Developing a Data-Driven System for Identifying the Most Vulnerable Communities in Puerto Rico

Katherine Ann Willyard
James Scurry
Heather King
Benjamin Gurrentz
Bethany DeSalvo
Social Economic and Housing Statistics Branch
U.S. Census Bureau

SEHSD Working Paper 2023-19
June 7, 2023

*This working paper is released to inform interested parties of ongoing research and to encourage discussion. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The U.S. Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. DRB #CBDRB-FY23-SEHSD012-004.

Abstract

By helping communities better anticipate, respond, resist, and recover from disasters, social vulnerability mapping strengthens community resilience. This paper explains common issues indices have when identifying socially vulnerable communities in Puerto Rico and how the U.S. Census Bureau's Community Resilience Estimates for Puerto Rico (CRE-PR) overcome these concerns. We demonstrate how small area modeling produces more precise and timely measures of the communities most and least vulnerable to disasters. The vulnerability of three groups in Puerto Rico are described: the total population, the population in the most vulnerable tracts, and the population in the least vulnerable tracts. We also analyze the similarities and differences in areas identified as the most vulnerable by CRE-PR and another commonly used index. Decision-makers and disaster planners may benefit from CRE-PR, which provides estimates of vulnerability that permit statistical comparisons, to make data-driven decisions that increase the resilience of underserved communities in Puerto Rico.

Acknowledgements

We would like to thank Census statistical, fact checking and subject matter expert review team for their contributions: Ryan Janicki, Yerís H. Mayol-García, and Thomas Welton.

Suggested Citation

K. Willyard, Scurry J., King H., Gurrentz B., DeSalvo B., "Developing a Data-Driven System for Identifying the Most Vulnerable Communities in Puerto Rico," *SEHSD Working Paper Series*, 2023-17, U.S. Census Bureau, Washington, D.C., 2023.

Developing a Data-Driven System for Identifying
the Most Vulnerable Communities in Puerto Rico

1 | INTRODUCTION

Accurate social vulnerability mapping can increase resilience by helping communities better prepare for, respond to and recover from disasters.[1] Social vulnerability is the risk of hazards to the physical and socially built environment, while community resilience is the capacity of individuals and households to absorb the stresses from a disaster.[2] Because most of its population lives in coastal municipalities, Puerto Rico is especially vulnerable to hurricanes.[3] To increase the resilience of Puerto Rico, accurate and timely social vulnerability mapping is needed.

This paper explains common issues indices have when identifying socially vulnerable communities in Puerto Rico and how the U.S. Census Bureau's Community Resilience Estimates for Puerto Rico (CRE-PR) provides an enhanced method of identifying communities most vulnerable to a disaster. Through modeling and using auxiliary data sources, CRE-PR enhances survey estimates and reduces margins of error to produce more precise and timely estimates that can be used to make statistical comparisons across space or over time.

2 | THE NEED TO UTILIZE MICRODATA

The Puerto Rico Community Survey (PRCS) is part of the U.S. Census Bureau's American Community Survey (ACS). PRCS is the largest household survey in Puerto Rico, providing comprehensive information on population, social, economic and housing characteristics. Because of its comprehensiveness,[4] detailed geographic-level of information[5] and compliance to high statistical quality standards,[6] many measures of social vulnerability rely on publicly available PRCS 5-year direct domain population estimates.[7]

---

[1] Source: S. Van Zandt, W. G. Peacock, D. W. Henry, H. Grover, W. Highfield, S. Brody, *Mapping Social Vulnerability to Enhance Housing and Neighborhood Resilience* (2012; https://doi.org/10.1080/10511482.2011.624528)

[2] Source: J. H. Masterson, W. G. Peacock, S. Van Zandt, H. Grover, L.F. Schwarz, J. Cooper, *Planning for Community Resilience: A Handbook for Reducing Vulnerability to Disasters* (2014; https://link.springer.com/book/10.5822/978-1-61091-586-1)

[3] Source: U.S. National Oceanic and Atmospheric Administration, *Making a Difference in Puerto Rico* (2023; https://coast.noaa.gov/states/puerto-rico.html)

[4] Source: U.S. Census Bureau, *Understanding the PRCS: The Basics* (2020; https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_prcs_handbook_2020_ch01.pdf)

[5] Source: U.S. Census Bureau, *Geographic Areas Covered in the PRCS* (2020; https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_prcs_handbook_2020_ch03.pdf)

[6] Source: U. S. Census Bureau, *Statistical Quality Standard D1: Producing Direct Estimates from Samples* (2021; https://www.census.gov/about/policies/quality/standards/standardd1.html)

[7] Source: J. K. Summers, L. Smith L, L. C. Harwell, K. D. Buck, *Conceptualizing Holistic Community Resilience to Climate Events: Foundation for a Climate Resilience Screening Index* (2017; https://doi.org/10.1002/2016GH000047).

A direct domain estimate approach to survey data involves a sampling design used to develop a sampling frame from the population with a probability, which depends on known design variables such as stratum indicator variables and size measures of clusters. A domain, U_i, is a subset of the target population, Y_i , such that $Y = \sum_{j \in U} y_j$. For instance, a domain can be considered the total number of individuals living within a Census tract without health insurance. To make population inferences, the values associated with the selected sample, y_i, are observed and summed up over all sample elements, s. PRCS's multiyear estimation methodology involves pooling data across years and adjusting base weights, ω_j, using generalized regression estimation (GREG) techniques and auxiliary demographic information on total population by age, sex and race/ethnicity from the U.S. Census Bureau Populations Estimates Program (PEP).[8] The detailed notation of a GREG estimation is as follows:

$$\hat{Y}^*_{iGR} = \sum_{j \in s} \omega^*_{ij} y_{ij},$$

where $\omega^*_{ij} = \omega_j g^*_{ij}$

with $g^*_{ij} = 1 + (X_i - \hat{X}_i)^T \left(\sum_{j \in s} \omega_j x_{ij} x^T_{ij} / c_j \right)^{-1} x_{ij}/c_j$

PRCS goes through substantive efforts to identify, reduce, and measure error.[9] PRCS estimates meet the U.S. Census Bureaus' high quality statistical standards which include the publication of sampling error.[10] Sampling error is the uncertainty that comes from the fact that a survey is based on a sample, rather than all housing units or individuals. The amount of error is directly related to the size of the sample, as well as the variability. In order to make comparisons and develop overall rankings of survey data, the sampling error should be incorporated.[11]

Common indices do not incorporate survey error measures when creating estimates, nor do they provide estimates of error.[12,13,14] One of the most prevalently used social vulnerability index is the Centers for Disease Control and Prevention's Social Vulnerability Index (CDC-

---

[8] Source: U.S. Census Bureau, *American Community Survey and Puerto Rico Community Survey Design and Methodology* (2022; https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/2022/acs_design_methodology_report_2022.pdf)

[9] Source: U. S. Census Bureau, *Understanding Error and Determining Statistical Significance* (2020; https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_general_handbook_2020_ch07.pdf).

[10] Source: L. Bowers, W. Basel, D. Powers, *Evaluating 2012-2014 Trends in Health Insurance Coverage for All U.S. Counties* (2016; https://www.census.gov/library/working-papers/2016/demo/SEHSD-WP2016-16.html).

[11] Source; M. Klein, T. Wright, J. Jerzy Wieczorek A joint confidence region for an overall ranking of populations (2020; https://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/KleinWrightWieczorek2020.pdf)

[12] Source: B. Flanagan, E. Hallisey, *The Social Vulnerability Index and Toolkit* (2013; https://svi.cdc.gov/Documents/Publications/CDC_ATSDR_SVI_Materials/SVI_30April2013.pdf).

[13] Source: Hazards and Vulnerability Research Institute, *The SoVI® Recipe* (2016; https://www.sc.edu/study/colleges_schools/artsandsciences/centers_and_institutes/hvri/documents/sovi/sovi_recipe_2016.pdf).

[14] Source: U.S. Environmental Protection Agency, *EJSCREEN Technical Documentation* (2019; https://www.epa.gov/sites/default/files/2021-04/documents/ejscreen_technical_document.pdf).

SVI). CDC-SVI relies upon publicly available PRCS 5-year direct survey estimates and area-level aggregations of vulnerability indicators through percentile ranking methods of survey estimates as true parameter values. For Puerto Rico, CDC-SVI produces an overall vulnerability score and scores for three specific themes: (1) socioeconomic status, (2) household characteristics, and (3) housing type and transportation. Each of the themes are based on the treatment of survey estimates as true parameter values to create percentile ranks, rather than survey estimates subject to sampling error. For example, the socioeconomic status theme uses five PRCS survey estimates: (1) percentage of persons below 150% poverty estimate, (2) unemployment rate estimate, (3) percentage of housing cost-burdened occupied housing units with annual income less than $75,000 estimate, (4) percentage of persons age 25+ with no high school diploma estimate, and (5) Percentage uninsured in the total civilian noninstitutionalized population estimate. Each of these survey estimates are treated as true percentile ranks.

To demonstrate statistical problems, we describe CDC-SVI 2020 methods to calculate the vulnerability of two tracts in Puerto Rico: Census Tract 1504 in Fajardo Municipio and Census Tract 9802 in San Juan. For the socioeconomic status theme unemployment indicator, Census Tract 1504 in Fajardo Municipio has a 2016-2020 PRCS unemployment rate of 29.5% (+/-7.4%) that is higher than the for all Puerto Rico (15.1% and an error of +/-0.3%) but, this tract is not flagged in CDC-SVI as vulnerable. The CDC-SVI 2020 index treats values as true parameter values, so this tract's unemployment rate is treated as only in the top 89.8% of estimates. On the other hand, the tract with the highest unemployment percentile ranking score, Census Tract 9802 in San Juan, has an unemployment rate of 100% (+/- 80.2%) that is not statistically different from either the Puerto Rico average or Census Tract 1504 in Fajardo Municipio, but is flagged as vulnerable. In the CDC-SVI 2020, after percentile ranks are obtained for each indicator, they are aggregated by theme and overall to create theme specific and overall percentile ranking values, again, by treating the values as true parameter values rather than survey estimates subject to sampling error. Finally, areas are flagged as vulnerable if the theme specific or overall percentile ranking value is within the 90th percentile.

There is a statistical process involved in ranking populations based on sample survey data.[15] While PRCS describes a process from comparing sample survey estimates using margins of error,[16] many indices treat survey estimates as true parameter values rather than survey estimates subject to sampling error.[11,12,13] To correctly interpret an estimate, the margin of error should be incorporated.[4] Since indices that rely on publicly available data don't utilize the margins of error, they don't reflect the statistical intricacies of survey data in their analysis. In addition, indices that don't produce margins of error along with their estimates are not as practically useful as those that do. Without the production of margins of error along with estimates, a statistically significant difference between places or across time cannot be found. This means that they cannot be used for accurate data driven decision making.

Communities are being overloaded with information about the vulnerability of populations that is often contradicting. One index might declare a place resilient, and another might declare it vulnerable. Because other indices do not correctly statistically rank survey data

---

[15] Source: T. Wright, M. Klein, J. Wieczorek, R*anking Populations Based on Sample Survey Data* (2014; https://www.census.gov/library/working-papers/2014/adrm/rrs2014-12.html).

[16] Source: *U. S. Census Bureau, Making Comparisons with ACS Data* (2020; https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_general_handbook_2020_ch04.pdf).

and they often make statistically inaccurate claims of difference. Having multiple estimates of population vulnerability without estimates of error makes it difficult for communities to understand data and act accordingly. Since Puerto Rico has a population that changes dramatically after natural disasters, like after Hurricane Maria,[17] but still faces the same risk of experiencing a natural disaster, timely and accurate estimates of vulnerable populations are needed.

3 | COMMUNITY RESILIENCE ESTIMATES FOR PUERTO RICO

Due to increased need for timely and consistent measures of vulnerability due to the COVID-19 pandemic, 2019 Community Resilience Estimates (CRE) for the United States was first released in July 2021[18] using ACS 1-year survey estimates combined with auxiliary population data along with established U.S Census Bureau small area estimation methods. CRE are modeled estimates of social vulnerability in the population, based on the number of vulnerability indicators individuals within the population have.[19] Working with local technical and subject matter experts, CRE methods were then revised to develop CRE-PR in May 2023.

CRE-PR is created by first tagging vulnerability indicators within PRCS microdata. The following ten vulnerability indicators are used in 2019 CRE-PR: (1) households with an income-to-poverty ratio less than 130 percent, (2) Only one or no individuals living in the household are aged 18-64, (3) Household crowding defined as more than 0.75 persons per room, (4) Household with a limited education defined households where no one over the age of 16 has a high school diploma, (5) No one in the household is employed full-time, year-round, but the flag is not applied if all residents of the household are aged 65 or older, (6) Individual with a disability posing a constraint to significant life activity, (7) Individual with no health insurance, (8) Individual aged 65 or older, (9) Household without a vehicle, and (10) Households without broadband internet access.

CRE-PR methods slightly differ from CRE methods. The two key modifications are as follows: (1) While 2019 CRE for the United States tags communication barrier households defined as those with either limited English-speaking households or households with a limited education defined households where no one over the age of 16 has a high school diploma, CRE-PR only uses education status for the flag. This is because Spanish, not English is the

[17] Source: J. Schachter, A. Bruce, *Estimating Puerto Rico's Population After Hurricane Maria* (2020; https://www.census.gov/library/stories/2020/08/estimating-puerto-rico-population-after-hurricane-maria.html) e

[18] Source: R. C. Sawyer, B. DeSalvo, *New Census Bureau Tool Will Now Consistently Update Communities on Their Vulnerable Populations* (2021; https://www.census.gov/library/stories/2021/09/measuring-communities-resilience-in-the-face-of-adversity.html#:~:text=New%20Census%20Bureau%20Tool%20Will%20Now%20Consistently%20Update%20Communities%20on%20Their%20Vulnerable%20Populations&text=The%20U.S.%20Census%20Bureau's%20Community,with%20disasters%20and%20other%20emergencieshttps://www.census.gov/library/stories/2019/09/measuring-communities-resilience-in-the-face-of-adversity.html).

[19] Source: U.S. Census Bureau, *Community Resilience Estimates* (2020; https://www.census.gov/programs-surveys/community-resilience-estimates.html).

predominant language in Puerto Rico.[20] So, households with limited English-speaking abilities do not experience the same vulnerabilities as those located in the United States; and (2) Race and ethnicity were managed differently for CRE-PR. Since intercensal population estimates for Puerto Rico are not broken down by race/ethnicity, race is imputed at the tract-level using 2010 Decennial Census population counts. Also, CRE-PR modeling was done by age and race instead of age and race/ethnicity. The following four racial categories were used: White alone, Black alone, two or more races, and any other race.[21]

Individuals within the PRCS microdata are then described as low-risk (0 vulnerability indicators), moderate-risk (1-2 vulnerability indicators), or high-risk (3 or more vulnerability indicators). Next, using traditional direct survey methods, tabulations for states, municipios, and tracts for the number of people at low-, moderate-, and high-risk for different age, race and assigned urban concentration index bin by Census division categories are estimated. These traditional direct survey estimates are then used to inform the small area model.

When a domain has an insufficient sample-size to make direct survey estimates of adequate precision, it is considered a small area.[22] By combining survey data with auxiliary data, through small area modeling techniques, survey data can "borrow strength" from the additional information to make more precise estimates.[23] Small area estimation methods can enhance survey estimates to make more precise estimates than direct survey estimation techniques alone. For example, in comparison to 2005 ACS 1-year direct survey estimates of county poverty, the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program produced a 56 percent decline in standard error over all counties and gains were the greatest among counties with smaller ACS sample sizes.[24]

CRE-PR follows an area-level approach from small area estimation: a direct survey estimate is averaged with an indirect estimate to produce a composite estimate. The average is a weighted average, and the estimates are less volatile than either of the two original estimates alone. Here, direct estimates refer to PRCS estimates for the number of people at low-, moderate-, and high-risk as described in the paragraph above. CRE-PR fits and empirically optimal shrinkage model through post-stratification. The indirect, or synthetic, estimates for the number of people at low-, moderate-, and high-risk at the tract-level are developed from applying modelled proportions to auxiliary population data from U.S. Census Bureau's Population Estimates Program (PEP). Variances for direct survey estimates are smoothed using a

[20] Source: J. Velez. Understanding Spanish-language maintenance in Puerto Rico: A Political Will Meets the Demographic Imperative (2009; https://www.degruyter.com/document/doi/10.1515/ijsl.2000.142.5/html)

[21] Source: U.S. Census Bureau, "Census 2020 State Profile: Puerto Rico" (2021; https://www.census.gov/library/stories/state-by-state/puerto-rico-population-change-between-census-decade.html)

[22] Source: National Center for Science and Engineering Statistics, *National Patterns of R&D Resources: Future Directions for Content and Methods* (2013; https://nap.nationalacademies.org/catalog/18317/national-patterns-of-rd-resources-future-directions-for-content-and)

[23] Source: J. N. K. Rao, I. Molina, *Small Area Estimation* (2015; https://onlinelibrary.wiley.com/doi/book/10.1002/9781118735855)

[24] Source: W. Bell, W. Basel, C. Cruse, L. Dalzell, J. Maples, B. O'Hara, D. Powers, "Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties" (2007; https://www.census.gov/library/working-papers/2007/demo/bell-01.html)

generalized variance function (GVF). The weight given to an indirect estimate when producing the composite estimate is the ratio of the GVF variance of the direct estimate to the total variance (i.e., the sum of the GVF variance and the estimated variance of the indirect estimate). The weight for the direct estimate is the complement, (i.e., one minus the weight for the indirect estimate). As a result, when survey methods are more precise, the direct survey estimate receives a greater weight; when direct survey methods are less precise, the indirect modeled estimate receives a greater weight. This allows CRE-PR to produce reliable estimates of the number of people in each tract that are low-, moderate-, or high-risk.

The detailed notation of the composite estimator at the tract level is as follows:

$$\begin{aligned} w_{t,g} &= \frac{\hat{v}_{t,g}}{\hat{v}_{t,g} + \widehat{MSE}_{t,g}} \\ \tilde{\theta}_{t,g} &= w_{t,g} r_{t,g} + \left(1 - w_{t,g}\right)\hat{R}_{t,g}. \end{aligned}$$

For each tract t and vulnerability group g (low-, moderate-, high-), where:

$w_{t,g}$ = shrinkage weight

$\hat{v}_{t,g}$ = GVF-estimated sampling variance

$\widehat{MSE}_{t,g}$ = estimated mean square error (i.e., model variance)

$\tilde{\theta}_{t,g}$= composite estimate

$\hat{R}_{t,g}$ = direct survey estimate

$r_{t,g}$ = indirect (model) estimate

CRE creates more precise estimates than direct survey estimates alone. **Fig. 1** describes the amount that the relative error of 2019 PRCS direct estimates of the high-risk population are reduced through the small area modeling techniques employed to create 2019 CRE-PR. In comparison to 2019 PRCS direct estimates, on average, small area modeling reduces the coefficient of variation of high-risk population estimates by 37 percent. Small area modeling reduces the relative error of all estimates. While for 11 tracts, the relative error of estimates is cut by less than 20 percent, for 227 tracts, the relative error of estimates is cut by at least by 40 percent.

4 | METHOD

This paper uses 2019 CRE-PR as a case study to demonstrate how small area estimates of high-risk populations can be used to make statistical comparisons in the vulnerability of populations in Puerto Rico. We describe the high-risk population across Puerto Rico and the population located in the most vulnerable communities (i.e., tracts with high-risk population rates above the Puerto Rico average using a t-test with a 90% confidence interval).

Then we analyze similarities and differences in communities identified as vulnerable using 2019 CRE-PR and 2020 CDC-SVI. First, we explore the correlations between CRE-PR high risk population rates for tracts in Puerto Rico to CDC-SVI vulnerability measures. Second, we map the differences in areas tagged as most and least vulnerable. Then, using f-tests and t-tests with a 90% confidence interval, we determine: (1) for areas identified as most vulnerable by CDC-SVI but not CRE-PR, if the amount that the relative error of 2019 PRCS direct estimates of the high-risk population are reduced through the small area modeling techniques employed to create 2019 CRE-PR, and (2) for areas similarly tagged as most vulnerable by CRE-PR and CDC-SVI, if the relative error of 2019 PRCS direct estimates of the high-risk population are reduced through the small area modeling techniques employed to create 2019 CRE-PR. We suspect the average reduction in relative error to be higher for areas tagged as most vulnerable by CDC-SVI overall themes but not CRE-PR. This is because index methods, like the CDC-SVI, tend to be skewed by areas with high sampling error that are subject to the greatest improvement through small area estimation methods.[25] On the contrary, we suspect the average reduction in relative error to be lower for areas tagged as most vulnerable by both CDC-SVI overall themes and CRE-PR.

5 | RESULTS

Summary statistics on low-, moderate-, and high-risk populations across Puerto Rico and in tracts above (i.e., most vulnerable tracts) or below (i.e., least vulnerable tracts) the Puerto Rico average are described in **Table 1**. Across Puerto Rico, 46% (+/-1%) of the population are high-risk, 39% (+/-1%) are moderate-risk and 15% (+/- 2%) are low-risk. While 317,496 people live in tracts that are less vulnerable (i.e., tracts with a high-risk population rate below the Puerto Rico average), 253,341 people live in tracts that are more vulnerable (i.e., tracts with a high-risk population rate above the Puerto Rico average).

A map of tract high-risk population rates is presented in **Figure 2**. The Eastern coastline has a greater concentration of tracts where less than half of the population is high-risk. Alternatively, the Western coastline has a greater concentration of tracts where more than half of the population is high-risk. A map of most and least vulnerable tracts across Puerto Rico is presented in **Figure 3**. Least vulnerable tracts in Puerto Rico are concentrated in the affluent suburbs of San Juan. On the other hand, the most vulnerable tracts are concentrated in the Western half of Puerto Rico.

Correlations between tract CRE-PR high-risk population rates and CDC-SVI theme and overall scores are described in **Table 2**. CRE-PR is moderately correlated with the CDC-SVI overall vulnerability score (0.52), socioeconomic status vulnerability score (0.49) and household characteristics (0.46) but has little correlation with the housing type and transportation vulnerability score (0.21). Within CDC-SVI themes, because the overall vulnerability score is an aggregate of each of its components, the overall vulnerability score is highly correlated with each of its components.

Similarities and differences between tracts described as most vulnerable by both CRE-PR and CDC-SVI are presented in **Figure 4**. The greatest similarity between areas tagged as most vulnerable with CRE-PR and CDC-SVI is the overall vulnerability score. But even then, only 25 tracts are tagged as most vulnerable in both, while 63 tracts are most vulnerable in CDC-SVI but not CRE-PR and 60 are most vulnerable in CRE-PR but not CDC-SVI.

[25] Source: K. Willyard, G. Amaro, R. C. Sawyer, B. DeSalvo, W. Basel, *An Evaluation of Social Vulnerability and Community Resilience Indices and Opportunities for Improvement through Community Resilience Estimates* (2022; https://www.census.gov/content/dam/Census/library/working-papers/2022/demo/sehsd-wp2022-25.pdf)

Small area modeling techniques provide a significant improvement in reducing the relative error of a population’s estimated risk, however, this improvement was not uniform across all places. Areas tagged as most vulnerable by CDC-SVI but not CRE-PR are more ripe for improvement, while areas tagged as most vulnerable by both are less so. Differences in the amount of relative error reduced through small area modeling between areas identified as most vulnerable by CDC-SVI overall themes but not CRE-PR are described in **Table 3**. We find the relative error is higher for areas tagged as most vulnerable by CDC-SVI overall themes but not CRE-PR. On average with a 90% confidence interval, the average relative error reduced through small area modeling for areas identified as most vulnerable by CDC-SVI overall theme score but not by CRE-PR is 43.54% - 53.13% while all other tracts are reduced on average by 35.50% - 37.60%.

Differences in the amount of relative error reduced through small area modeling between areas identified as most vulnerable by both CDC-SVI themes and CRE-PR are described in **Table 4**. We find the average relative error is lower for areas tagged similarly between CDC-SVI overall themes and CRE-PR. The relative error reduced for areas tagged similarly is 34.53% - 36.61% while all other tracts are reduced on average by 45.31% - 52.23%.

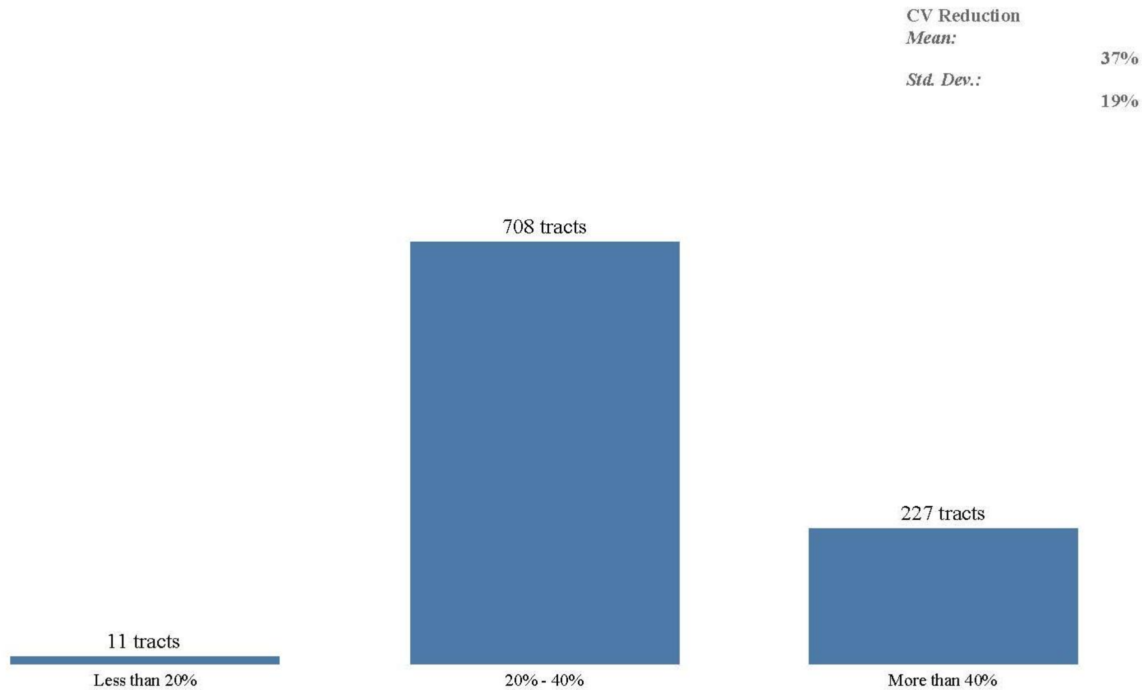
6 | DISCUSSION

CRE-PR provides a major advancement in the measurement of how at-risk every neighborhood in Puerto Rico is to the impacts of disasters. CRE-PR provides more timely and accurate data, which is critical to help communities better plan for and respond to disasters. While other indices rely upon 5-year PRCS data, CRE-PR methodology combines the 1-year PRCS estimates with other data sources to provide more timely, precise, and stable estimates than any other index or estimate that uses publicly available data.

While there are some similarities, CRE-PR tags many areas as more vulnerable differently than CDC-SVI. Areas tagged as high-risk by CDC-SVI but not CRE-PR have a greater reduction in relative error, in comparison to all other areas. This supports our hypothesis that the larger survey error that small area modeling reduces is driving these areas to be inappropriately flagged as high-risk. Areas tagged as high-risk by both CDC-SVI and CRE-PR have a smaller reduction in relative error, in comparison to all other areas. This supports our hypothesis that areas are tagged similarly because there is less survey error for small area methods to enhance.

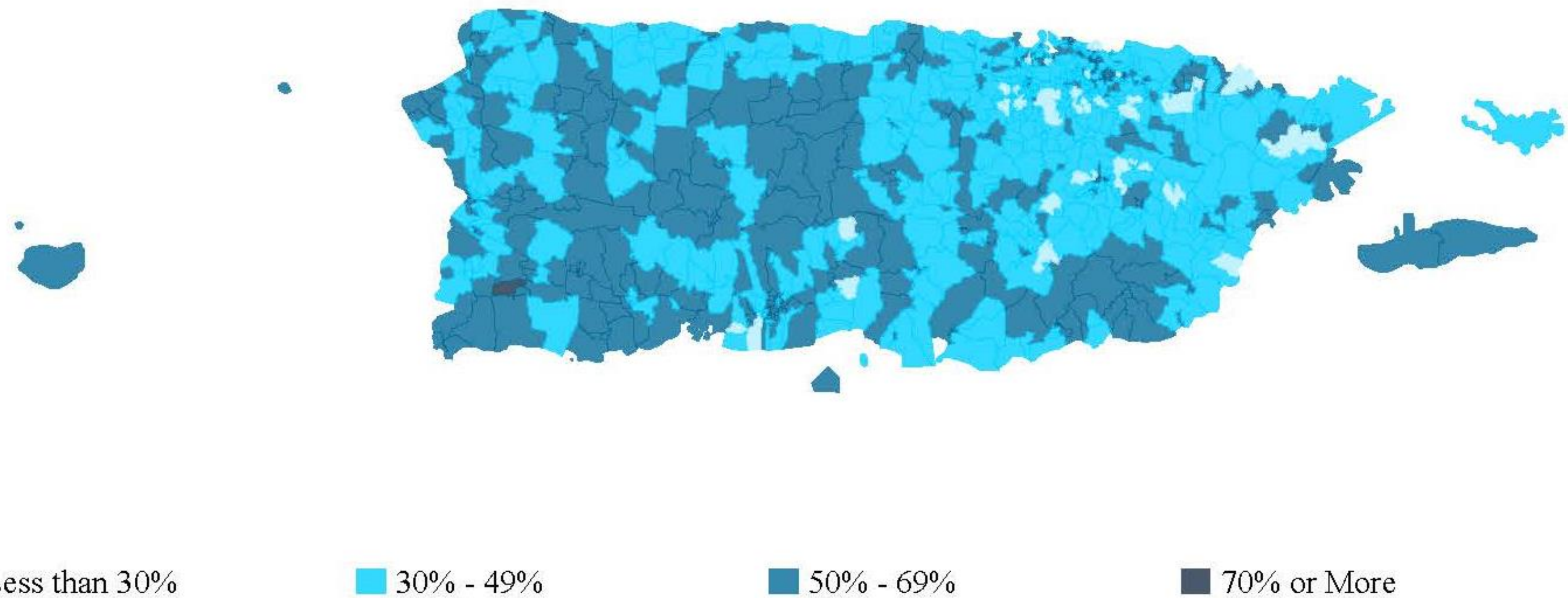
In addition to providing a methodological advancement, CRE-PR provides a practical advancement to emergency planning, mitigation, response and recovery. Because CRE-PR produces measures of error along with estimates, it can be used to measure significant differences in vulnerability between places or over time. CRE-PR can help decision-makers make more accurate data-driven choices to increase the resilience of underserved communities across Puerto Rico.

Fig. 1. Description of Percent Reduction in the Relative Error of High-Risk Population Estimates for Populated Census Tracts. Individuals with 3 or more vulnerability indicators are high-risk.



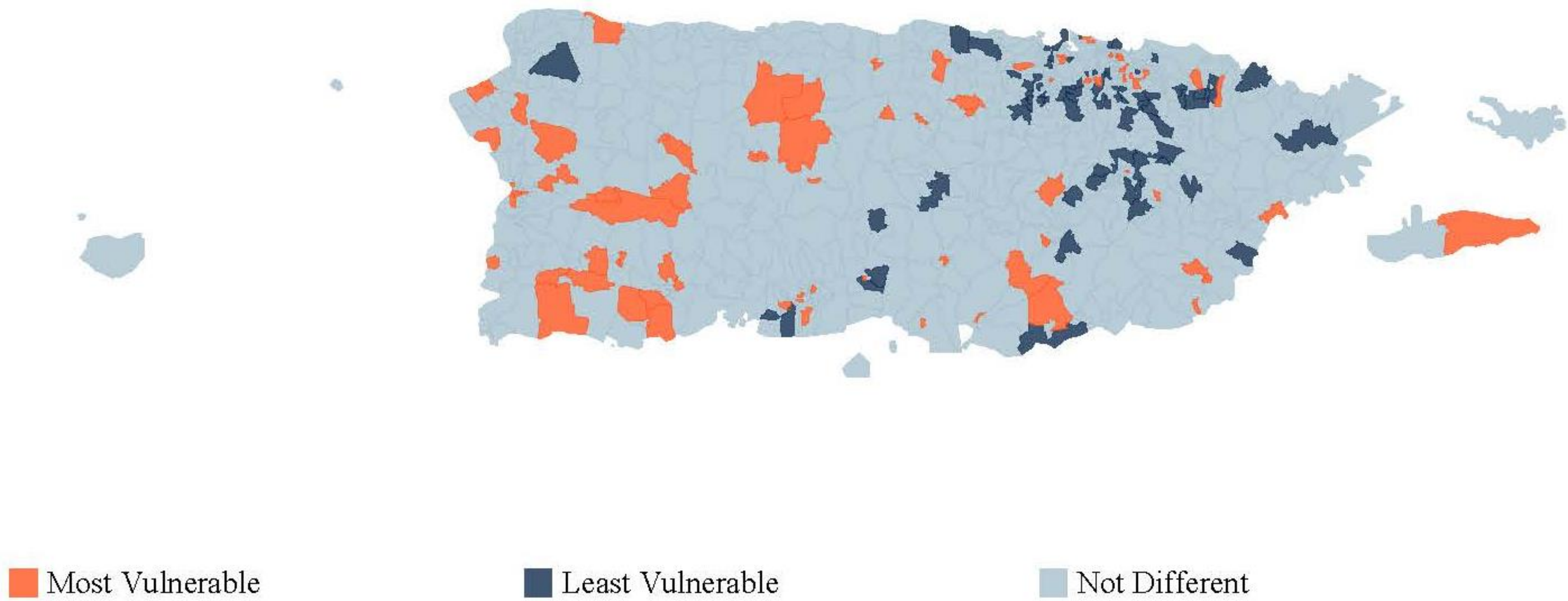
Source: 2019 Community Resilience Estimates for Puerto Rico.

Fig. 2. Map of High-Risk Population Rate Estimates for Populated Census Tracts. Individuals with 3 or more vulnerability indicators are considered high-risk.



Source: 2019 Community Resilience Estimates for Puerto Rico.

Fig. 3. Most and Least Vulnerable Census Tracts in Puerto Rico. Individuals with 3 or more vulnerability indicators are considered high-risk. The most vulnerable tracts have a high-risk population rate higher than the Puerto Rico average. The least vulnerable tracts have a high-risk population rate lower than the Puerto Rico average.

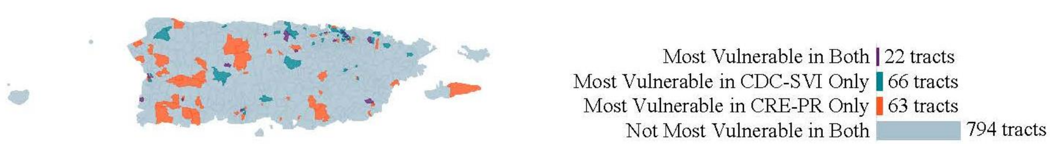


Source: 2019 Community Resilience Estimates for Puerto Rico.

Fig. 4. Similarities and Differences in Census Tracts Identified as Most Vulnerable by CRE-PR and CDC-SVI. Only the 895 tracts in Puerto Rico that are included in both estimates are displayed in 2019 geography. Individuals with 3 or more vulnerability indicators are considered high-risk. CRE-PR identifies the most vulnerable tracts as those with a high-risk population rate higher than the Puerto Rico average. CDC-SVI flags the most vulnerable tracts based on the 90th percentile rank

Comparison of CRE-PR with CDC-SVI:
Overall Vulnerability Score Flags

Socioeconomic Status Vulnerability Score Flags



Household Characteristics Vulnerability Score Flags

Housing Type and Transportation Vulnerability Score Flags

Source: 2019 Community Resilience Estimates for Puerto Rico and 2020 Centers for Disease Control Social Vulnerability Index.

Table 1. Summary Statistics for Low-, Moderate-, and High-Risk Populations in Most Vulnerable, Least Vulnerable, and All Census Tracts in Puerto Rico. Individuals with 3 or more vulnerability indicators are considered high-risk. Margins of error are shown after estimates in parenthesis. The most vulnerable communities are Puerto Rico Census tracts with a high-risk population rate above the Puerto Rico rate. The least vulnerable communities are Census tracts with a high-risk population rate below the Puerto Rico rate.

	All Census Tracts	**Most Vulnerable Census Tracts**	**Least Vulnerable Census Tracts**
	Population	**Population**	**Population**
Total:	3,180,019	253,341	317,496
By Risk Category	**Population Rate**	**Population Rate**	**Population Rate**
Low-Risk	15% (+/-2%)	10% (+/-12%)	25% (+/-5%)
Moderate-Risk	39% (+/-1%)	26% (+/-6%)	47% (+/-4%)
High-Risk	46% (+/-1%)	64% (+/-3%)	28% (+/-6%)

Source: 2019 Community Resilience Estimates for Puerto Rico.

Table 2. Correlations Between Puerto Rician Tract High-Risk Population Rates and Vulnerability Scores. Individuals with 3 or more vulnerability indicators are considered high-risk. CDC-SVI vulnerability scores are based on percentile ranks.

	Pearson's R Coefficient				
	CRE-19	CDC-SVI20: Overall	CDC-SVI20: Socioeconomic	CDC-SVI20: Household	CDC-SVI20: Housing
2019 CRE-PR: High-Risk Population Rate					
CDC-SVI20 Overall Vulnerability Score	.52				
CDC-SVI20 Socioeconomic Status Vulnerability Score	.49	.78			
CDC-SVI20 Household Characteristics Vulnerability Score	.46	.69	.43		
CDC-SVI20 Housing Type and Transportation Vulnerability Score	.21	.66	.28	.17	

Source: 2019 Community Resilience Estimates for Puerto Rico and 2020 Centers for Disease Control Social Vulnerability Index.

Table 3. Two Sample T-Test and Analysis of Variance F-Test Comparing the Reduction in Relative Error for High-Risk Population Estimates through Small Area Modeling for Tracts Flagged as Most Vulnerable by CDC-SVI Overall Theme Score and Not by CRE-PR. Individuals with 3 or more vulnerability indicators are considered high-risk. Individuals with 3 or more vulnerability indicators are considered high-risk. CRE-PR identifies the most vulnerable tracts as those with a high-risk population rate higher than the Puerto Rico average. CDC-SVI flags the most vulnerable tracts based on the 90th percentile rank of aggregated indicators.

Tracts Flagged as Most Vulnerable by CDC-SVI Only	T-Test Method	90% Confidence Level Mean		90% Confidence Level Std. Dev.					
0		-0.3760	-0.3550	0.1765	0.1913				
1		-0.5313	-0.4354	0.2680	0.2680	**Variances**	**D.F.**	**T Value**	**Pr > |t|**
Diff (1-2)	Pooled	0.0776	0.1581	0.1800	0.1946	Equal	893	4.82	<.0001
Diff (1-2)	Satterthwaite	0.0687	0.1669			Unequal	68.22	4.00	0.0002

ANOVA	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.8130	0.8130	23.24	<.0001
Error	893	31.24	0.0350		
Corrected Total	894	32.05			

R-Square	Coeff Var	Root MSE	Average CV Reduction
0.0254	-50.03	0.1870	-0.3738

Source	DF	Anova SS	Mean Square	F Value	Pr > F
CDC-SVI Only	1	0.8130	0.8130	23.24	<.0001

Source: 2019 Community Resilience Estimates for Puerto Rico and 2020 Centers for Disease Control Social Vulnerability Index.

Table 4. Analysis of Variance F-Test Comparing the Reduction in Relative Error for High-Risk Population Estimates through Small Area Modeling for Tracts Flagged as Vulnerable by Both CDC-SVI Overall Vulnerability Score and CRE-PR. Individuals with 3 or more vulnerability indicators are considered high-risk. Individuals with 3 or more vulnerability indicators are considered high-risk. CRE-PR identifies the most vulnerable tracts as those with a high-risk population rate higher than the Puerto Rico average. CDC-SVI flags the most vulnerable tracts based on the 90th percentile rank of aggregated indicators.

Tracts Flagged as Vulnerable by CDC-SVI and CRE-PR	T-Test Method	90% Confidence Level Mean		90% Confidence Level Std. Dev.		Variances	D.F.	T Value	Pr > |t|
0		-0.5223	-0.4531	0.2097	0.2590				
1		-0.3661	-0.3453	0.1682	0.1829				
Diff (1-2)	Pooled	-0.1614	-0.1026	0.1770	0.1914	Equal	893	-7.40	<.0001
Diff (1-2)	Satterthwaite	-0.1681	-0.0959			Unequal	145.1	-6.05	<.0001

ANOVA	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.850	1.850	54.69	<.0001
Error	893	30.20	0.0338		
Corrected Total	894	32.05			

R-Square	Coeff Var	Root MSE	Average CV Reduction
0.0577	-49.19	0.1839	-0.3738

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Both	1	1.850	1.850	54.69	<.0001

Source: 2019 Community Resilience Estimates for Puerto Rico and 2020 Centers for Disease Control Social Vulnerability Index.