Research Article

Health effects of heat vulnerability in Rio de Janeiro: a validation model for policy applications

Diogo Prosdocimi[1,2] · Kelly Klima[2]

Received: 13 June 2020 / Accepted: 21 October 2020 / Published online: 5 November 2020

Abstract

Extreme heat events can lead to increased risk of heat-related deaths. Furthermore, urban areas are often hotter than their rural surroundings, exacerbating heat waves. Unfortunately, validation is difficult; to our knowledge, most validations, even if they control for temperatures, really only validate a social vulnerability index instead of a heat vulnerability index. Here we investigate how to construct and validate a heat vulnerability index given uncertainty ranges in data for the city of Rio de Janeiro. First, we compare excess deaths of certain types of circulatory diseases during heat waves. Second, we use demographic and environmental data and factor analysis to construct a set of unobserved factors and respective weightings related to heat vulnerability, including a Monte Carlo analysis to represent the uncertainty ranges assigned to the input data. Finally, we use distance to hospital and clinics and their health record data as an instrumental variable to validate our factors. We find that we can validate the Rio de Janeiro heat vulnerability index against excess deaths during heat waves; specifically, we use three types of regressions coupled with difference in difference calculations to show this is indeed a heat vulnerability index as opposed to a social vulnerability index. The factor analysis identifies two factors that contribute to >70% of the variability in the data; one socio-economic factor and one urban form factor. This suggests it is necessary to add a step to existing methods for validation of heat vulnerability indices, that of the difference-in-difference calculation.

Keywords Heat vulnerability · Validation · Heat wave · Heat-related deaths

1 Introduction

Extreme heat events can have major impacts on people's lives. Brazil, with its predominantly tropical monsoon climate [2], is one of the world's hottest countries [40] and historically has had significant increases in heat wave frequency [26]. The types of heat waves that Brazil experiences can lead to increased risk of heat-related deaths [4, 12, 24, 29], and numbers of deaths are more likely due to climate change [7, 46].

Furthermore, the urban heat island effect (where urban areas are hotter than surrounding rural areas) exacerbates heat waves [64, 65], especially in very densely populated areas [45]. Unfortunately for researchers, there are relatively few populated areas with hyperlocal areas spanning extreme socioeconomic disparity. This can make it difficult to understand whether extreme heat deaths are due to changes in local temperature or changes in urban form (and population living there). We are aware of one study in the United States, which shows that the types of characteristics leading to increased vulnerability differ between

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42452-020-03750-7) contains supplementary material, which is available to authorized users.

✉ Kelly Klima, kklima@rand.org | [1]Pardee RAND Graduate School, Santa Monica, CA, USA. [2]Engineering and Applied Sciences, RAND Corporation, 1776 Main Street, Santa Monica, CA 90401-3208, USA.

urban and rural areas; however, this study actually showed lower deaths during hotter heat waves (see [43] which used temperature and humidity to define the apparent temperature) suggesting that more work needs to be done to understand the differences between rural and urban areas. While regression techniques exist to address these problems, it can sometimes ease the calculation burden to focus on hyperlocal areas. Perhaps unique to Brazil is the presence of favelas (slums), where the richest individuals in Brazil live a couple of meters away from the population living in these conditions (Leblon—the richest neighborhood in Rio—and is less than 3 km away from Rocinha—the largest favela in Latin America) [35]. To our knowledge, few studies have been done to understand how Brazil's unique characteristics, such as favelas, might contribute to our understanding heat vulnerability.

Intertwined with weather and the environment is heat vulnerability. An individual's heat vulnerability is known to increase with a number of factors, such as age [1], income [38], or ethnicity [55]. Generally speaking, two approaches exist to assess heat distress and deaths caused by heat exposure: regression and factor analysis [3]. In the first method, a researcher obtains data on deaths and regresses out the characteristics leading to increased vulnerability (e.g., [42, 62]). In the second method, a research attempts to develop heat vulnerability measures using factor analysis or equal weights with generalized models, and then validate those models [53]. Unfortunately, due to limitations on personally identifiable information and the need to aggregate and anonymize, in many countries it can be difficult to obtain fine resolution health record data to validate these models. Conversely, in Brazil, health record data are publicly available. Furthermore, since in Brazil health emergency calls and health care are both free and universal (and do not require enrollment), often patients experiencing any heat-related symptom or condition are taken to the nearest hospital or clinic for urgent care. Thus, since the data are unclouded with differences in hospitalization choices, it is likely that hospitals closer to high heat vulnerability neighborhoods would have higher rates of heat-related illnesses (e.g., cardiovascular disease, myocardial infarction, and chronic pulmonary disease hospitalization and deaths) during heatwaves.

Furthermore, consider the studies that have attempted a validation. Bao et al. [3] conduct a literature review of heat vulnerability index construction and validation. They find that the temperature level is the most well-documented contributor to heat-related deaths, and that it is difficult to validate a heat vulnerability index. This suggests that it is difficult to determine whether the indices are a general social vulnerability index (which measures vulnerability) or a heat vulnerability index (which has nonuniformly more deaths as temperatures rise). Indeed, this is the conclusion reached by multiple authors (e.g., [16, 53]).

Given this, our study considers three research questions in Brazil. First, what are the health effects of heat waves in Rio de Janeiro? Second, will a vulnerability index built over socio-economic and urban form variables retain urban form characteristics? Third, are we able to validate the proposed heat vulnerability index against health record data during heat waves? We use three types of regressions coupled with difference in difference calculations to test whether our index is indeed a heat vulnerability index as opposed to a social vulnerability index. Furthermore, we test our finding across multiple uncertainties, including that of the heat wave definition and that of the input data to the heat vulnerability index.

2 Methods

2.1 Heat-related deaths

We compared heat wave data and data characterizing deaths over the period February 2007 to February 2016. First, consider the heat wave definition and data. Within the literature, the definition of heat wave events (and thus heat wave intensity) varies widely as a certain temperature threshold [20, 22, 27], a function of extreme percentiles of temperatures [14, 25, 51], or a particular heat index [47, 61]. In this study, we used temperature and relative humidity data from Instituto Nacional de Meteorologia [36] to calculate the incidence of heat waves in Rio de Janeiro. We conducted our analysis for two published definitions for heat waves. First, following Rothfusz, we use the temperature and relative humidity to calculate the heat index [54]. We considered heat hazard days with heat index higher than 32 °C (90 °F), which is defined by the National Weather Service (NWS) to be in the "Extreme Caution" range for likelihood of heat disorders with prolonged exposure or strenuous activity [33]. Second, and considering that a policy-maker might wish for a simpler data collection method, following [42], we considered a temperature-only definition of an extreme heat wave event being a period of two or more consecutive days with lower temperatures over 25 °C (77 °F) and higher temperatures over 35 °C (95 °F). For each definition, we included two additional days at the end of each extreme heat days, to account for health effects occurring after the event.[1]

[1] Future work could conduct additional sensitivity analyses; since the results are similar (insensitive) to the definitions of heat waves used, we explore other sensitivities in this paper.

Table 1 Socio-economic and urban form descriptive statistics for Rio de Janeiro

Variable	*Count within census block group: mean (standard deviation)*
Population of Household	2.94 (0.46)
Average Individual Income (R$)	1381 (1378)
Total Population	617 (296)
Variable	*Fraction within census block group: mean (standard deviation)*
Age > 60 years	0.002 (0.021)
Age < 5 years	0.843 (0.084)
Literate in Portuguese	0.181 (0.088)
Branco (White)	0.095 (0.059)
Preto (Black)	0.527 (0.213)
Amarelo (Yellow)	0.155 (0.107)
Pardo (Mixed)	0.028 (0.045)
Indigena (Native)	0.369 (0.168)
Per capita income <1/8 minimum wage	0.020 (0.050)
Per capita income equal to zero	0.023 (0.049)
No access to water network	0.974 (0.118)
No access to sewage network	0.003 (0.021)
No garbage service	0.019 (0.106)
Presence of garbage in the street	0.117 (0.168)
No access to energy service	0.026 (0.118)
No access to energy	0.981 (0.105)
No streetlight	0.019 (0.105)
No sidewalk	0.037 (0.144)
No pavement	0.037 (0.147)
No trees	0.044 (0.162)
Variable	*Normalized to 0 to 1: mean (standard deviation)*
Normalized Difference Vegetation Index (NDVI)	0.109 (0.267)

Now consider deaths. Due to the vast epidemiologic evidence (e.g., [5, 6, 60]), researchers suggest that heat waves increase the probability of death from certain circulatory system diseases including specific cardiovascular diseases of heart disease, congestive heart failure, and myocardial infarction. For example, in Brazil, 115 excess deaths per year have been shown to occur due to the link between acute myocardial infarction (heart attack) and increased temperatures [23]. In this paper, we use data from the Brazilian healthcare database called DATASUS [48] to assess socio-economic determinants of heat-related deaths. Within these data, from 2007 to 2016, 449 clinics have reported a total of approximately 168,000 deaths in Rio de Janeiro by diseases of the circulatory system (ranging from 1180 in February 2015 to 1650 in February 2008).

2.2 Heat vulnerability index

We created a heat vulnerability index including the urban form characteristics of Brazil's favelas. Literature has demonstrated in Brazil [8] and elsewhere [10] that there are socioeconomic characteristics that are indicative of increased deaths in heat wave events. We used a type of scaling called factor analysis to develop a predictive index for social vulnerability. The objective was twofold: (1) to construct a latent variable, or an index (or a scale), to spatially measure social vulnerability of a population; (2) to identify the underlying dimensions of the index to support public policy.

Based on literature, we hypothesized a number of variables that affect heat vulnerability (Table 1). All variables were coded so that higher values indicated higher vulnerability. We collected most socio-economic and urban form data from the 2010 Censo [35], a Brazilian decennial survey of all households in country (where *setor censitário* means "census tract", and has a size similar to a United States census block group[2]). A full description of the variables,

[2] There are a total of 10,233 *setores censitário* in Rio de Janiero. In our study, the number of people in a *setor censitário* is on average 617 people (see Table 2). Figure 2, which shows part of our results, depicts within the choropleth map the individual *setores censitário.*

including both the reason for inclusion and characteristics within our dataset, is given in the Electronic Supplementary Information.

Next, following existing literature [10, 31, 32, 50, 52], we then combined the variables to create factors. First, we tested for multicollinearity of the variables, and dropped variables with Pearson coefficient higher than 0.75 (see Electronic Supplementary Material, Table S.1). Second, we conducted a factor analysis [49] with varimax rotation, retaining the factors following the Kaiser rule (eigenvalue higher than one). Finally, we calculated the index from the factor scores. For ease of interpretation, we converted results into seven groups of one standard deviation (Minimum: < −2.5SD from the mean; Very Low: −2.5 to −1.5 SD from the mean; Low: −1.5 to −0.5 SD from the mean; Medium: −0.5 to 0.5 SD from the mean; High: 0.5 to 1.5 SD from the mean; Very High: 1.5 to 2.5 SD from the mean; Maximum: >2.5 ST from the mean).

While spatial clustering analyses have existed for many years (e.g., see summary texts on spatial autocorrelation such as [18, 19]), there is a recent interest in the literature on examining the vulnerabilities for spatial correlation. For example, a recent paper calculates a Moran's I analysis on the Center for Disease Control's social vulnerability index, heat-related emergency room visits, and heat mortality at the county level in the state of Georgia, U.S., finding significant levels of high clustering [41]. Conversely to this paper, our analysis is conducted at the *setor censitário* level, which is roughly equivalent to a U.S. census tract, and only within the urban area. Thus, while we do not expect to see spatial clustering from urban and rural differences (as these are not in our data set), we calculate a univariate Moran's I to determine whether there is clustering in the city itself.

Finally, we checked the sensitivity of the heat vulnerability index performing a 10,000 simulation to check the robustness of the results to measurement error in the variables (up to 1.96 standard deviation).

2.3 Validating heat vulnerability against excess deaths during heat wave

We tested whether the index is a good predictor of health effects in the population during a heat wave. Our hypothesis is that more vulnerable regions in the city may experience higher health effect due to heat events. That is, we should observe more deaths in regions with higher vulnerability.

Existing research has considered over-dispersed generalized linear modeling, finding that there is a relation between deaths and temperature in San Paolo, Brazil [57]. Here, since we consider both distance to hospitals and socio-economic vulnerability, we specified two parts of a model to validate our index: (1) a model to determine whether the vulnerability index is statistically significant, and (2) a logistic model predicting whether zero deaths observed is a certain zero. For each of these model approaches, we specified explanatory variables and an outcome variable, and then tested a linear regression, zero-inflated Poisson, and zero-inflated negative binomial (allowing for tests of overdispersion and excess of zero in the data).

In each of the models, we use three of the same explanatory variables. First, the number of heat waves takes the part of a "treatment" that we are aiming to measure the health impacts, and second the duration of heat waves is the intensity of the "treatment", or the dosage effect. These two variables allow us to make inferences about possible diminishing effects for longer heat waves or periods with more than one heat wave. The third explanatory variable, the heat vulnerability index, is the construct we are aiming to validate.

Following the literature (e.g., [13, 21, 37]), we use the relative distance from *setor censitário* to the nearest hospital (following the road network using GIS network analyst tools) as an instrument to patient hospital/clinic choice and, eventually, their death location. Distance from their residence to the nearest hospital as an instrument to patient choice for healthcare is reasonable for two reasons. First, the emergency medical services in Brazil direct patients to the nearest hospital for urgent issues [9]. Second, Brazil has a free and universal healthcare system (not even an opt-in is required) and even individuals enrolled in private health insurance are directed to the public system during emergencies [30]. The matching process provided the response variable (number of deaths in the nearby hospital) and one control variable (the population served by each hospital—that is, the sum of the populations of all *setores censitário* from each hospital). We also calculated the death rate using these two variables that would be an explanatory model for a second model specification.

Then for the three hybrid models, we specified explanatory variables as the heat wave variables, the vulnerability index, and the population count, where the outcome variable is the number of observed deaths. Also, given known effects of seasonal variations in certain types of deaths (e.g., [44]), we included in our model time fixed effects to account for seasonality in the death data. Note, the geographic unit, the *setor censitário*, is on average 617 people (see Table 1), and so excess deaths of 2–3 people per *setor censitário* is approximately 0.5% of the population.

An important caveat is to ensure that the index measures heat vulnerability and not a general socio-economic vulnerability. Thus, for the models, we clean our coefficients performing a simple differences-in-differences calculation between high and low vulnerable regions with and without occurrence of heat wave. The conclusion and discussion will use this parameter rather than the coefficients of the estimation models.

Finally, in addition to using multiple definitions and model specifications to test the robustness of our result, we used two other methods to check the robustness of our results. First, we checked the sensibility of our results using a Monte Carlo Simulation. We ran 1000 simulations to check the sensibility of the results to measurement error. We adopted the following steps: (1) we created new variables with the standard normal deviation of each socio-economic and physical characteristic; (2) we generated a random number based on a beta distribution and performed the appropriate transformations to have values between −1 and 1; (3) we multiplied the values in (1) and (2) with the value of each observation; (4) we performed the factor analysis and the validation count model for the data with this measurement error shock; (5) we repeated the first three procedures 1000 times. We recorded the resulting coefficients only for the two predictors of interest: heat vulnerability index and length of the heat wave.

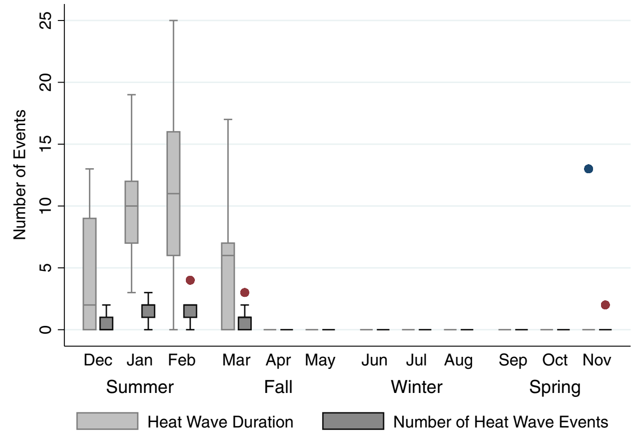

Fig. 1 The distribution of the occurrence and duration of heat waves in Rio de Janeiro by month. The box represents the interquartile range (25th to 75th percentile), the whiskers represent an additional 1.5 times the interquartile range, and the outliers are represented by dots

3 Results

3.1 Heat-related deaths

We find that, regardless of heat wave definition chosen, Rio de Janeiro has had excess deaths due to extreme heat. First, consider a heat wave definition that includes both temperature and humidity (per [54]), or defined as three or more consecutive days with heat index over 32 °C /90 °F. Our data show that Rio de Janeiro had 60 such events during the period of analysis (February/2007 to February /2016). Moreover, heat index exceeded 103 °F in 2 events, totaling 8 days. Figure 1 shows the distribution of the occurrence and duration of heat waves in Rio de Janeiro by month. Most heat waves occur between late spring and early autumn, with at least one heat wave occurring in each January and in each February. Heat wave duration spans 0–29 days, with an average duration of 8 days in December, 17 days in January, and 14 days in February.

Considering the heat waves and the two subsequent days after each event, we find that the number of deaths increases with number and duration. The coefficient for the variable heat wave duration is 0.003677 and statistically significant at 1% level, meaning that a heat event would increase the count of deaths by 0.4% (or a factor of exp.[0.003677]). That is, the marginal effect of each additional day in the length of a heat wave would lead to an increase in the order of 0.0301 in the death count. These results indicate that the number of deaths in each *setor censitário* (which, on average, consists of 617 people) increases from 2.59 to 2.89 during a 10-day heat wave, holding all other variables constant at the mean. We find that a 25 day long heat wave would lead to a death count of 3.40 in each *setor censitário*. This is approximately 0.5% of the population, which since Rio de Janiero is home to approximately six million people, means approximately an excess 30,000 deaths.

Next, consider a definition that might be easier for a policy-maker to obtain: a temperature-only definition. Here, consider minimum temperatures higher than 25 °C/77 °F and maximum temperatures higher than 35 °C/ 95 °F (per a combination of [17, 47]). Given this, Rio de Janeiro experienced 46 heat waves, totaling 319 days. We find that the number of deaths increases with number and duration. The coefficient for the variable heat wave duration is .003681 and statistically significant at 1% level, meaning that a heat event would increase the count of deaths by 0.4% (or a factor of exp.[0.003681]). That is, the marginal effect of each additional day in the length of a heat wave would lead to an increase in the order of 0.0230 in the death count. These results indicate that the number of deaths in each setor censitário increases from 2.66 to 2.85 during a 10-day heat wave, holding all other variables constant at the mean. We find that a 25 day long heat wave would lead to a death count of 3.15 in each setor censitário.

These results are qualitatively similar. We also checked the same model specification with multiple definitions of heat wave (e.g., [20, 22, 27, 28, 56]), and found that the results do not change for a simplified definition using only temperature or considering only more extreme events.

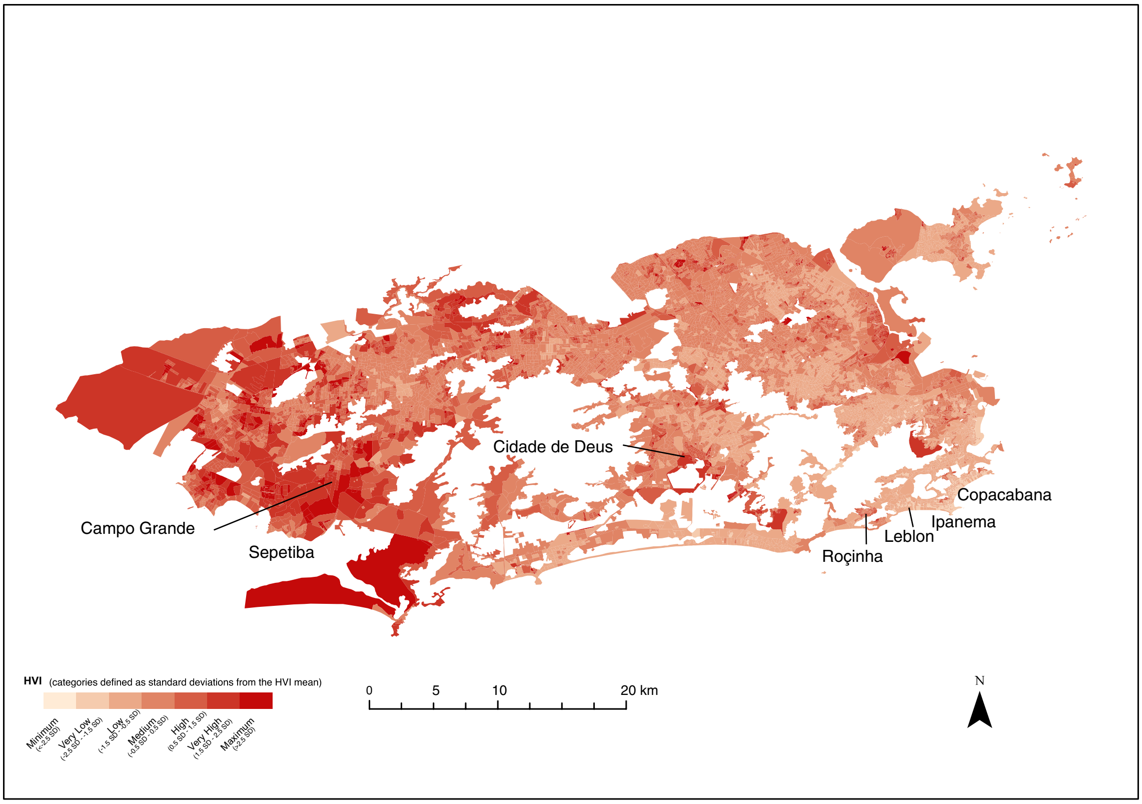

Fig. 2 Heat vulnerability map for Rio de Janeiro and the hospitals that reported deaths due to cardiovascular diseases. Legend items are as follows: Minimum: < −2.5SD from the mean; Very Low: −2.5 to −1.5 SD from the mean; Low: −1.5 to −0.5 SD from the mean; Medium: −0.5 to 0.5 SD from the mean; High: 0.5 to 1.5 SD from the mean; Very High: 1.5 to 2.5 SD from the mean; Maximum: >2.5 ST from the mean

3.2 Heat vulnerability index

We collected multiple socio-economic and urban form characteristics shown to be predictive of heat vulnerability. First, we performed multi-correlation tests with the socio-economic and environmental characteristics to determine relative unique contributions of variables. We found that only having children younger than 5 years old was highly correlated with being able to read and write in Portuguese (Pearson coefficient of 0.81), and so dropped the first one since it describes a characteristic with less amplitude.

Then, we performed factor analysis and the results revealed that two underlying factors capture 94.74% of the data variance in explaining vulnerability across populations in Rio de Janeiro (see Electronic Supplementary Material, Table S.2, Fig. S.1). The first dimension (factor 1) is a construct of socio-economic characteristics such as age, race, and alphabetization. The second dimension (factor 2) is a construct of urban form, including arborization, public lights, and presence of sidewalks and paved streets. Figure 2 shows the heat vulnerability map for Rio de Janeiro and the hospitals that reported deaths due to cardiovascular diseases. We see that favelas and slums neighborhoods have higher vulnerability given the high weight of urban form in the factor analysis. This is an expected finding given the importance for the composition of the index in the factor analysis of variables such as paved roads, presence of sidewalk, streetlight, trees and sewage. Favelas and slums in Brazil lack this kind of public infrastructure.

Other high vulnerability regions include traditionally poor and non-white neighborhoods, such as Guaratiba, Sepetiba and Santa Cruz. These suburb regions have more public infrastructure than the favelas and slums but may have more profound socio-economic disparities from the main regions in the city such as unemployment and less education.

We conducted a univariate Moran's I test on the heat vulnerability index at the *setor censitário,* and find across the city a statistical significance of 0.11 at $p < 0.0001$. Since Moran's I runs from −1 to 1 with 0 indicating no correlation and one indicating perfect correlation, our data show a small amount of autocorrelation. This suggests that we may proceed with using the heat vulnerability index; in our discussion, we describe how future research could examine the spatial autocorrelation.

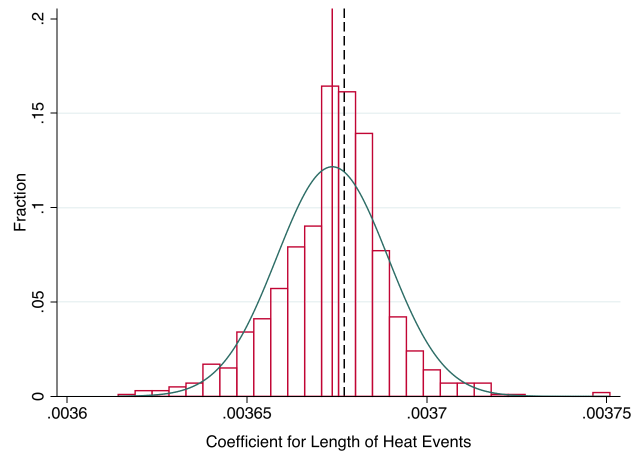

Fig. 3 Probability distribution function of Monte Carlo analysis of the coefficient for heat wave duration. Note the difference between the mean of the Monte Carlo results (solid line, 0.003673) and the initial result (dashed line, 0.003677)

To understand the sensitivity of our results to measurement error in the Brazilian census data, we performed 1000 simulations using the margin of error of Brazilian census data (and 1.96 standard deviation from the mean for the other variables). We found that our results did not change significantly due to uncertainty in the census data. As shown in Fig. 3, the simulation's mean is 0.003673 for the variable of length of heat waves while our main result is 0.003677. That is, the difference between the mean in less than 10^{-5}.

3.3 Validating heat vulnerability against excess deaths during heat wave

Finally, we use distance to hospital and clinics and their health record data (both diseases knowingly related and unrelated to heat waves) to validate the heat vulnerability index against excess deaths during heat waves.

In the first part of the model, we find statistically significant coefficients for heat vulnerability and heat wave variables (see Table 2). A zero-inflated model indicates that the vulnerability index is statistically significant at 99.9% level and one increase in the vulnerability index is associated with an increase of a factor of 1.05 (exp[0.049]). The effect is diminishing since the coefficient for squared heat vulnerability index is negative. The expected number of deaths increases a factor of 1.004 (exp[0.0037]) during the first heat event. The zero-binomial specification shows similar results than our main model, showing that standard deviations may not be biased and the statistical significance of the results is reliable. Our results also corroborate with findings that deaths due to cardiovascular diseases have seasonal variation with higher incidence during the winter than in the summer [58]).

The second part of the model, a logistic model predicting whether or not zero deaths observed is a certain zero, also corroborates the literature. Consider: longer the distance to hospital [15], less vulnerable populations [63], and absence of heat events [11] would decrease the probability of deaths (from cardiovascular diseases) within in a *setor censitário*. Figure 4 shows the first derivative of the response with respect of the length of heat waves for different levels of vulnerability levels. Low vulnerable *setores censitário* also face an increase in deaths during heat wave, but with lower magnitude. The number of deaths increases from 2.527 to 2.828 during a 10-day heat wave and to 2.483 during a 20-day heat wave. The graph also shows us the count of deaths by cardiovascular disease for *setores censitário* with different levels of vulnerability. For instance, very highly vulnerable census tracts experience 0.587 counts more than the medium vulnerability while low vulnerable census experience 0.253 less counts during a 5-day heat wave.

Combining these two results, we perform a difference-in-difference analysis to isolate the impact of vulnerability during the longest heat wave experienced in Rio de Janeiro (that lasted 25 days), as shown in Table 3. The first difference (the estimated count of deaths for a low vulnerable *setor censitário* and a highly vulnerable *setor censitário* when no event has occurred) is approximately 0.795. The second difference separately considers the subset of *setores censitário* within each vulnerability level. For each group, we find that the number of deaths increases as the length of the heat wave event increases (from no days to a 25 day length event). These two characteristics indicate that, at a minimum, the developed index is indicative of vulnerability to all types of events. To determine whether the index is indicative specifically of heat vulnerability, one would need to see that there is an even higher increase in the most vulnerable *setores censitário* than in the least vulnerable *setores censitário* when moving from zero events to a 25 day long event. We see that the difference in deaths across high and low heat vulnerability becomes 1.043 during the 25 day heat wave, which is an increase of 0.248 over the non-heat wave event. This indicates that the developed index is also representative of heat vulnerability.

4 Discussion

We show three findings in Rio de Janeiro. First, as shown in the literature [23], we find that heat waves are correlated with excess deaths of certain types of circulatory diseases. The factor analysis identifies two factors that contribute to

Table 2 Heat vulnerability index validation results showing estimation of number of deaths in Rio de Janeiro's *setores censitário*

Variable	Model 1: Linear regression; B (se)	Model 2: Zero-inflated Poisson B (se)	Model 3: Zero-inflated negative binomial B (se)
Heat vulnerability index	0.086 (0.082)	0.049*** (0.003)	0.039*** (0.004)
Heat vulnerability index, squared	−0.059 (0.047)	−0.030*** (0.002)	−0.026*** (0.002)
Distance to closest clinic	−0.000***(0.000)	−0.000*** (0.000)	−0.000*** (0.000)
Population (log)	0.137 (0.086)	0.008** (0.003)	0.046*** (0.004)
Population (log) squared	−0.000 (0.000)	0.000*** (0.000)	0.000 (0.000)
Length of Heat Events	0.004*** (0.001)	0.004*** (0.000)	0.003*** (0.001)
January	0.118*** (0.014)	0.018 (0.012)	0.051*** (0.013)
February	−0.283*** (0.014)	−0.090*** (0.011)	−0.098*** (0,013)
March	−0.133*** (0.014)	0.002 (0.011)	−0.019 (0.012)
April	−0.129*** (0.014)	0.002 (0.012)	−0.016 (0.013)
May	0.184*** (0.014)	0.100*** (0.011)	0.096*** (0.013)
June	0.229*** (0.014)	0.130*** (0.012)	0.121*** (0.013)
July	0.357*** (0.014)	0.139*** (0.012)	0.152*** (0.013)
August	0.283*** (0.014)	0.134*** (0.011)	0.133*** (0.013)
September	0.140*** (0.014)	0.081*** (0.011)	0.076*** (0.013)
October	0.099*** (0.014)	0.071*** (0.012)	0.061*** (0.013)
November	−0.077 (0.014)	−0.015 (0.011)	−0.010 (0.013)
Constant	1.677* (0.864)	1.889*** (0.031)	1.049*** (0.040)
Inflate			
Distance to closest clinic	–	0.000*** (0.000)	0.000*** (0.000)
z-scored HVI		−0.018*** (0.002)	−0.009*** (0.004)
# of Heat Events	–	0.108*** (0.005)	0.169*** (0.009)
Length of Heat Events	–	−0.012 *** (0.001)	−0.020*** (0.001)
Constant	–	0.288*** (0.003)	−0.712*** (0.012)
Model statistics			
R-squared	0.0046	n/a	n/a
log likelihood of null model	−3,737,891	−3,098,517	−1,890,717
log likelihood of full model	−3,735,360	−3,089,081	−1,889,815
AIC	7,470,755	6,178,205	3,779,677
BIC	7,470,970	6,178,467	3,779,951
df	17	22	23

B standardized estimate, *se* standard error, β unstandardized estimate

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

94.74% of the variability in the data; one socio-economic factor and one urban form factor. This indicates, as suggested in the literature (e.g., [43]), that there are differences between rural and urban vulnerability indices; in our index unregulated urbanization (within favelas and other slum-like neighborhoods,) is an important as socio-economic characteristics to determine vulnerability. Finally, we find that we can validate the Rio de Janeiro heat vulnerability index against excess deaths during heat waves; specifically, we use three types of regressions coupled with difference in difference calculations to show this is indeed a heat vulnerability index as opposed to a social vulnerability index. This suggests it is necessary to add a step to existing methods for validation of heat vulnerability indices, that of the difference-in-difference calculation.

These findings may have an important impact in policy for developing countries. While socio-economic risk may change only with long term policies, urban form may change more rapidly with re-urbanization policies. Therefore, reducing heat vulnerability might add up to the many benefits of urbanizing favelas and slum areas. In developing countries, urbanization is likely more effective and responsive in the short term. For example, urbanization policies such as paving streets, building proper sanitary sewer and arborizing favelas will reduce the vulnerability of the population living there. It may even be that, as

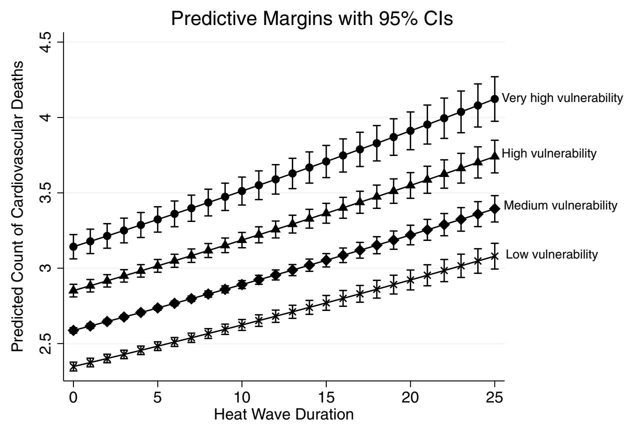

Fig. 4 The first derivative of the response (number of cardiovascular deaths) with respect of the length of heat waves for different levels of vulnerability levels

Table 3 Difference in difference model estimates for a 25-day heat wave between low and very highly vulnerable *setores censitário*

Heat wave length (days)	Low HVI	Very high HVI	Difference
0	2.35	3.14	0.79
25	3.08	4.12	1.04
Difference	0.73	0.98	0.25

shown in a flood study in the United States [39], that the order of these policies does not matter; rather the environmental justice disparities might be sufficiently high that any action will be helpful.

Given that our findings suggest there may be some reason to, within a city, separate favelas from other areas, this suggests research into improved understanding of the spatial autocorrelation within the city and how it may be affecting heat vulnerability and deaths. For instance, although the factor analysis takes into account similarities between regions, other methods exist to investigate similarities between regions (e.g., see summary texts on spatial autocorrelation such as [18, 19]). A recent report has found evidence of spatial clustering at the county level (which reflects urban and rural areas, [41]); other reports suggest there may be spatial clustering in Latin American cities (e.g., [34]). Thompson et al. [59] describe other methods to calculate vulnerability, and find that if one is limited to county level data, that a hierarchical generalized linear regression model with multiscalar indicators and spatial components performs better than methods that lack consideration of spatial dynamics. While our study did not face the limitation of aggregating to the county level (recall the *setores censitário* is of similar size to a census tract in the United States), due to the stark differences between very closely neighboring areas, it may be of interest to explore other vulnerability method calculations to determine whether a model that performs even better can be found.

Alternatively, consider that our distance metric (the way to get from here to there) was distance along roads as calculated using GIS network analyst tools. This may be an appropriate metric when driving or taking public transit to a local hospital. However, recall our findings regarding favelas; people living in favelas might travel to the hospital might in a qualitatively manner different from those living in more wealthy communities. Alternative methods to calculate distance might be able to more closely model actual behavior in favelas, and thus improve model fidelity.

5 Conclusion

In this paper, we investigated how to construct and validate a heat vulnerability index given uncertainty ranges in data for the city of Rio de Janeiro. First, we compare excess deaths of certain types of circulatory diseases during heat waves. Second, we use demographic and environmental data and factor analysis to construct a set of unobserved factors and respective weightings related to heat vulnerability, including a Monte Carlo analysis to represent the uncertainty ranges assigned to the input data. Finally, we use distance to hospital and clinics and their health record data as an instrumental variable to validate our factors. We find that we can validate the Rio de Janeiro heat vulnerability index against excess deaths during heat waves; specifically, we use three types of regressions coupled with difference in difference calculations to show this is indeed a heat vulnerability index as opposed to a social vulnerability index. The factor analysis identifies two factors that contribute to >70% of the variability in the data; one socioeconomic factor and one urban form factor. This suggests it is necessary to add a step to existing methods for validation of heat vulnerability indices, that of the difference-in-difference calculation.

Acknowledgements This work was supported by the National Science Foundation's funded SHADE: Surface Heat Assessment for Developed Environments (PREEVENTS- 1664091).

Authors' contributions D.P. did the majority of the data collection, analysis, and generation of results. K.K. created the research questions and mentored D.P. through the writing.

Funding This work was supported by the National Science Foundation's funded SHADE: Surface Heat Assessment for Developed Environments (PREEVENTS-1664091).

Availability of data and material Data available from authors upon request.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest or competing interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Authors consent for this to be published in Climatic Change (if reviewers and editors approve).

Code availability The majority of the work was completed in GIS and Stata.

References

1. Aldrich N, Benson WF (2008) Disaster preparedness and the chronic disease needs of vulnerable older adults. Prev Chronic Dis 5(1):1–7
2. Alvares CA et al (2013) Köppen's climate classification map for Brazil. Meteorol Z 22(6):711–728
3. Bao J, Li X, Yu C (2015) The construction and validation of the heat vulnerability index, a review. Int J Environ Res Public Health 12(7):7220–7234
4. Barrow MW, Clark KA (1998) Heat-related illnesses. Am Fam Physician 58(3):749–756
5. Basagana X, Sartini C, Barrera-Gomez J et al (2011) Heat waves and cause-specific mortality at all ages. Epidemiology. 22(6):765–772
6. Basu R (2009) High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. Environ Health 8(1):40
7. Bathiany S, Dakos V, Scheffer M, Lenton TM (2018) Climate models predict increasing temperature variability in poor countries. Sci Adv 4(5):eaar5809
8. Bell ML et al (2008) Vulnerability to heat-related mortality in Latin America: a case-crossover study in Sao Paulo, Brazil, Santiago, Chile and Mexico City, Mexico. Int J Epidemiol 37(4):796–804
9. Bieger A, Borges G, Kranz S, McGowan C, Meehan K, Mancuso LG, de Macedo Guimarães LB (2009) Increasing the efficiency of a Brazilian Emergency response Call Center. In: 2009 systems and information engineering design symposium. IEEE, pp 125–130
10. Bradford K, Abrahams L, Hegglin M, Klima K (2015) A heat vulnerability index and adaptation solutions for Pittsburgh, Pennsylvania. Environ Sci Technol 49(19):11303–11311
11. Braga ALF, Zanobetti A, Schwartz J (2002) The effect of weather on respiratory and cardiovascular deaths in 12 US cities. Environ Health Perspect 110(9):859–863
12. Breslin K (1994) Global climate change: beyond sunburn. Environ Health Perspect 102(5):440–443
13. Card D, Fenizia A, Silver D (2018) The health effects of cesarean delivery for low-risk first births. No. w24493. National Bureau of Economic Research
14. Cesar L, Libonati R, Trigo R, Peres L, de Avelar M, Lucas J (2017) Relation between Heat Wave events and mortality rates for the State of Rio de Janeiro. VIISIC Simposio Internacional de Climatologia. September 18–22
15. Chang RR, Klitzner TS (2002) Can regionalization decrease the number of deaths for children who undergo cardiac surgery? A theoretical analysis. Pediatrics 109(2):173–181
16. Chuang WC, Gober P (2015) Predicting hospitalization for heat-related illness at the census-tract level: accuracy of a generic heat vulnerability index in phoenix, Arizona (USA). Environ Health Perspect 123(6):606–612
17. CIMSS (2016) What is a heat wave? The Weather Guys http://wxguys.ssec.wisc.edu/2016/07/25/heat-wave/. Last accessed 23 Feb 2020
18. Cliff AD, Ord JK (1981) Spatial processes: models and applications. Taylor & Francis, London
19. Cressie N (2015) Statistics for spatial data. John Wiley & Sons, New York
20. Curriero FC, Heiner KS, Samet JM, Zeger SL, Strug L, Patz JA (2002) Temperature and mortality in 11 cities of the Eastern United States. Am J Epidemiol 155(1):80–87. https://doi.org/10.1093/aje/155.1.80
21. Dafny L (2009) Estimation and identification of merger effects: an application to hospital mergers. J Law Econ 52(3):523–550
22. Dash SK, Mamgain A (2011) Changes in the frequency of different categories of temperature extremes in India. J Appl Meteorol Climatol 50(9):1842–1858. https://doi.org/10.1175/2011JAMC2687
23. de Castro Martins Ferreira L et al (2019) Ambient temperature and mortality due to acute myocardial infarction in Brazil: an ecological study of time-series analyses. Sci Rep 9(1):1–10
24. Ellis FP (1972) Mortality from heat illness and heat-aggravated illness in the United States. Environ Res 5(1):1–58
25. Frich A, Alexander LV, Della-Marta P, Gleason B, Haylock M, Tank AMGK, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19:193–212
26. Geirinhas JL et al (2018) Climatic and synoptic characterization of heat waves in Brazil. Int J Climatol 38(4):1760–1776
27. Glickman TS (2000) Glossary of meteorology. American Meteorological Society, Boston. isbn:978-1-878220-49-3
28. Glossary – NOAA's National Weather Service. Weather.gov. 25 June 2009. Retrieved 17 July 2013
29. Gosling S, Lowe J, McGregor G, Pelling M, Malamud B (2009) Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. Clim Chang 92(3):299–341
30. Gragnolati M, Lindelow M, Couttolenc B (2013) Twenty years of health system reform in Brazil: an assessment of the Sistema Único de Saúde. The World Bank, Washington, DC
31. Harlan SL, Brazel AJ, Prashad L, Stefanov WL, Larsen L (2006) Neighborhood microclimates and vulnerability to heat stress. Soc Sci Med 63(11):2847–2863
32. Harlan SL, Declet-Barreto JH, Stefanov WL, Petitti DB (2013) Neighborhood effects on heat deaths: social and environmental predictors of vulnerability in Maricopa county, Arizona. PLoS One 121(2):197–204

33. Heat Index. Weather.gov. Available at https://www.weather.gov/safety/heat-index. Retrieved 30 June 2019
34. Inostroza L, Palme M, de la Barrera F (2016) A heat vulnerability index: spatial patterns of exposure, sensitivity and adaptive capacity for Santiago de Chile. PLoS One 11(9):e0162464
35. Instituto Brasileiro de Geografia e Estatísitica. Data from "IBGE | Censo Demográfico 2010". Available at http://censo2010.ibge.gov.br. Deposited 13 June 2019
36. Instituto Nacional de Meteorologia, Data from "BDMEP - Banco de Dados Meteorológicos para Ensino e Pesquisa". Available at http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. Deposited 13 June 2019
37. Kahn JM, Ten Have TR, Iwashyna TJ (2009) The relationship between hospital volume and mortality in mechanical ventilation: an instrumental variable analysis. Health Serv Res 44(3):862–879
38. Kim Y, Joh S (2006) A vulnerability study of the low-income elderly in the context of high temperature and mortality in Seoul. Korea Sci Total Environ 371(1–3):82–88
39. Klima K, El Gammal E, Kong D, Prosdocimi D (2019) Creating a water risk index to improve community resilience. IBM J Res Dev 64(1/2):16–11
40. Kottek M et al (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15(3):259–263
41. Lehnert EA, Wilt G, Flanagan B, Hallisey E (2020) Spatial exploration of the CDC's Social Vulnerability Index and heat-related health outcomes in Georgia. Int J Disaster Risk Reduct 46:101517
42. Madrigano J, Ito K, Johnson S, Kinney PL, Matte T (2015) A case-only study of vulnerability to heat wave–related mortality in New York City (2000–2011). Environ Health Perspect 123(7):672–678
43. Maier G et al (2014) Assessing the performance of a vulnerability index during oppressive heat across Georgia, United States. Weather Clim Soc 6(2):253–263
44. Marti-Soler H et al (2014) Seasonal variation of overall and cardiovascular mortality: a study in 19 countries from different geographic locations. PLoS One 9(11):e113500
45. Mehrotra S, Bardhan R, Ramamritham K (2018) Urban informal housing and surface urban Heat Island intensity: exploring spatial association in the city of Mumbai. Environ Urban ASIA 9(2):158–177
46. Melillo JM, Richmond TT, Yohe G (2014) Climate change impacts in the United States. Third national climate assessment, p 52
47. Metzger KB, Ito K, Matte TD (2010) Summer heat and mortality in New York City: how hot is too hot? Environ Health Perspect 118(1):80–86
48. Ministério da Saúde do Brasil, Data from "Departamento de Informática do Sistema Único de Saúde (DATASUS)." Available at: http://www.datasus.gov.br. Deposited 13 June 2019
49. Nunnally JC, Bernstein IH (1994) Psychological theory. MacGraw-Hill, New York, NY, pp 131–147
50. Pett MA, Lackey NR, Sullivan JJ (2003) Making sense of factor analysis: the use of factor analysis for instrument development in health care research. Sage, Thousand Oaks
51. Pezza AB, Rensch P, Cai WJ (2012) Severe heat waves in Southern Australia: synoptic climatology and large scale connections. Clim Dyn 38:209–224
52. Reid CE et al (2009) Mapping community determinants of heat vulnerability. Environ Health Perspect 117(11):1730–1736
53. Reid CE et al (2012) Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study. Environ Health Perspect 120(5):715–720
54. Rothfusz LP (1990) The heat index equation. National Weather Service Technical Attachment (SR 90–23)
55. Schwartz J (2005) Who is sensitive to extremes of temperature? A case only analysis. Epidemiology. 16(1):67–72
56. Sherwood S (2012) Exceedence of heat index thresholds for 15 regions under a warming climate using the wet-bulb globe temperature. Int J Climatol 32:161–177
57. Son J-Y et al (2016) The impact of temperature on mortality in a subtropical city: effects of cold, heat, and heat waves in São Paulo, Brazil. Int J Biometeorol 60(1):113–121
58. Stewart S, Keates AK, Redfern A, McMurray JJV (2017) Seasonal variations in cardiovascular disease. Nat Rev Cardiol 14(11):654
59. Thompson CM, Dezzani RJ, Radil SM (2019) Modeling multiscalar influences on natural hazards vulnerability: a proof of concept using coastal hazards in Sarasota County, Florida. GeoJ:1–22
60. Turner LR, Barnett AG, Connell D, Tong S (2012) Ambient temperature and cardiorespiratory morbidity: a systematic review and meta-analysis. Epidemiology 23(4):594–606
61. Willett KM, Sherwood S (2012) Exceedance of heat index thresholds for 15 regions under a warming climate using the wet-bulb globe temperature. Int J Climatol 32:161–177
62. Wolf T, McGregor G, Analitis A (2014) Performance assessment of a heat wave vulnerability index for greater London, United Kingdom. Weather Clim Soc 6(1):32–46
63. Wu P, Lin C, Lung S, Guo H, Chou C, Su H (2011) Cardiovascular mortality during heat and cold events: determinants of regional vulnerability in Taiwan. Occup Environ Med 68(7):525–530
64. Zhang DL, Shou YX, Dickerson RR (2009) Upstream urbanization exacerbates urban heat island effects. Geophys Res Lett 36(24):L24401
65. Zhang K, Wang R, Shen C, Da L (2010) Temporal and spatial characteristics of the urban heat island during rapid urbanization in Shanghai. China. Environ Monit Assess 169(1–4):101–112