SVI against this time-series of outcome data to see how the predictability of each index varied through the pandemic's evolution. It is beyond the scope of this document to delve into these analyses, but the current CCVI is able to explain the variance (r-squared values) of many of these outcomes to a higher degree than either the prior release of the CCVI and the SVI. The results of these analyses are still ongoing and will be released in various formats in the coming weeks and months.