



ORIGINAL ARTICLE

Vulnerability interactive geographic viewer against COVID-19 at the block level in Colombia: Analytical tool based on machine learning techniques

Oscar Espinosa^{1,2} | Jhonathan Rodríguez^{1,2} | Adriana Robayo¹ | Lelio Arias³ | Sandra Moreno³ | Mariana Ospina³ | David Insuasti⁴ | Juan Oviedo³

¹Instituto de Evaluación Tecnológica en Salud, Colombia

²Universidad Nacional de Colombia, Colombia

³Departamento Administrativo Nacional de Estadística, Colombia

⁴Departamento Nacional de Planeación, Colombia

Correspondence

Oscar Espinosa, Instituto de Evaluación Tecnológica en Salud and Universidad Nacional de Colombia, Carrera 49a # 91-91, Bogotá D.C., Colombia.
Email: oaepinosaa@unal.edu.co

Abstract

To mitigate the effects of the coronavirus disease 2019 (COVID-19) pandemic, different countries have developed computational tools and dashboards that generate value for decision-making in public health. We aimed to build an interactive geographic viewer for vulnerability to COVID-19 at the block level in Colombia to identify the location of populations that, because of sociodemographic characteristics and health conditions, could have more complications from COVID-19 infections. The vulnerability levels of the different blocks of 1,102 municipal capitals were calculated. Additionally, the institutions that provide health services and hotels were georeferenced, and changes in people's mobility dynamics in large cities were identified.

KEYWORDS

Colombia, COVID-19, geovisualization, machine learning, vulnerability

JEL CLASSIFICATION

I15, I18, C49



1 | INTRODUCTION

Since the first coronavirus disease 2019 (COVID-19) case recorded in December 2019 in the City of Wuhan in China, all countries in the world have faced a situation of major consequence for their health systems, especially their ability to deal with the rapidly spreading respiratory virus, which has a high fatality rate in older adults and people with preexisting medical comorbidities. COVID-19 was officially declared a pandemic by the World Health Organization (on March 11, 2020), and the number of cases continues to rise. Most countries have taken drastic decisions to contain the contagion, including the temporary closure of schools and universities, the prohibition of mass events, and mandatory quarantines for all their inhabitants, leaving only essential economic sectors open with very high biosecurity measures (World Health Organization, 2021).

The data revolution has created a growing volume of information and a diversity of methodologies under an ever-changing scenario in which the need for a solid evidence base for the decision-making process is highlighted. Also, the unfolding of the global pandemic has reconfigured a context in which factors such as precision, quality, and feasibility must be harmonized. Moreover, COVID-19 has transformed the socioeconomic outlook and will have long-lasting effects in various areas such as demand and consumption patterns, global value chains, employment dynamics, income distribution, social structures, and the environment (Baldwin & Mauro, 2020; Nicola et al., 2020). By understanding the broad scope of the crisis, the value of official statistics as a public asset has been highlighted; nonetheless, to achieve this, national statistical offices and agencies for health technology assessment must address technological, ethical, and communication challenges (data infrastructure, the integration of new sources such as geospatial information, and big data, among others). By doing so, countries will be able to have statistical information that meets the data needs of users but empowers them to create the enabling conditions for a policy dialogue under an economic reactivation context.

From the perspective of the Colombian executive branch, considering the social isolation measures adopted to limit the spread of COVID-19, one of the priorities was to identify the vulnerable populations that could benefit from state programs (whether they were cash transfer based or other).

Therefore, the objective of this study was to construct a geographic viewer for vulnerability to COVID-19 at the block level for Colombia. Vulnerability is understood as a concept that groups different aspects of health (social, demographic, and economic, among others), using data science techniques, to discern the location of populations that, due to their sociodemographic characteristics and health conditions, could have more complications if they are infected with COVID-19. The purpose of this analytical tool is to provide relevant information for public policy decision-making by the government. Thus, the research question was the following: What geographic areas at the block level in Colombia would be more vulnerable – due to health, economic, and sociodemographic effects – if their resident populations were to become infected with COVID-19?

The structure of the remainder of the article is as follows: Section 2 presents a summarized literature review. Section 3 shows the methodological design, the models used, and the data. Section 4 contains the results and model interpretation. Finally, Section 5 contains a discussion and Section 6 the conclusions.

2 | LITERATURE REVIEW

To contribute to an understanding of the dynamics of COVID-19, some countries have had the idea to create geographic viewers that combine quantitative cartographic elements with computer systems. Recent developments in web technologies have significantly changed the way geospatial information is extracted, refined, analyzed, transmitted, published, shared, and viewed in the health sciences.

Although at the international level no geographic viewer of vulnerability to COVID-19 deploys a degree of disaggregation at the block level for an entire country – our computational development is unique, pioneering,¹ innovative, and disruptive in its approach and spatial disaggregation – some interesting initiatives have been created.



Among these is the COVID-19 Pandemic Vulnerability Index dashboard (<https://covid19pvi.niehs.nih.gov>) developed by Marvel et al. (2021) for the United States, where county-level data display viral spread and which communities are at risk. The dashboard presents infection rates, vaccination rates, population density, and some health and environmental variables; each county's vulnerability profile is calculated using the Toxicological Prioritization Index (ToxPi) framework.

Also for the United States there is the computational tool of the COVID-19 Healthcare Coalition (2020), which built a dashboard at the level of zip code tabulation areas (<https://c19hcc.org/resource/vulnerable-population>), measuring vulnerability through three main dimensions: medical risks, health resource risks, and social risks (understood as social determinants of health).

Similarly, South Africa's official statistical agency built a South African COVID-19 Vulnerability Index Dashboard (<https://bit.ly/3eW5ysG>) at the municipality level (Statistics South Africa, 2020). Based on the Alkire–Foster method, a vulnerability index is calculated using data from the 2011 census and other indicators on an individual's participation in the labor market, the density of the home, access to health, and use of chronic medication, among other sociodemographic aspects.

In another example, the United Nations Population Fund created the COVID-19 Population Vulnerability Dashboard (<https://covid19-map.unfpa.org>), where indicators such as percentage of older people, density per household and per room, medical doctors per 10,000 people, nurses and midwives per 10,000 people, hospital beds per 1,000 people, and intensive care unit (ICU) beds per 100,000 people are shown (United Nations Population Fund, 2020).

Although other geographic viewers (or dashboards) have been developed, most are only for specific regions or provinces of a country, not for all of the administrative political divisions throughout the national territory. Likewise, there are studies that, although they do not generate interactive geovisualization, do calculate vulnerability indices with different levels of disaggregation and make use of various statistical techniques.

Among these, four developments are worth highlighting: (1) an investigation by Daras et al. (2021) for small areas in England through multivariate Poisson regression models; (2) the analysis of Acharya and Porwal (2020) for the different states and districts of India using the percentile ranking method; (3) the approximation of Tiwari et al. (2021) at the county level for the United States using random forest methodology; and (4) the research of Macharia et al. (2020) at the subnational level of Kenya, which applies standardization and weighting of multiple epidemiological and social indicators.

3 | METHOD

3.1 | Datasets

At the beginning of 2019, the National Administrative Department of Statistics (*Departamento Administrativo Nacional de Estadística*–DANE in Spanish) presented to the country the results of its largest statistical operation, the National Population and Housing Census (CNPV 2018 in Spanish or NPHC 2018 in English), which marked essential milestones such as the design and implementation of an electronic census, an innovation in the logistics domain by designing a 'routes method' that increases the coverage of the statistical operation, an ethnic approach, and the collection of personal identification variables. The latter would become one of the significant assets of the NPHC 2018 owing to the ease of integration with other statistical operations such as household surveys and administrative records. In addition to personal identification, the composition of households and places of residence immediately became the information elements in greatest demand when the pandemic began.

The availability and timing of these input sources allowed for a reinforced collaboration among DANE, the Institute for Health Technology Assessment (*Instituto de Evaluación Tecnológica en Salud*), the National Planning Department (*Departamento Nacional de Planeación*), and the Ministry of Health and Social Protection (*Ministerio de Salud y Protección Social*) in the characterization of the population that presented the most significant risk if infected



with COVID-19 and that would require a higher priority in the medical services of hospitalization and intensive care. To this end, we based our efforts on the sociodemographic characteristics of the deceased at the global level to identify patterns regarding age, sex, comorbidities, and residential conditions.

In this manner, all adults over 60 years of age were identified as susceptible people. Also, the preexistence of hypertension, diabetes, chronic respiratory diseases, ischemic heart disease, weak immune systems, and any type of cancer make up some of the individual characteristics most frequently presented by those who have died from COVID-19 (Albitar et al., 2020; Espinosa et al., 2020; Li et al., 2021; Noor & Islam, 2020; Williamson et al., 2020; Zhou et al., 2020).

Additionally, if a person had these characteristics, interaction with the groups with the greatest contagion could put them at greater risk. Therefore, thanks to household composition data provided by the NPHC 2018, it was possible to locate the elderly living with people 20–30 years of age (with a higher risk of contagion) and those who cohabited with people 30–50 years of age (with a medium risk of contagion).

Once the isolation measures were taken at the national and local level, older adults who live alone or with people who belong to their family nucleus were at risk because they had to necessarily expose themselves to be able to stock up on food. Consequently, interactions of any kind could increase the risk of contagion, so determining the places with the highest demographic density and overcrowded conditions in rooms and bedrooms could help policy-makers to plan and distribute resources.

This understanding and the availability of the NPHC 2018 data, the records of the Unique Database of Health Affiliation (*Base de Datos Única de Afiliación en Salud*–BDUA), and individual records of health service provision (*Registros Individuales de Prestación de Servicios de Salud*–RIPS), allowed us to construct, through deterministic integration (by identification document), 13 indicators calculated at the block level in all the municipal seats of the 1,102 municipalities in the country. According to the NPHC 2018, these municipalities represent the residences of 75.5% of the population, or approximately 36.5 million people. With this information, the next step consisted of determining the areas with the greatest vulnerability: the areas with the highest levels in the constructed indicators, which led to the vulnerability index.

This index aims to help policy-makers determine the quantity and location of people with the most significant susceptibility to the virus and, therefore, prepare for the health system's possible saturation. Through an interactive geovisualization platform, the index helps decision-makers and contributes to the generation of self-care awareness because the results are available to all people, where they can easily find their residence block and understand how they and their community could be affected if the virus spreads in their area.

3.2 | Model

As a proposal to identify blocks with possible high levels of vulnerability, a *k*-means cluster analysis was proposed (Hartigan & Wong, 1979; Trevor et al., 2008), which groups the blocks according to their sociodemographic and comorbidities characteristics. Subsequently, an index was built using the information from the centroids of each group that assigns levels of vulnerability. The steps followed to build the index were as follows:

- a. Determine the proportion of individuals per block with the following comorbidities, using the NPHC 2018 and the RIPS. The pathologies identified as risk factors that can generate complications in people with COVID-19, with their respective ICD-10 codes, are the following:
 - Hypertension (I10, I11, I12, I13, I15)
 - Diabetes (E10, E11, E12, E13, E14)
 - Ischemic heart disease (I20, I21, I22, I23, I24, I25)
 - Chronic pulmonary (J40, J41, J42, J43, J44, J45, J46, J47)
 - Cancer (C00–C97)



- b. Determine the proportion of households and people at the block level with the following indicators, using people registered in the NPHC 2018:
 - Individuals over 60 years of age
 - Homes with overcrowded rooms and bedrooms
 - Households with high and medium intergenerational risk² per block
- c. Determine the average number of people over 60 years of age per block who live in single and family households.³
- d. Determine the population density at the block level.
- e. Construct a database that has 407,277 rows that represent all the blocks in the seats of 1,102 municipalities of the country with the variables described above. The analysis concentrates on these seats and excludes the 20 nonmunicipalized areas of the country. The municipal capitals under analysis are grouped according to their population as follows:
 - Municipal seats with more than 1,000,000 inhabitants
 - Municipal seats with 100,000–999,999 inhabitants
 - Municipal seats with 20,000–99,999 inhabitants
 - Municipal seats with fewer than 19,999 inhabitants
- f. Apply the unsupervised learning technique *k*-means per municipal capital to determine the number of groups defined as the greater the number of inhabitants in a municipal seat, the greater the number of vulnerability categories:
 - Five groups⁴ for municipal seats with more than 1,000,000 inhabitants
 - Four groups for municipal seats of 100,000–1,000,000 inhabitants
 - Three groups for municipal seats of 20,000–100,000 inhabitants
 - Two groups for municipal seats with fewer than 20,000 inhabitants
- g. Determine the ordinality level⁵ of the selected clusters and thus assign a vulnerability level, using the centroids of each of the municipal capitals. In the first place, the maximum value of each of the variables in each centroid is obtained, a label is assigned, the number of maximums in each of the groups is counted, and, finally, it is established that the group with the greatest vulnerability is the group with the highest number of maximums; in the event of a tie, it is resolved randomly.

To determine the next level of vulnerability, the previously selected group is excluded. This procedure is repeated $k - 1$ times to determine the level of vulnerability of the k centroids of the header.
- h. Display the vulnerability level assigned to each of the 407,277 blocks in the analysis.

4 | RESULTS AND MODEL INTERPRETATION

After we conducted the analyses for large cities such as Bogotá, D.C., and Medellín, variables were filtered. The construction of the index was consolidated as described in the step-by-step of the previous section to unify the vulnerability assignment criterion in all the municipal seats.

For the final analysis, 13 variables of interest were used. Table 1 presents the values of the centroids that determined vulnerability thematically. It is important to emphasize that vulnerable groups do not represent contagion risks. As detailed, with these groups we considered sociodemographic variables (mostly older adults) and comorbidities, which allowed us to presume that, if a block has a higher prevalence in these variables, the proportion of the population that lives in it has a greater vulnerability in case of being infected with COVID-19. Mainly, for municipalities with fewer than 20,000 inhabitants, high vulnerability is only georeferencing the blocks with the highest prevalence of older adults.

The analysis of the 23 main municipalities in the country⁶ shows that 13 have more than 20% of their blocks with a medium-high or high level of vulnerability. The worrying situation of Popayán is evident, because it is a

**TABLE 1** Values of the centroids for the analyzed variables

Variables	Centroids				
	Group 0	Group 1	Group 2	Group 3	Group 4
Population density block ⁷	0.049	0.074	0.038	0.063	0.051
Proportion of persons 60 years of age and older	0.264	0.183	0.279	0.220	0.115
Proportion of households overcrowded in rooms	0.003	0.010	0.001	0.006	0.012
Proportion of households overcrowded in bedrooms	0.007	0.026	0.003	0.017	0.030
Proportion of households at high intergenerational risk	0.029	0.030	0.028	0.030	0.022
Proportion of households at medium intergenerational risk	0.063	0.057	0.058	0.063	0.039
Proportion of persons diagnosed with hypertension	0.179	0.155	0.173	0.171	0.103
Proportion of persons diagnosed with diabetes	0.053	0.050	0.048	0.054	0.033
Proportion of persons diagnosed with ischemic heart disease	0.025	0.018	0.024	0.021	0.011
Proportion of persons diagnosed with chronic lung diseases	0.032	0.032	0.033	0.034	0.021
Proportion of persons diagnosed with cancer	0.053	0.039	0.053	0.045	0.025
Over 60 years old in single-person households	21.41	3.14	55.96	8.40	0.36
Persons 60 years of age and older in non-family households	1.08	0.21	2.49	0.43	0.06
Number of maximums	3	1	4	3	2

⁷Ratio between the number of people actually registered and the area in square meters.

territory that has more than 30% of blocks in high vulnerability. On the other hand, Pasto and Armenia are the municipalities with the highest percentage of blocks with low vulnerability (Table 2).

The geospatial information platform, which can identify the vulnerability of the block to COVID-19 contagion (Figure 1), can be found at the following open access link: <https://geoportal.dane.gov.co/visor-vulnerabilidad>.

Layers, transparencies, and 3D views can be enabled, as shown in Figure 2. Within these layers, thanks to the collaboration of the United Nations Development Programme, the geographical mobility layer published and calculated by Grandata is available, which shows percentage differences of events carried out by mobile users outside the home, taking as the day of reference March 2, 2020 (United Nations Development Programme, 2020).

As a result, the layer allows users to visualize the high (or low) degree of compliance with the compulsory social isolation of a population in each urban sector. Positive values (represented with red colors) indicate high mobility in relation to March 2, 2020, while negative values, represented with blue colors, indicate low mobility compared with this same reference date. For this publication, the data are available for all municipal seats in Colombia.

5 | DISCUSSION AND POLICY RECOMMENDATION

Under the premise that everything that happens happens somewhere, the relation between the different natural and socioeconomic phenomena with the space in which these phenomena unfold results in an indissoluble link, without which it would be impossible to analyze and, therefore, understand the behavior of these phenomena from their distribution in geographical space, where the interactions between human beings and the surrounding environment are established.



TABLE 2 Block-level percentages according to vulnerability level by municipality

Municipalities	Vulnerability				
	Low	Middle-low	Middle	Middle-high	High
Armenia	60.27%	23.21%	6.44%	8.69%	1.40%
Barranquilla	53.33%	24.44%	12.93%	7.86%	1.44%
Bucaramanga	58.17%	27.93%	NA	11.50%	2.40%
Bogota	38.88%	31.65%	18.78%	6.80%	3.88%
Cali	48.81%	2.78%	33.85%	14.19%	0.37%
Cartagena	55.36%	23.98%	0.52%	18.93%	1.21%
Cúcuta	51.08%	24.08%	2.65%	18.91%	3.29%
Florencia	53.46%	24.16%	NA	17.47%	4.91%
Ibagué	52.88%	45.85%	NA	1.27%	NA
Manizales	49.77%	33.06%	3.86%	11.29%	2.03%
Medellín	40.81%	33.08%	0.87%	16.40%	8.84%
Montería	59.16%	9.62%	NA	24.34%	6.88%
Neiva	52.35%	36.59%	NA	9.08%	1.98%
Pasto	67.67%	1.89%	NA	20.62%	9.82%
Pereira	50.14%	23.30%	NA	12.65%	13.90%
Popayan	39.32%	4.02%	NA	26.22%	30.44%
Quibdo	52.39%	29.10%	NA	4.78%	13.74%
Riohacha	13.10%	55.69%	NA	28.02%	3.19%
Santa Marta	57.77%	24.52%	NA	15.34%	2.37%
Sincelejo	51.11%	10.40%	NA	23.47%	15.03%
Tunja	52.70%	1.66%	NA	32.61%	13.03%
Valledupar	54.92%	24.99%	NA	17.14%	2.95%
Villavicencio	48.20%	24.57%	NA	22.92%	4.31%

Depending on the level of geographical/territorial disaggregation, the dissemination of the results obtained from the statistical operations is generally done alphanumerically through tables and graphs that synthesize the information. Suppose the integration of statistical and geospatial information is added to the parameters above through the spatialization and visualization of statistical information through thematic maps. In that case, effective communication instruments are obtained for understanding trends and patterns of the different phenomena of interest. This disaggregation would allow users to explore, in a more focused way, the detailed spatial behavior of the different thematic variables constructed from the source information, guaranteeing, in turn, statistical privacy using the anonymization of the information.

Under these premises, we have developed an interactive geographic viewer that displays a vulnerability index of census source by block, which allows users to locate the population most likely to have complications if they become infected with COVID-19. The geographic viewer allows for visualization, through choropleth representations on the blocks of the municipal capitals of the country, different thematic variables of interest for decision-making focused on the current situation. These variables could include the multidimensional poverty index; groups by level of vulnerability; percentage of adults over 60 years; percentage of adults older than 70 years; location of medical centers and hotels; and the mobility index of the population.

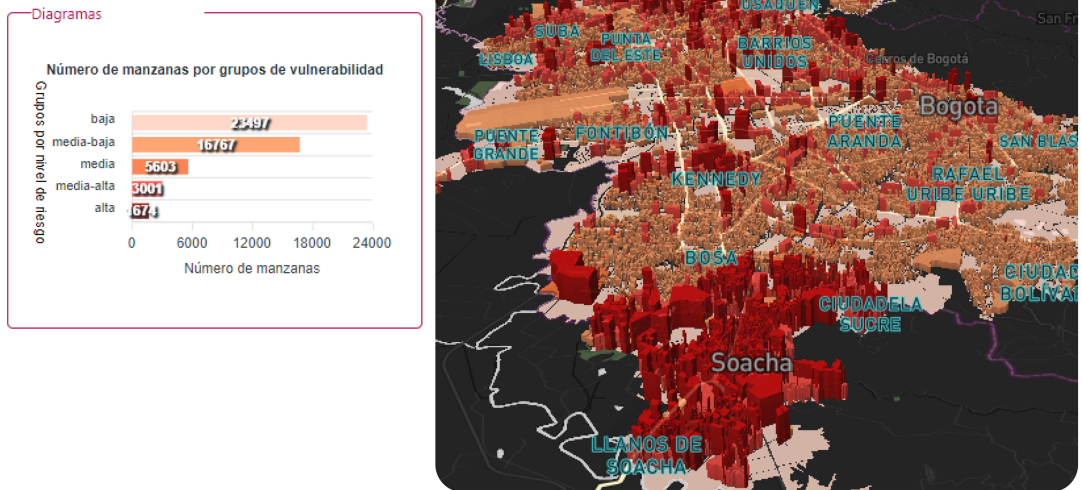


FIGURE 1 Geovisualization per block vulnerability (city: Bogotá, D.C.)

The provision of this tool has allowed different types of users to access the information strikingly and dynamically, as well as the possibility of making inquiries about sites of particular interest. In addition, the tool permits the downloading of information in geographic coverage format to those users interested in carrying out particular analyses.

Based on the geographic viewer presented in this study, we have identified the following public policy recommendations:

- Policy-makers must consider the design, construction, and continuous updating of this type of geographic information system as a relevant part of their decision-making instruments to develop monitoring strategies for different public health problems that feature different geographic behaviors.
- The complex interrelations among health events and socioeconomic and demographic conditions in a territory as diverse as Colombia make it necessary to spatially explore the impact of this type of communicable disease (from a comprehensive approach). This could serve to support decision-making on epidemiological surveillance, health promotion, and disease prevention.
- The developed geographic viewer offers useful information for prioritizing population segments that have not yet been vaccinated against COVID-19. This tool could be used to locate people who are highly vulnerable and have not yet been immunized, which would allow targeting and awareness campaigns aimed at the populations at greatest risk.
- From the mobility layer, information can be obtained that is useful for the issuance of policies on the authorization and restriction of rules on passenger mobility within cities.
- From the layers of the multidimensional poverty index, showing adults over 60 and over 70 years of age, it is possible to contribute to the geolocation at the block level of people belonging to the old age life cycle who have economic needs and require government support.
- From the layer of installed capacity, health care centers, hospitals, and hotels, it is possible to generate contingency plans to provide sufficient infrastructure within areas of greatest vulnerability.

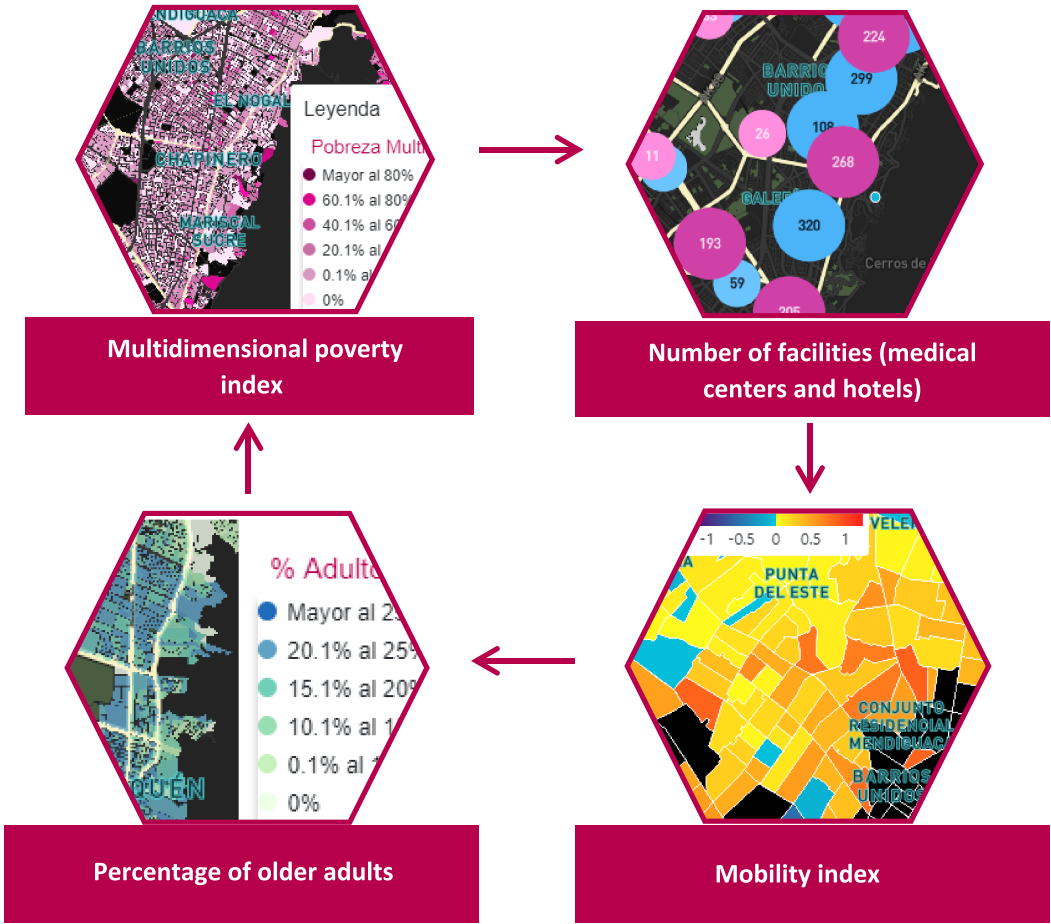


FIGURE 2 Different layers of the per-block geographic viewer

6 | CONCLUSION

In this study, levels of vulnerability to COVID-19 for individuals residing within the different blocks of 1,102 municipal capitals were calculated for Colombia. Additionally, the institutions that provide health services and hotels in these areas were georeferenced, and changes in people’s mobility dynamics in large cities were identified. This computational tool will help the executive branch to identify vulnerable populations for assistance under different state programs.

Although the vulnerability index was useful to raise awareness among citizens to abide by the existing social isolation measures and to provide information for the national and local authorities to determine how to better allocate budget support to health care centers and hospitals, due to its prompt availability and ease of interpretation, some of the current challenges include the insertion of new indicators and the need to reduce the publication lag of the index to help plan and mitigate the impact of the pandemic in the new normality.

Two possible limitations of this study are that the records of diagnoses of disease conditions used in the analysis refer to the population with effective coverage under the Colombian health system, so it is not possible to take into account the real incidence and prevalence of chronic disease (however, such a bias is thought to be limited) and that people were able to change their place of residence from that registered in the NPHC of 2018. As a future line of



work, we expect to update the COVID-19 geographic vulnerability viewer with more recent data from clinical records and the inclusion of new disease conditions that have proved to be risk factors for COVID-19.

Dealing with these challenges requires establishing adaptative cultures as an opportunity to strengthen coordination spaces and mechanisms in the statistical community. Thus, the COVID-19 pandemic can be seen as a window of opportunity for innovation from the statistical perspective, as it has fostered the strengthening of traditional partnerships and highlighted the value of statistics as a public asset for the well-being of humanity.

ACKNOWLEDGMENTS

We are grateful to participants at the VIII International Monetary Fund Statistical Forum, the XL Health Economics Conference, the IX International Congress of Ethical Economics, the VI ISPOR Colombia Congress, and the VII Congress of the Colombian Association of Health Economics for comments and suggestions.

ENDNOTES

- ¹ With an official launch on April 15, 2020, it was one of the first such computational tools worldwide (https://www.dane.gov.co/files/comunicados/Comunicado_visor_niveles_vulnerabilidad_covid19_Abril-15-2020.pdf).
- ² High risk: generational households consisting of adults over 60 years of age and population in the age group with the greatest contagion (20–29 years); medium risk: generational households made up of adults over 60 years of age and population in the second most contagious age group (30–59 years).
- ³ Elderly individuals living in a home with other people who are not their relatives.
- ⁴ Number of groups selected applying the elbow technique in the cities of Bogotá, D.C., and Medellín (Trevor et al., 2008).
- ⁵ By not having an index at the block level, the idea presented in DeCaprio et al. (2020) is taken from the index built at the person level.
- ⁶ The 23 municipalities presented are those used by DANE to estimate the country's unemployment rate.

DATA AVAILABILITY STATEMENT

https://geoportal.dane.gov.co/descargas/vulnerabilidad/VULNRB_IPMxMZ.rar

ORCID

Oscar Espinosa  <https://orcid.org/0000-0003-4893-0509>

Jhonathan Rodríguez  <https://orcid.org/0000-0002-3599-6595>

Adriana Robayo  <https://orcid.org/0000-0001-6895-4992>

Lelio Arias  <https://orcid.org/0000-0003-4706-2244>

Sandra Moreno  <https://orcid.org/0000-0003-3469-7785>

Mariana Ospina  <https://orcid.org/0000-0001-5722-4767>

David Insuasti  <https://orcid.org/0000-0002-3498-4576>

Juan Oviedo  <https://orcid.org/0000-0001-6619-2891>

REFERENCES

- Acharya, R., & Porwal, A. (2020). A vulnerability index for the management of and response to the COVID-19 epidemic in India: An ecological study. *The Lancet Global Health*, 8(9), e1142–e1151. [https://doi.org/10.1016/S2214-109X\(20\)30300-4](https://doi.org/10.1016/S2214-109X(20)30300-4)
- Africa, S. S. (2020). South African COVID-19 vulnerability index. Pretoria.
- Albitar, O., Ballouze, R., Ooi, J., & Sheikh Ghadzi, S. (2020). Risk factors for mortality among COVID-19 patients. *Diabetes Research and Clinical Practice*, 166, 108293. <https://doi.org/10.1016/j.diabres.2020.108293>
- Baldwin, R., & Mauro, B. (2020). *Economics in the time of COVID-19*. London: CEPR Press.
- COVID-19 Healthcare Coalition. (2020). Tracking vulnerable population by region. Retrieved July 26, 2021, from <https://c19hcc.org/resource/vulnerable-population>



- Daras, K., Alexiou, A., Rose, T., Buchan, I., Taylor-Robinson, D., & Barr, B. (2021). How does vulnerability to COVID-19 vary between communities in England? Developing a small area vulnerability index (SAVI). *Journal of Epidemiology and Community Health*, 75(8), 729–734. <https://doi.org/10.1136/jech-2020-215227>
- DeCaprio, D., Gartner, J., McCall, C., Burgess, T., Garcia, K., Kothari, S., & Sayed, S. (2020). Building a COVID-19 vulnerability index. *Journal of Medical Artificial Intelligence*, 3, 15–15. <https://doi.org/10.21037/jmai-20-47>
- Espinosa, O., Dos Santos-Zanetti, A., Fornaciari-Antunes, E., Gulin-Longhi, F., De Matos, T., & Battaglini, P. (2020). Prevalence of comorbidities in patients and mortality cases affected by SARS-CoV2: A systematic review and meta-analysis. *Revista do Instituto de Medicina Tropical de São Paulo*, 62, e43. <https://doi.org/10.1590/s1678-9946202062043>
- Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28(1), 100. <https://doi.org/10.2307/2346830>
- Li, J., Huang, D., Zou, B., Yang, H., Hui, W., Rui, F., Yee, N. T. S., Liu, C., Nerurkar, S. N., Kai, J. C. Y., Teng, M. L. P., Li, X., Zeng, H., Borghi, J. A., Henry, L., Cheung, R., & Nguyen, M. H. (2021). Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. *Journal of Medical Virology*, 93(3), 1449–1458. <https://doi.org/10.1002/jmv.26424>
- Macharia, P., Joseph, N., & Okiro, E. (2020). A vulnerability index for COVID-19: Spatial analysis at the subnational level in Kenya. *BMJ Global Health*, 5(8), e003014. <https://doi.org/10.1136/bmjgh-2020-003014>
- Marvel, S., House, J., Wheeler, M., Song, K., Zhou, Y.-H., Wright, F., Chiu, W. A., Rusyn, I., Motsinger-Reif, A., & Reif, D. M. (2021). The COVID-19 pandemic vulnerability index (PVI) dashboard: Monitoring county-level vulnerability using visualization, statistical modeling, and machine learning. *Environmental Health Perspectives*, 129(1), 017701. <https://doi.org/10.1289/EHP8690>
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193. <https://doi.org/10.1016/j.ijsu.2020.04.018>
- Noor, F., & Islam, M. (2020). Prevalence and associated risk factors of mortality among COVID-19 patients: A meta-analysis. *Journal of Community Health*, 45(6), 1270–1282. <https://doi.org/10.1007/s10900-020-00920-x>
- Tiwari, A., Dadhania, A., Ragunathrao, V., & Oliveira, E. (2021). Using machine learning to develop a novel COVID-19 vulnerability index (C19VI). *Science of the Total Environment*, 773, 145650. <https://doi.org/10.1016/j.scitotenv.2021.145650>
- Trevor, H., Robert, T., & Jerome, F. (2008). *The elements of statistical learning* (2nd ed.). Stanford: Springer.
- United Nations Development Programme. (2020). UNDP GRANDATA. Retrieved February 28, 2021, from <https://covid.grandata.com/methodology>
- United Nations Population Fund. (2020). COVID-19 population vulnerability dashboard. Retrieved July 25, 2021, from <https://covid19-map.unfpa.org/>
- Williamson, E., Walker, A., Bhaskaran, K., Bacon, S., Bates, C., Morton, C., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R., ... Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 584(7821), 430–436. <https://doi.org/10.1038/s41586-020-2521-4>
- World Health Organization. (2021). Coronavirus disease (COVID-19) pandemic. Retrieved March 17, 2021, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet*, 395(10229), 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)

How to cite this article: Espinosa, O., Rodríguez, J., Robayo, A., Arias, L., Moreno, S., Ospina, M., Insuasti, D., & Oviedo, J. (2021). Vulnerability interactive geographic viewer against COVID-19 at the block level in Colombia: Analytical tool based on machine learning techniques. *Regional Science Policy & Practice*, 13(S1), 187–197. <https://doi.org/10.1111/rsp3.12469>