American Journal of Epidemiology

Associations between Flood Risk and United States Census Tract-Level Health Outcomes

Journal:	*American Journal of Epidemiology*
Manuscript ID	AJE-00535-2023.R1
Manuscript Type:	Original Contribution
Key Words:	Bayesian spatial analysis, climate change, social vulnerability

Associations between Flood Risk and United States Census Tract-Level Health Outcomes

Alvin Sheng[1], Brian J Reich[1], and Kyle P Messier[2,3]

[1]North Carolina State University, Department of Statistics
[2]National Institute of Environmental Health Sciences, Division of Translational Toxicology, Predictive Toxicology Branch
[3]National Institute of Environmental Health Sciences, Division of Intramural Research, Biostatistics and Computational Biology Branch

Abstract

Human-induced climate change has led to more frequent and severe flooding throughout the globe. We examined the association between flood risk and the prevalence of coronary heart disease, high blood pressure, asthma, and poor mental health in the United States, while taking into account different levels of social vulnerability. We aggregated flood risk variables from First Street Foundation by census tract and used principal component analysis to derive a set of five interpretable flood risk factors. The dependent variables were census-tract level disease prevalences generated by the Centers for Disease Control and Prevention. Bayesian spatial conditional autoregressive models were fit on this data to quantify the relationship between flood risk and health outcomes under different stratifications of social vulnerability. We showed that three flood risk principal components had small but significant associations with each of the health outcomes, across the different stratifications of social vulnerability. Our analysis gives the first United States-wide estimates of the associated effects of flood risk on specific health outcomes. We also show that social vulnerability is an important moderator of the relationship between flood risk and health outcomes. Our approach can be extended to other ecological studies that examine the health impacts of climate hazards.

Keywords: Bayesian spatial analysis, flood risk, health outcomes, social vulnerability.

1 Introduction

According to the IPCC, it is unequivocal that human influence has led to widespread and rapid changes in the atmosphere, ocean, cryosphere, and biosphere [18]. The global mean sea level increased by 0.20 m (90% confidence interval [0.15 m, 0.25 m]) between 1901 and 2018, and human influence was very likely (above 90% likelihood, using IPCC's calibrated language for quantified uncertainty [19]) to be the main driver of the increase since at least 1971. Furthermore, human-induced climate change is affecting many weather and climate extremes throughout the globe. For example, human influence is likely (above 66% likelihood, using IPCC's calibrated language) to be the main driver of the increase in the frequency and intensity of heavy precipitation events since the 1950s.

There are several observed health impacts of climate change. According to the IPCC, weather and climate extreme events have exposed millions of people to reduced food and water security [21]. Additionally, increasing temperatures, trauma from weather and climate extreme events, and loss of livelihoods and culture have led to mental health challenges; increased exposure to wildfire smoke, atmospheric dust, and aeroallergens have been associated with cardiovascular and respiratory distress. The vulnerability to these climate hazards is exacerbated by inequity related to gender, ethnicity, low income, and current and historical forms of slavery and colonialism [21].

There is extensive literature on the health effects of acute climate hazards, especially flooding. Du et al. [9] conducted a review of more than 500 references to classify the immediate, medium-term, and long-term health impacts of floods. To our knowledge, however, there is no literature on the health effects of the burden of flood risk. Related to the health effects of persistent flood risk are the long-term, indirect effects of floods like malnutrition and psychological distress arising from social and economic disruption [9]. Health effects from flood risk may also arise via eco-anxiety, i.e., a "chronic fear of environmental doom" [8], or allostatic load, i.e., the cost of chronic stress responses toward repeated challenges like climate hazards [15]. These complex, potentially causal pathways justify the need to quantify health effects associated with flood risk, and the risk of climate hazards more generally.

A frequently-used approach for quantifying associations between climate hazards and health outcomes is a cross-sectional study design. Because the data is often provided in the form of spatially aggregated units with this design, statistical analysis must take into account residual spatial correlation, i.e., spatial correlation of the response variable among nearby units not accounted for by the covariates. For instance, to measure the association between dengue fever counts and weather variables and the socioeconomic index, Hu et al. [13] fitted a Bayesian hierarchical model that included spatial terms to account for residual correlation among neighboring local government areas. In this study, we used the same modeling approach as in Hu et al. [13], i.e., Bayesian conditional autoregressive models.

As our contribution to the literature on the health effects of climate hazards, we measured associations between flood risk and prevalences of coronary heart disease (CHD), high blood pressure (HBP), asthma, and poor mental health (PMH). We used flood risk data from a model developed by the First Street Foundation (FSF), which provides flood risk estimates at the property level throughout the United States (US) [11]. We also accounted for secondary exposures such as social vulnerability, air pollution, and weather. Finally, we stratified the analysis on selected factors of social vulnerability to examine interactions between the associated effects of flood risk and social inequity.

2 Methods

Study area. The analytic units are the census tracts of the contiguous US, of which there are 72,539 in total, based on the 2010 Census [25]. Each census tract generally has a population between 1,200 and 8,000 people, with the optimum size being 4,000 people according to the US Census Bureau's objective of providing a stable set of geographical units for the presentation of statistics. [1]. Census tract statistics for the total population, number of housing units, and number of households are shown in Table 1, which are taken from the 2014-2018 American Community Survey (ACS) [4]. Statistics on the number of neighboring census tracts are also shown, which are given by Spatial Structures in the Social Sciences, Brown University [23]. Less than 1% of census tracts had missing health outcome data or adjacency information.

Response variables. The prevalences (in %) of four health outcomes among adults at the census tract level were estimated by the Centers for Disease Control and Prevention (CDC). The prevalences are small area estimates based on the Behavioral Risk Factor Surveillance System (BRFSS) 2017 or 2018 data [6]. To acquire estimates for small areas such as census tracts when the survey was done over larger areas, the CDC used a multilevel statistical modeling framework [7]. See Table 2 for the summary statistics and Figure S1 for the choropleth maps of the small area estimates for each outcome.

Flood risk variables. FSF has developed climate risk models for flooding (including rainfall, river flooding, and coastal surge flooding) at the property level throughout the US [10]. Their risk estimates are used by private individuals, industries, and the government to determine the potential impact of flooding on their homes, housing market inefficiencies, and vulnerable communities, etc. [11].

For each property, the FSF flood risk model estimates the annual probability of flooding in the current year (2021) or in the climate-adjusted future (2051), based on IPCC Representative Concentration Pathway (RCP) 4.5 [10]. IPCC RCP 4.5 represents the middle-ground scenario of carbon emissions considered by the IPCC. This scenario assumes that atmospheric carbon dioxide levels will stay below 550 ppm during this century, leading to less than a 2°C increase of the global average temperature. The model also estimates a Flood Factor (FF) score for the property, which is an integer indicating the level of flood risk in terms of the likelihood of flooding and the depth of that flood. It ranges from 1 to 10, 1 being minimal flood risk and 10 being extreme flood risk. We aggregated these variables by census tract by calculating the proportion of properties in the census tract that has greater than a given annual probability of flooding (0.2%, 1%, or 20%) in a given year (2021 or 2051). We also calculated the average, standard deviation, and coefficient of variation of the FF scores for the properties in the census tract, as well as conditional averages (average of scores between 2 and 10; average of scores for properties with at least a 0.2% or 1% annual probability of flooding in 2021). Finally, we calculated the proportion of properties in the census tract with a given FF score, ranging from 1 to 10. See Table 2 for the summary statistics of the 22 flood risk variables.

Other variables. We obtained the census tract-level social vulnerability index (SVI), which is generated by the CDC from the 2014-2018 ACS and developed to reflect a community's resilience to natural disasters or human-made hazardous events [5]. The SVI includes covariates related to age, i.e., the percentage of persons aged 17 and younger and the per-

Variable	Mean	S.d.	Min	Max
Total population in census tract	4434	2275	0	70271
Number of housing units	1874	911	0	26436
Number of households	1645	809	0	21337
Number of neighboring census tracts	6.301	2.09	1	49

Table 1: Summary statistics (mean, standard deviation, mean, and maximum) for US census tracts.

centage of persons aged 65 and older in the census tract. There are four themes of SVI, each consisting of multiple variables: socioeconomic status, household composition, minority status/language, and housing/transportation (see Supplementary Materials Table 1).

We also obtained data on the annual average concentrations of six pollutants, CO, NO_2, O_3, $PM10$, $PM2.5$, and SO_2 [14], daily maximum temperature and maximum relative humidity [2], and the prevalence of smoking among adults in 2018 [6]. We averaged the pollutant concentrations across the years 2000-2016 and averaged the daily maximum temperature and relative humidity across the years 2005-2020 for each census tract. Summary statistics for these variables are shown in Supplementary Materials Table 1.

Data analysis. We reduced 22 flood risk variables to 5 principal components via principal component analysis (PCA) such that 80% of the variance was accounted for. We standardized each flood risk variable to have a mean of zero and a standard deviation of one before the PCA, so that the flood risk variables contribute equally to the PCA. We reduced 6 pollutant concentrations to 3 principal components in the same manner. Then, we standardized the PCs and other covariates to have a mean of zero and a standard deviation of one, so that the PCA loadings are comparable across principal components (see Supplementary Materials Table 2 and Supplementary Materials Table 3) and the regression coefficients can be meaningfully compared across all variables.

Across the covariates and response variables, a median of 0.97% of values were missing and a maximum of 7.36% of values were missing. We imputed the missing values of each covariate with the mean of that covariate. The models described below will predict the missing values of the responses.

Separate linear regression models were developed in a Bayesian framework for each of the four prevalences of health outcomes: CHD, HBP, asthma, and PMH. The Bayesian hierarchical model (BHM) for the observed prevalence Y_k for the kth census tract, $k = 1, \ldots, 72539$, is given by

$$Y_k \sim N(\mathbf{x}_k^\intercal \boldsymbol{\beta} + \phi_k, \nu^2) \tag{1}$$
$$\mathbf{x}_k^\intercal \boldsymbol{\beta} = \sum_j x_{kj}\beta_j,$$

where $\mathbf{x}_k$ is a vector containing the intercept and covariates for census tract k, $\boldsymbol{\beta}$ is a vector of the corresponding coefficients, and ϕ_k is a spatially structured random effect for census tract k. Linear regression is a straightforward approach that is valid for this application. First, the response variables are derived from small-area estimation rather than raw counts, for which Poisson regression is often used [7]. Second, the examination of the confidence

Variable	Mean	S.d.	Min	Max
Health Outcomes				
Percentage of adults with coronary heart disease	6.67	2.21	0.50	36.00
Percentage of adults with high blood pressure	32.35	7.30	4.90	70.30
Percentage of adults with asthma	9.90	1.58	5.40	20.60
Percentage of adults with poor mental health	14.26	3.41	5.20	35.50
Flood Risk				
Percentage of properties in 2020 with greater than 0.2% annual probability of flooding	17.08	20.71	0	100
Percentage of properties in 2050 with greater than 0.2% annual probability of flooding	18.57	22.42	0	100
Percentage of properties in 2020 with greater than 1% annual probability of flooding	10.97	15.29	0	100
Percentage of properties in 2050 with greater than 1% annual probability of flooding	12.35	17.63	0	100
Percentage of properties in 2020 with greater than 20% annual probability of flooding	2.88	7.59	0	100
Percentage of properties in 2050 with greater than 20% annual probability of flooding	3.55	9.71	0	100
Average FF score	1.86	1.14	1	10
Standard deviation of FF score	1.53	0.77	0.00	6.36
Coefficient of variation of FF score	0.86	0.32	0.00	1.50
Average FF score between 2-10	5.64	1.35	2	10
Average FF score of properties in 2020 with greater than 0.2% annual probability of flooding	5.84	1.32	2	10
Average FF score of properties in 2020 with greater than 1% annual probability of flooding	6.76	1.11	3	10
Percentage of properties with FF score 1	81.41	22.42	0	100
Percentage of properties with FF score 2	1.03	3.63	0	100
Percentage of properties with FF score 3	2.71	6.14	0	100
Percentage of properties with FF score 4	3.61	9.44	0	100
Percentage of properties with FF score 5	1.27	3.12	0	100
Percentage of properties with FF score 6	4.81	8.49	0	100
Percentage of properties with FF score 7	1.52	2.58	0	100
Percentage of properties with FF score 8	0.31	1.09	0	100
Percentage of properties with FF score 9	1.68	4.67	0	100
Percentage of properties with FF score 10	1.64	6.21	0	100

Table 2: Summary statistics (mean, standard deviation, minimum, and maximum) for health outcomes and flood risk variables.

bounds of the model-based responses indicates that the variability of estimated responses among census tracts are similar, regardless of their population sizes. Lastly, even though linear regression assumes that the response is not bounded, exploration of the data suggests that the distribution of each response can be reasonably modeled as Gaussian within the ranges of the covariates considered, in that it is approximately normal (as indicated by the qq-plots) and far from the extremes of 0 and 100 in terms of the standard deviation.

The random effects ϕ_k are residuals that exhibit spatial autocorrelation, after accounting for the flood risk and secondary exposures. Its distribution is defined by the following full conditional distribution:

$$\phi_k|\phi_i, i \neq k \sim N\left(\frac{\sum_{i=1,i\neq k}^{K} w_{ki}\phi_i}{\sum_{i=1,i\neq k}^{K} w_{ki}}, \frac{\tau^2}{\sum_{i=1,i\neq k}^{K} w_{ki}}\right), \tag{2}$$

where w_{ki} is 1 if census tracts k and i share a boundary and 0 if they do not share a boundary. The adjacencies among the census tracts are given by Spatial Structures in the Social Sciences, Brown University [23]. This is the intrinsic conditional autoregressive (CAR) prior [3]. The combination of this prior with a spatially unstructured random error, included in the model through the variance parameter ν^2, makes the full model (Equation 1) a convolution model as proposed by Besag et al. [3]. More flexible priors were tried; however, they do not scale well to large spatial data that encompass the contiguous US. These priors were compared to the intrinsic CAR prior for subsets of the US, but the results had negligible differences. The Markov Chain Monte Carlo (MCMC) algorithm for the above BHM is implemented in the R package `CARBayes` [16].

To examine how social vulnerability moderates the associated effects of flood risk, we stratified the models by dichotomizing the data according to the median of a given stratification variable based on the SVIs. We considered six stratification variables: percentage of persons below poverty level, percentile rank for socioeconomic status (SVI Theme 1), percentile rank for household composition (SVI Theme 2), percentile rank for minority status/language (SVI Theme 3), percentile rank for housing/transportation (SVI Theme 4), and the overall percentile rank of social vulnerability (all SVI Themes). The CDC computes the percentile rank for a given theme in a census tract by calculating the percentile rank for each variable within the theme, adding the percentile ranks, and finding the percentile rank of the sum [4]. See Supplementary Materials Table 1 for the variables in each theme. To verify whether stratification by the median leads to distinct levels of social vulnerability, we examined the distributions of the SVIs that make up the stratification variables. As shown by Supplementary Materials Figure S2, most SVIs are skewed, and so there is a large distinction between census tracts below the median and those above the median for each stratification variable.

For each stratification variable across the census tracts, $\mathbf{u}_l, l = 1, \ldots, 6$, we stratified the model by including an interaction between every variable and the lower stratum indicator $I(u_{kl} \leq Med(\mathbf{u}_l))$ and the upper stratum indicator $I(u_{kl} > Med(\mathbf{u}_l))$, where $Med(\mathbf{u}_l)$ is the median of $\mathbf{u}_l$. The low SV strata corresponding to the six stratification variables indicate lower social vulnerability, and the upper SV strata indicate higher social vulnerability. The set of variables corresponding to the theme used to stratify the data is not included as covariates in the model. Then, the linear predictor in the BHM (Equation 1) for stratification variable l is augmented as follows:

$$\mathbf{x}_{kl}^{\intercal}\boldsymbol{\beta}_l = \sum_j x_{kj} I(u_{kl} \leq Med(\mathbf{u}_l))\beta_{lj0} + \sum_j x_{kj} I(u_{kl} > Med(\mathbf{u}_l))\beta_{lj1} \tag{3}$$

In total, there are 24 models for the four health outcomes and six stratification variables. For each model, we ran the MCMC algorithm for three chains, where each chain was run for 10,000 burn-in iterations that were discarded and 100,000 iterations that were thinned by 2. In total, for each BHM, 150,000 iterations were used for inference. Convergence was confirmed by examining the effective sample sizes, trace plots, and posterior density plots for the parameters.

The BHMs were used to find posterior medians and 95% credible intervals (CI) for the $\boldsymbol{\beta}$ coefficients for the flood risk PCs. To emphasize that inferences for primary and secondary exposures should be interpreted differently [26], we provide figures of results for the non-flood risk covariates separately in the Supplementary Materials.

For the sensitivity analyses, several adjustments were made to the BHMs stratified on all SVI Themes to determine the robustness of findings. The first sensitivity analysis removes the SVIs from the models instead of stratifying on them; the second sensitivity analysis does not include the spatially structured random effects ϕ_k in the models; and the last sensitivity analysis uses flood factor groupings rather than flood risk PCs in the models. Sensitivity analyses results are provided in Supplementary Materials Section 2.

All analyses were carried out in the statistical software R, version 4.1.2 [22]. See Supplementary Materials Section 3 for the R packages used in this analysis. The following GitHub repository contains the code used to generate the analysis and plots for this paper: `https://github.com/AlvinSheng/flood-risk-health-effects`.

3 Results

To reduce collinearity and maintain interpretability of parameters, dimension reduction via principal component analysis (PCA) was performed on the flood risk variables. PCs with negative correlation to flood risk were multiplied by -1 such that all final PCs are positively correlated with higher flood risk. PCA loadings are available in Supplementary Materials Table 2. Conveniently, the first 5 PCs that explain 80% of the variability in flood risk have unique and complementary interpretations. To aid further discussion, we give each PC a short nickname. (1) PC 1, referred to as ‘consistent flood risk’, is positively correlated with any indicator of flood risk. Only “Percentage of properties with FF score 1” and “coefficient of variation of FF score” have negative loadings. (2) PC 2, referred to as ‘average FF score’, emphasizes the average FF score variables with the largest positive loadings. (3) PC 3, referred to as ‘severe flood risk’, is positively correlated with common or severe flood risk, highlighting “Percentage of properties with greater than 20% annual probability of flooding” and “Percentage of properties with FF score 10” with large positive loadings. (4) PC 4, referred to as ‘high/low risk difference’, indicates the difference between severe flood risk percentages and mild flood risk percentages: percentages corresponding to FF scores

6-8 have large positive loadings, while percentages corresponding to FF scores 2 and 3 have large negative loadings. (5) PC 5, referred to as 'skewed flood risk', emphasizes the low and high extremums of flood risk with large positive loadings for FF scores 2-3 and 7-9 and large negative loadings for FF scores 4-6. Supplementary Materials Section 1 contains additional exposition on the interpretation of the flood risk PCs.

Figure 1A shows the choropleth map of 'consistent flood risk' (PC 1) in the contiguous US. Most census tracts have low or negative consistent flood risk. However, coastal regions in states such as Louisiana and Florida tend to have higher consistent flood risk. There are also inland census tracts with high consistent flood risk, especially those in Central Appalachia. There are census tracts with missing flood risk variables (e.g., those in the Midwest and Mountain West); the health outcomes for these census tracts are estimated through the mean imputation of missing variables and the smoothing of health outcomes among neighboring census tracts via the CAR prior.

Figure 1B-E shows choropleth maps of predicted prevalences for the health outcomes, derived from BHMs stratified on all SVI Themes. Each choropleth map displays the prediction $\mathbf{x}_{k6}^{\intercal}\boldsymbol{\beta}_6 + \phi_k$ for each census tract k; this is the mean parameter of the Gaussian in Equation 1 but with the augmented linear predictor in Equation 3. Figure S3 in the Supplementary Materials shows the spatially structured random effects ϕ_k for the health outcomes. None of the smoothed prevalences shown in Figure 1B-E lie outside of the interval $[0, 100]$; additionally, scatterplots between residuals and smoothed prevalences and qq-plots for the residuals indicate that the assumptions of a linear association between the response and covariates, as well as Gaussian and homoskedastic errors, are met. Therefore, modeling prevalences with a linear model is valid for our application. The CHD choropleth map is characterized by a prevalence of around 10% in many census tracts, interspersed with clusters of census tracts with low prevalences. The HBP choropleth map has a similar spatial pattern as the CHD choropleth map, but with relatively smaller differences between census tracts. Choropleth maps for asthma and PMH have similar spatial patterns, with especially high prevalences for census tracts in Arizona and New Mexico.

For cardiovascular health outcomes, 'consistent flood risk' (PC 1), 'average FF score' (PC 2), and 'high/low risk difference' (PC 4) were the most important flood risk factors. For CHD, 'consistent flood risk' (PC 1) was significantly associated with increased prevalence for all high SV strata, but only a few low SV strata; 'average FF score' (PC 2) was significantly associated with increased prevalence for most of the low SV strata, but none of the high SV strata 2a. For HBP, 'average FF score' (PC 2) was significantly associated with increased prevalence for most of the low SV strata; 'high/low risk difference' (PC 4) was significantly associated with decreased prevalence for all low SV strata (Figure 2b).

For the other two health outcomes, 'consistent flood risk' (PC 1) and 'high/low risk difference' (PC 4) were the most important flood risk factors. For asthma, 'consistent flood risk' (PC 1) was significantly associated with decreased prevalence for most low SV strata and only a few high SV strata; 'high/low risk difference' (PC 4) was significantly associated with increased prevalence for most low and high SV strata (Figure 2c). For PMH, 'high/low risk difference' (PC 4) was significantly associated with increased prevalence for most low SV strata (Figure 2d).

Figure S4 (in the Supplementary Materials) shows the flood risk linear predictions for the health outcomes, derived from BHMs stratified on all SVI Themes. That is, each choropleth map displays the prediction $\sum_{j=1}^{5} x_{kj} I(u_{k6} \leq Med(\mathbf{u}_6))\beta_{6j0} + \sum_{j=1}^{5} x_{kj} I(u_{k6} > Med(\mathbf{u}_6))\beta_{6j1}$ for each census tract k, where $j = 1, \ldots, 5$ are the indices corresponding to the five flood risk PCs (Equation 3). For CHD and HBP, health effects associated with flood risk were prominent along parts of the coastline, with up to a 1.5 or 3 point increase in prevalence, respectively (Figures S4a and S4b). For Asthma and PMH, health effects associated with flood risk were not as prominent, with less than a 0.2 or 0.45 point change in prevalence, respectively (Figures S4c and S4d).

In the Supplementary Materials, Figures S5, S6, S7, and S8 show the associated effect of secondary exposures on health outcomes, and Section 2 contains results of the sensitivity analyses (Figures S9, S10, S11) and the discussion thereof.

4 Discussion

Our analysis shows the first US-wide estimates of the association between flood risk and human health outcome prevalences, including CHD, HBP, asthma, and PMH. We have shown that flood risk indicators were associated with small but significant impacts on all four health outcomes. Although the health effects associated with flood risk were small relative to those of secondary exposures (Figures S5, S6, S7, and S8), our results show that flood risk is nonetheless an important source of variation for these health outcomes.

The complex phenomena of flood risk can be represented by five PCs. Three of the five flood risk PCs had significant associations with health outcomes that were consistent across six stratifications of the data. Thus, the most important summaries of flood risk with respect to health outcomes were 'consistent flood risk' (PC 1), 'average FF score' (PC 2), and 'high/low risk difference' (PC 4).

Quantifying the interaction between social vulnerability and impacts of flooding is necessary for incorporating social equity into flood mitigation policy. To this end, Tate et al. [24] mapped the hotspots where high flood exposure and high social vulnerability converge in the contiguous US. We provide additional evidence showing social vulnerability to be a significant moderator for the associated effects of flood risk; moreover, our analysis provides continuous estimates of the associated effects within each SV strata while accounting for additional confounders.

There are a few findings of note from the stratified analysis. First, 'consistent flood risk' (PC 1) was significantly associated with decreased prevalence of asthma for most low SV strata (Figure 2c). The low SV strata is associated with higher socioeconomic status and other factors that lead to resilience to environmental hazards; this shows how accounting for interactions with the SVI can explain possibly counterintuitive effects. Second, for HBP, 'severe flood risk' (PC 3) was significantly associated with decreased prevalence for some high SV strata; additionally, 'high/low risk difference' (PC 4) was significantly associated with decreased prevalence for all low SV strata and some high SV strata (Figure 2b). Both PC 3 and PC 4 are positively correlated with severe flood risk indicators but negatively correlated with mild or moderate flood risk indicators (Supplementary Materials Table 2); if mild or moderate flood risk is more informative with regard to worse health outcomes than severe flood risk, then PC 3 and PC 4 may be associated with protective effects. This points to a limitation of using PCA to summarize flood risk: some PCs may not have straightforward

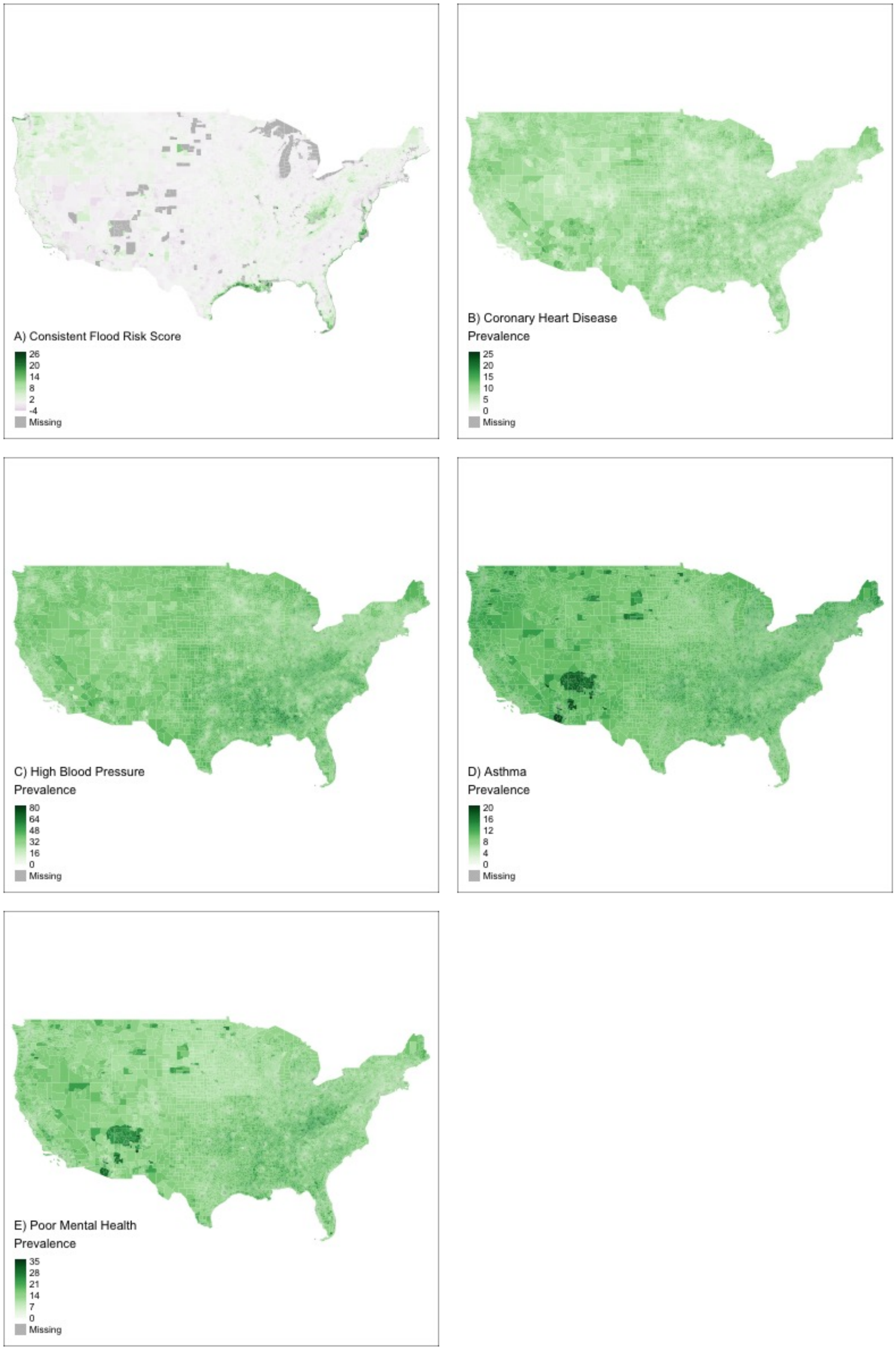
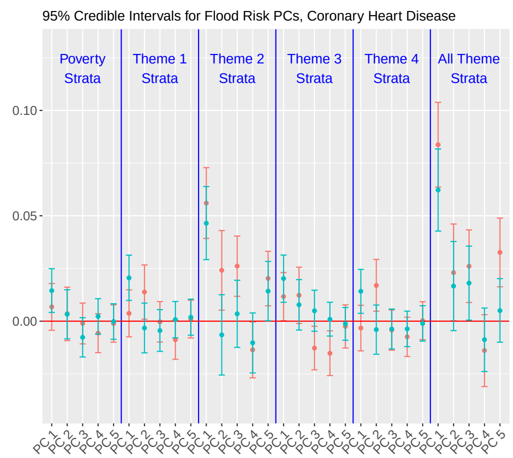
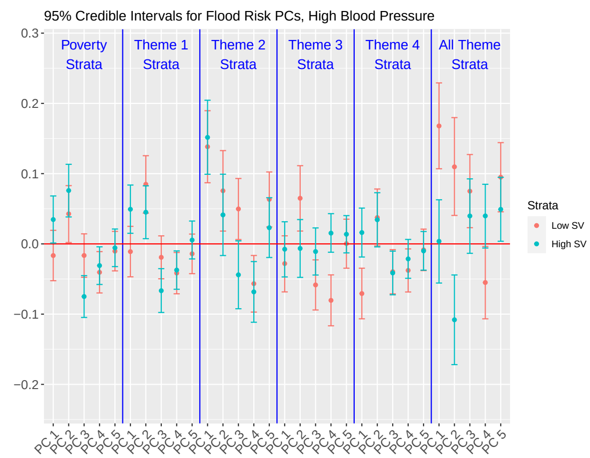


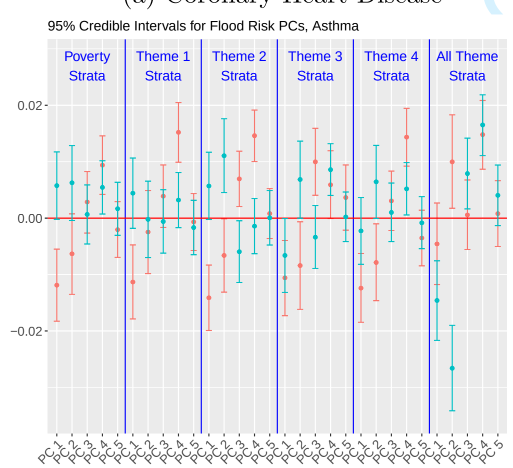
Figure 1: Choropleth maps for Consistent Flood Risk (PC 1) and smoothed prevalences for the four health outcomes, derived from BHMs stratified on all SVI Themes. The maps have a common colorscale with varying tick breaks and labels, which are shown in the top-right. (A) Consistent Flood Risk/PC1 (B-E) Smoothed Predictions of CHD, HBP, asthma, and Poor mental health, respectively.



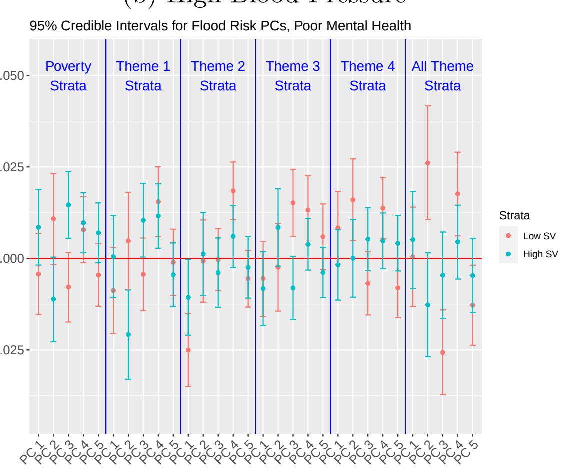
(a) Coronary Heart Disease



(b) High Blood Pressure



(c) Asthma



(d) Poor Mental Health

Figure 2: Credible intervals (95%) of flood risk PCs for each of the four health outcomes and each of the six stratification variables (24 stratified models in total). Each stratified model contains all of the variables as shown in Supplementary Materials Figures S5, S6, S7, and S8, but only coefficients for the flood risk variables are shown here. Theme 1: Socioeconomic Status, Theme 2: Household Composition, Theme 3: Minority Status, Theme 4: Housing Type.

interpretations.

Lastly, models stratified on all SVI Themes occasionally had exaggerated or flipped estimates as compared to the other stratified models. This can be seen by the estimates for 'consistent flood risk' (PC 1) for health outcomes CHD and HBP (Figures 2a and 2b). Also, for each of the health outcomes HBP, Asthma, and PMH, there was a relatively large separation between the 'average FF score' (PC 2) coefficient estimate for the high stratum and that for the low stratum (Figures 2b, 2c, and 2d). The greater variability of coefficient estimates is partly due to having fewer covariates in the models (e.g., models stratified on all SVI Themes had 26 β coefficients each, whereas models stratified on SVI Theme 1 had 46 β coefficients each). Additionally, removing all SVI covariates and only stratifying on the overall percentile rank of social vulnerability has the potential to introduce confounding. Thus, stratifying the data in multiple ways serves to check the robustness of results.

In addition to the limitations of ecological inference [12], another limitation of this study is the temporal mismatch between the exposure and outcomes. Our models estimated associations between past health effects and the risk of future floods. Like environmental justice and social vulnerability indices, flood risk represents a burden that can manifest through psychosocial stress and behavior modifications [17, 20].

In their survey on the health impacts of floods, Du et al. [9] outlined the need for more evidence on the strength of the association between flooding and specific health effects; our work is an important addition to this literature. We have distilled the various measures of flood risk into a few informative dimensions and quantified their association with four health outcomes, stratified on social vulnerability. Overall, flood risk is associated with low magnitude but significant health outcomes at the census tract level, after controlling for plausible confounding and spatial random effects. Through the sensitivity analyses, we show the effects of various adjustments to our model, which highlight some robustness of results and the folly of poorly specified models (e.g., not accounting for the effect-modification of social vulnerability). Our approach can be straightforwardly extended to other ecological studies that examine the health impacts of climate hazards over a large area.

Acknowledgements

This work is supported by the National Institute of Environmental Health Sciences, Division of Translational Toxicology, Division of Intramural Research, and the Spatiotemporal Exposures and Toxicology group under project number ZIA ES103368-02.

References

[1] Census glossary, February 2022. URL `https://www.census.gov/programs-surveys/geography/about/glossary.html`.

[2] J. T. Abatzoglou. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1):121–131, 2012. URL `https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_GRIDMET`.

[3] J. Besag, J. York, and A. Mollié. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*, 43:1–59, 1991.

[4] Centers for Disease Control and Prevention and Agency for Toxic Substances and Disease Registry, Geospatial Research, Analysis, and Services Program. CDC SVI documentation 2018, January 2020. URL https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html.

[5] Centers for Disease Control and Prevention and Agency for Toxic Substances and Disease Registry, Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index 2018 Database US, April 2021. URL https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html.

[6] Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, and Division of Population Health. PLACES: Local data for better health, census tract data 2020 release, January 2021. URL https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/4ai3-zynv.

[7] Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, and Division of Population Health. PLACES: Methodology, 2023. URL https://www.cdc.gov/places/methodology/index.html.

[8] S. Clayton, C. Manning, K. Krygsman, M. Speiser, et al. Mental health and our changing climate: Impacts, implications, and guidance. *Washington, DC: American Psychological Association and ecoAmerica*, 2017.

[9] W. Du, G. J. FitzGerald, M. Clark, and X.-Y. Hou. Health impacts of floods. *Prehospital and Disaster Medicine*, 25(3):265–272, 2010. URL https://doi.org/10.1017/S1049023X00008141.

[10] First Street Foundation. Methodology: Flood model overview, 2022. URL https://firststreet.org/research-lab/published-research/flood-model-methodology_overview/.

[11] First Street Foundation. First Street Foundation mission, 2022. URL https://firststreet.org/mission/.

[12] D. A. Freedman. Ecological inference and the ecological fallacy. *International Encyclopedia of the Social & Behavioral Sciences*, 6(4027-4030):1–7, 1999.

[13] W. Hu, A. Clements, G. Williams, S. Tong, and K. Mengersen. Spatial patterns and socioecological drivers of dengue fever transmission in queensland, australia. *Environmental Health Perspectives*, 120(2):260–266, February 2012.

[14] S.-Y. Kim, M. Bechle, S. Hankey, L. Sheppard, A. A. Szpiro, and J. D. Marshall. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLoS ONE*, 15(2), 2020.

[15] B. Kudielka and C. Kirschbaum. Stress and health research. In N. J. Smelser and P. B. Baltes, editors, *International Encyclopedia of the Social Behavioral Sciences*, pages 15170–15175. Pergamon, Oxford, 2001. ISBN 978-0-08-043076-8. doi: https://doi.org/10.1016/B0-08-043076-7/03818-3. URL https://www.sciencedirect.com/science/article/pii/B0080430767038183.

[16] D. Lee. *CARBayes version 5.2.3: An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors*. University of Glasgow.

[17] N. Maizlish, T. Delaney, H. Dowling, D. A. Chapman, R. Sabo, S. Woolf, C. Orndahl, L. Hill, and L. Snellings. California healthy places index: frames matter. *Public Health Reports*, 134(4):354–362, 2019.

[18] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Climate change 2021: The physical science basis. Working group I contribution to the sixth assessment report of the intergovernmental panel on climate change. Technical report, IPCC, 2021.

[19] M. D. Mastrandrea, C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, E. Kriegler, K. J. Mach, P. R. Matschoss, G.-K. Plattner, G. W. Yohe, and F. W. Zwiers. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties, July 2010. URL https://www.ipcc.ch/site/assets/uploads/2017/08/AR5_Uncertainty_Guidance_Note.pdf.

[20] C. Owusu, B. Flanagan, A. M. Lavery, C. E. Mertzlufft, B. A. McKenzie, J. Kolling, B. Lewis, I. Dunn, E. Hallisey, E. A. Lehnert, et al. Developing a granular scale environmental burden index (ebi) for diverse land cover types across the contiguous united states. *Science of the Total Environment*, 838:155908, 2022.

[21] H.-O. Pörtner, D. C. Roberts, M. M. B. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. Climate change 2022: Impacts, adaptation and vulnerability. Working group II contribution to the sixth assessment report of the intergovernmental panel on climate change. Technical report, IPCC, 2022.

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

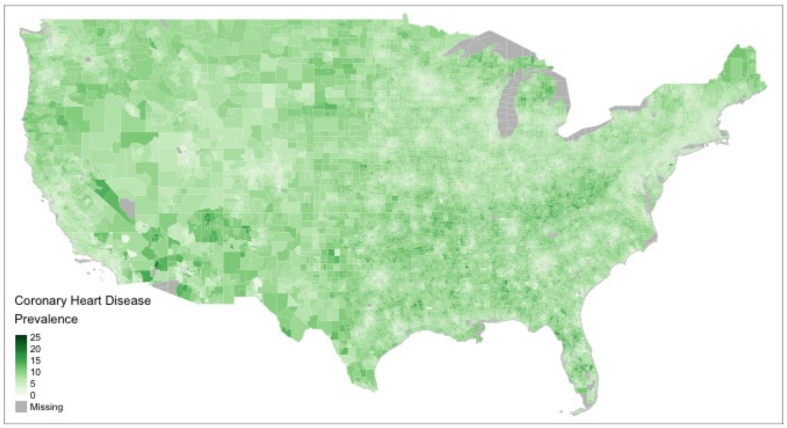
[23] Spatial Structures in the Social Sciences, Brown University. U.S. census-tract adjacency file. URL https://s4.ad.brown.edu/projects/diversity/index.htm.

[24] E. Tate, M. A. Rahman, C. T. Emrich, and C. C. Sampson. Flood exposure and social vulnerability in the united states. *Natural Hazards*, 106(1):435–457, 2021. URL https://doi.org/10.1007/s11069-020-04470-2.

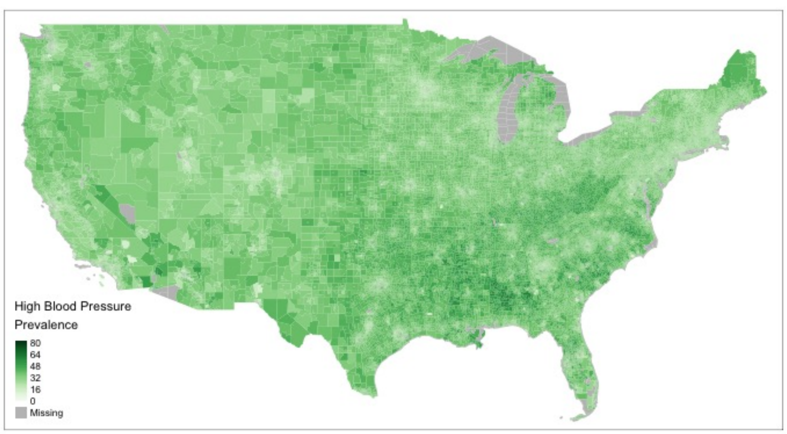
[25] United States Census Bureau, Geography Division. TIGER/Line shapefiles. URL https://www.census.gov/cgi-bin/geo/shapefiles/index.php.

[26] D. Westreich and S. Greenland. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *American journal of epidemiology*, 177(4):292–298, 2013.

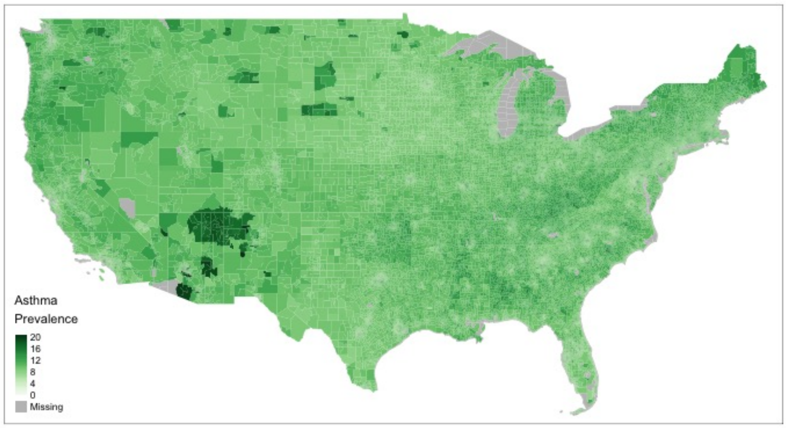
Supplementary Materials for Associations between Flood Risk and United States Census Tract-Level Health Outcomes



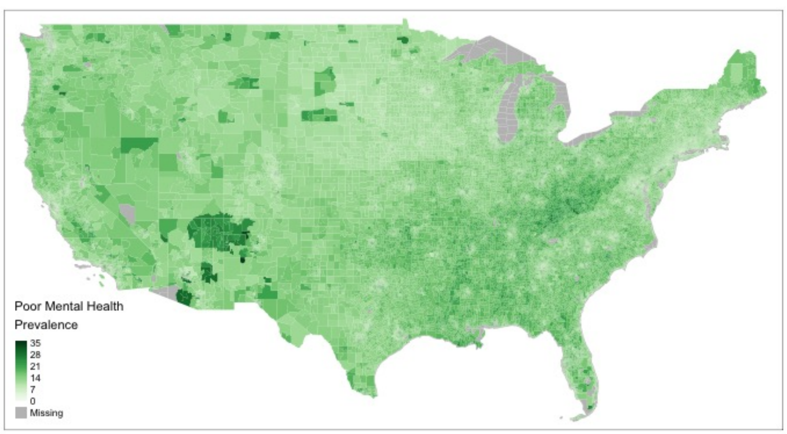
(a) Coronary Heart Disease



(b) High Blood Pressure



(c) Asthma



(d) Poor Mental Health

Figure S1: Choropleth maps for the prevalences (in %) of the four health outcomes among adults at the census tract level, calculated via small area estimation by the CDC [3].

Variable	Mean	S.d.	Min	Max
Social Vulnerability Index				
Theme 1: Socioeconomic Status				
Percentage of persons below poverty level	15.28	11.93	0	100
Percentage of persons unemployed	6.378	4.67	0	100
Per capita income	$32,258.07	$16,848.70	$42.00	$227,064.00
Percentage of persons without high school diploma	13.03	10.56	0	100
Theme 2: Household Composition and Disability				
Percentage of persons aged 65 and older	15.98	8.02	0	100
Percentage of persons aged 17 and younger	21.97	6.83	0	87.6
Percentage of persons older than 5 with a disability	13.37	5.88	0	100
Percentage of single-parent households	9.18	6.44	0	100
Theme 3: Minority Status and Language				
Percentage of minorities	37.96	30.03	0	100
Percentage that speak English less than well	4.13	6.81	0	100
Theme 4: Housing Type and Transportation				
Percentage of housing in structures with 10 or more units	12.25	18.45	0	100
Percentage of mobile homes	6.06	10.76	0	100
Percentage of occupied housing units with more people than rooms	3.52	5.18	0	100
Percentage of households with no vehicle available	9.39	12.24	0	100
Percentage of persons in group quarters	2.66	9.53	0	100
Percentage of persons uninsured	9.37	7.09	0	100
Air Pollution				
Outdoor concentration for CO (ppm)	0.36	0.09	0.21	1.93
Outdoor concentration for NO2 (ppb)	10.20	5.66	1.09	33.08
Outdoor concentration for O3 (ppb)	47.32	5.17	29.37	60.51
Outdoor concentration for PM10 ($\mu g/m^3$)	20.25	5.42	3.88	49.35
Outdoor concentration for PM2.5 ($\mu g/m^3$)	10.46	2.32	2.43	18.69
Outdoor concentration for SO2 (ppb)	2.19	0.98	0.58	9.01
Weather				
Mean daily maximum temperature (Kelvin)	293.3	4.80	280.4	305.5
Mean daily maximum relative humidity	83.97	8.98	36.70	98.56
Percentage of adults that smoke	18.28	5.87	3.20	51.70

Table 1: Summary statistics (mean, standard deviation, mean, and maximum) for moderators and secondary exposures.

Flood Risk Variable	PC1	PC2	PC3	PC4	PC5
Annual Probabilities of Flooding					
Percentage of properties in 2020 with greater than 0.2% annual probability of flooding	0.10	-0.08	-0.04	-0.01	-0.09
Percentage of properties in 2050 with greater than 0.2% annual probability of flooding	0.10	-0.09	-0.06	-0.11	-0.02
Percentage of properties in 2020 with greater than 1% annual probability of flooding	0.11	-0.01	0.00	0.16	-0.03
Percentage of properties in 2050 with greater than 1% annual probability of flooding	0.11	-0.03	-0.01	0.14	-0.11
Percentage of properties in 2020 with greater than 20% annual probability of flooding	0.08	0.08	0.24	-0.11	0.03
Percentage of properties in 2050 with greater than 20% annual probability of flooding	0.09	0.07	0.25	-0.03	0.12
FF Summary Statistics					
Average FF score	0.11	-0.02	0.05	0.02	-0.03
Standard deviation of FF score	0.06	0.09	-0.29	-0.22	-0.02
Coefficient of variation of FF score	-0.03	0.14	-0.29	-0.18	-0.06
Average FF score between 2 to 10	0.04	0.17	-0.08	0.01	-0.18
Average FF score of properties in 2020 with greater than 0.2% annual probability of flooding	0.04	0.17	-0.09	-0.07	-0.13
Average FF score of properties in 2020 with greater than 1% annual probability of flooding	0.03	0.17	-0.07	-0.14	-0.08
FF Percentages					
Percentage of properties with FF score 1	-0.10	0.09	0.06	0.11	0.02
Percentage of properties with FF score 2	0.02	-0.07	-0.05	-0.46	0.40
Percentage of properties with FF score 3	0.03	-0.10	-0.08	-0.42	0.15
Percentage of properties with FF score 4	0.03	-0.12	-0.02	-0.12	-0.20
Percentage of properties with FF score 5	0.05	-0.07	-0.19	0.02	-0.19
Percentage of properties with FF score 6	0.07	-0.07	-0.20	0.23	-0.25
Percentage of properties with FF score 7	0.05	0.03	-0.24	0.25	0.36
Percentage of properties with FF score 8	0.05	0.04	-0.12	0.25	0.58
Percentage of properties with FF score 9	0.07	0.06	0.13	0.10	0.22
Percentage of properties with FF score 10	0.07	0.07	0.31	-0.14	-0.10

Table 2: Loadings of the first five principal components from the PCA of the flood risk variables, which account for 80% of the variance. The flood risk covariates have been scaled both before and after the PCA, so the loadings in Table 2 are directly comparable. Important positive and negative loadings are highlighted in blue and red, respectively, to aid discussion.

Pollutant	PC1	PC2	PC3
CO	0.27	0.35	-0.14
NO2	0.30	0.16	-0.22
O3	0.12	-0.48	0.72
PM10	0.25	0.27	0.38
PM2.5	0.25	-0.28	0.00
SO2	0.14	-0.60	-0.54

Table 3: Loadings of the first three principal components from the PCA of the air pollutant concentrations, which account for 80% of the variance.

1 Interpretation of the Flood Risk Principal Components

This section serves as additional exposition on the interpretation of the flood risk principal components as presented in Section 3 of the main article.

According to the loadings in Table 2, PC 1 is positively correlated with all flood risk variables except "Percentage of properties with FF score 1," which indicates protectiveness against flooding, and "Coefficient of variation of FF score," which is not directly related to flood risk. Since the loadings are positive with no loading being especially larger than the rest, PC 1 has been labelled as 'consistent flood risk.' For PC 2, the largest loadings are assigned to the FF summary statistics, especially those that begin with "Average FF score," which all have loadings of 0.17. Therefore, PC 2 has been labelled as 'average FF score.' For PC 3, the largest positive loadings are assigned to severe flood risk indicators, as mentioned in Section 3 of the main article. In addition, the negative loadings -0.19, -0.20, and -0.24, respectively, are assigned to FF scores 5-7; however, since "Percentage of properties with FF score 10" is assigned the biggest loading of 0.31, PC 3 has been labelled as 'severe flood risk.'

PCs 4 and 5 are less straightforward to interpret than PCs 1-3. For PC 4, FF scores 6-8 have the loadings 0.23, 0.25, and 0.25, respectively, whereas FF scores 2 and 3 have loadings -0.46 and -0.42. Because PC 4 emphasizes both ends of the spectrum in opposite directions, it has been labelled as 'high/low risk difference.' For PC 5, FF scores 2-3 have positive loadings as large as 0.40, FF scores 4-6 have negative loadings as low as -0.25, and FF scores 7-9 have positive loadings as large as 0.58. Therefore, PC 5 has been labelled as 'skewed flood risk.'

2 Sensitivity Analyses

The sensitivity analyses show the impact of several modifications to the models stratified on all SVI Themes. Supplementary Materials Figure S9 shows the flood risk coefficient estimates for models that omit the SVIs instead of stratifying on them; the estimates can be considered as the average of the two strata. The results show the importance of stratification: a flood risk PC that was deemed not statistically significant on average may be significant for one of the strata. Several effects that were not statistically significant in Figure S9 were significant for the low SV strata and/or the high SV strata in Figure 2 of the main article.

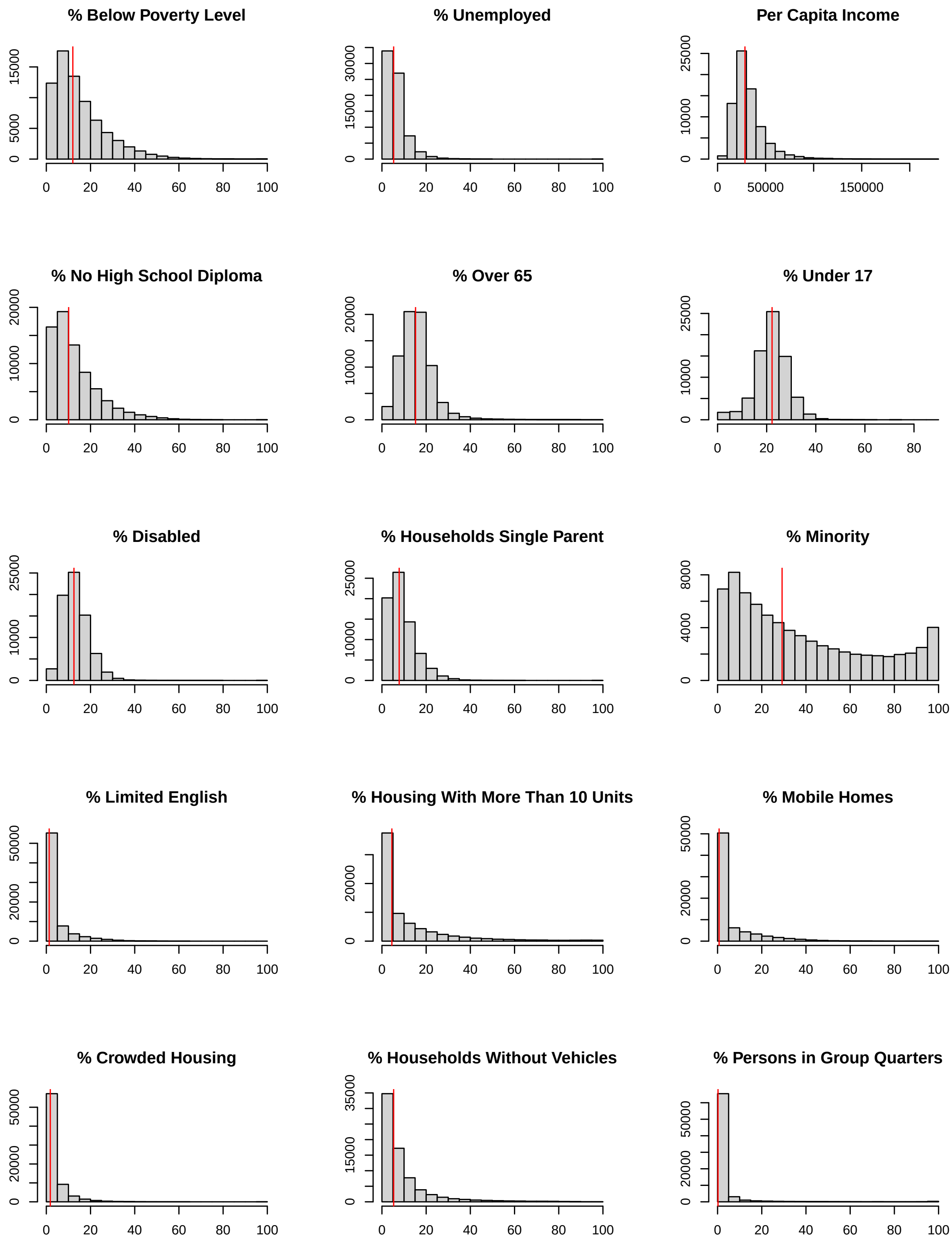
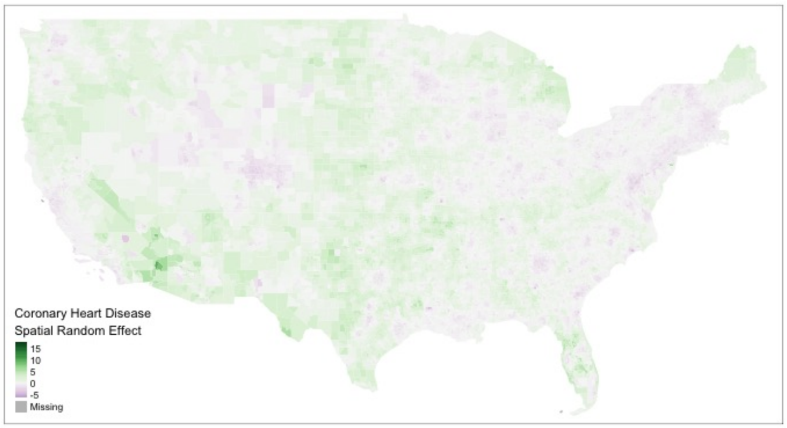
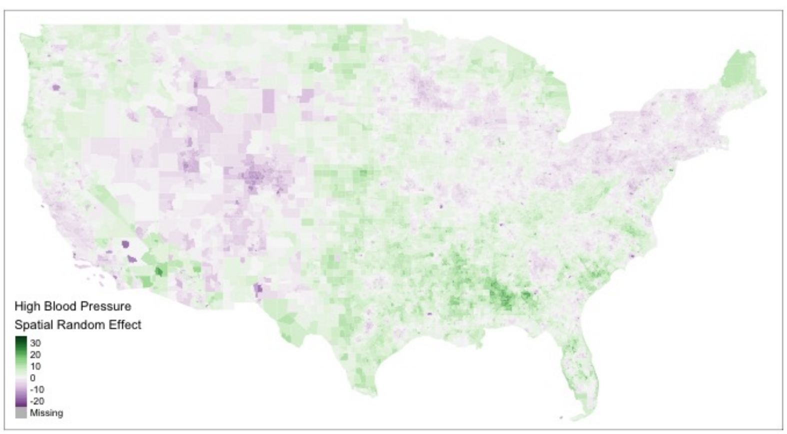


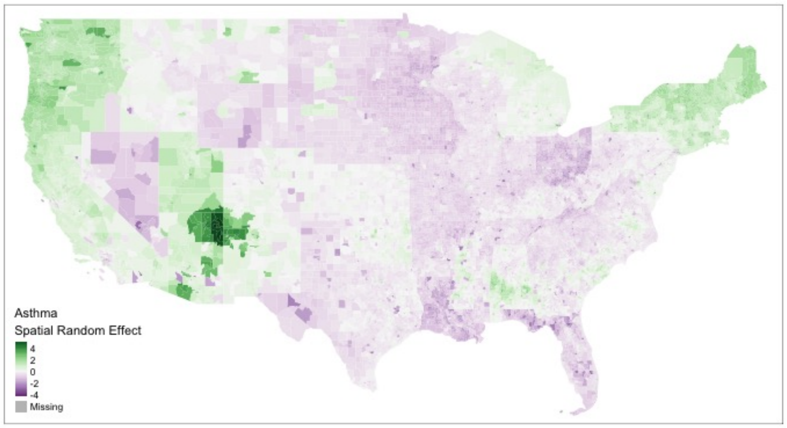
Figure S2: Distributions of the social vulnerability indices involved in the stratification of the models. The red vertical lines are the medians of the variables.



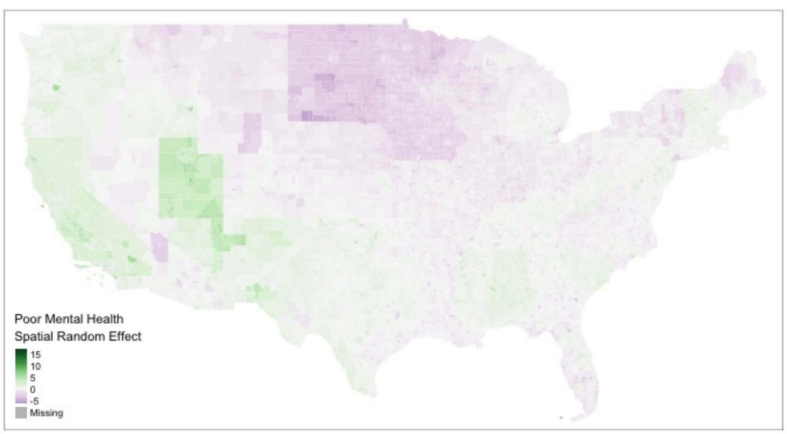
(a) Coronary Heart Disease



(b) High Blood Pressure

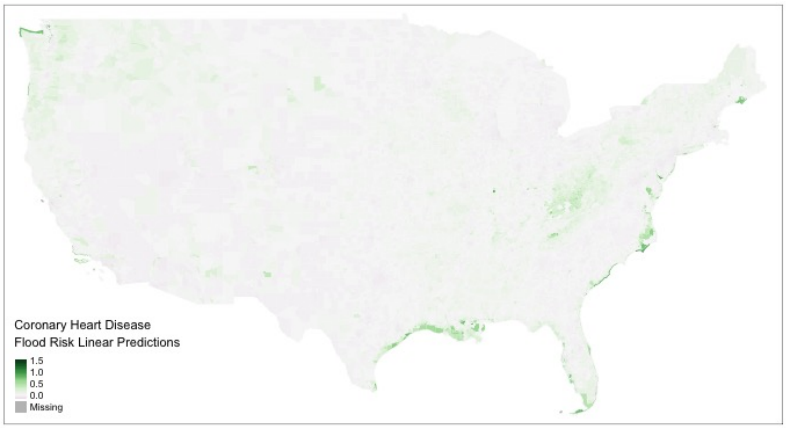


(c) Asthma

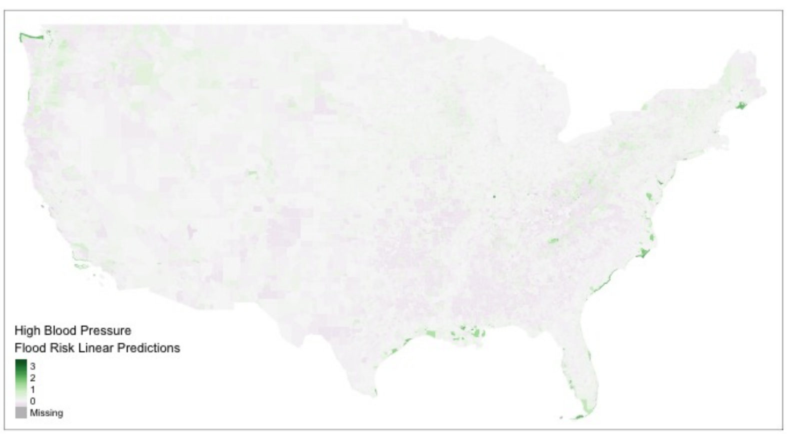


(d) Poor Mental Health

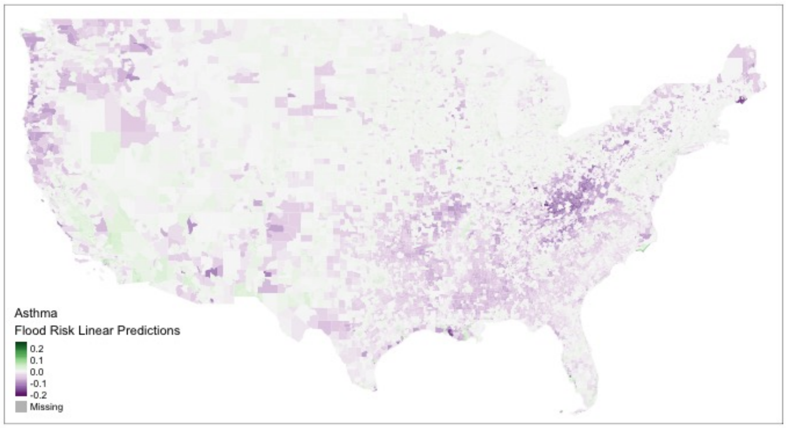
Figure S3: Choropleth maps for the spatially structured random effects for the four health outcomes, derived from BHMs stratified on all SVI Themes.



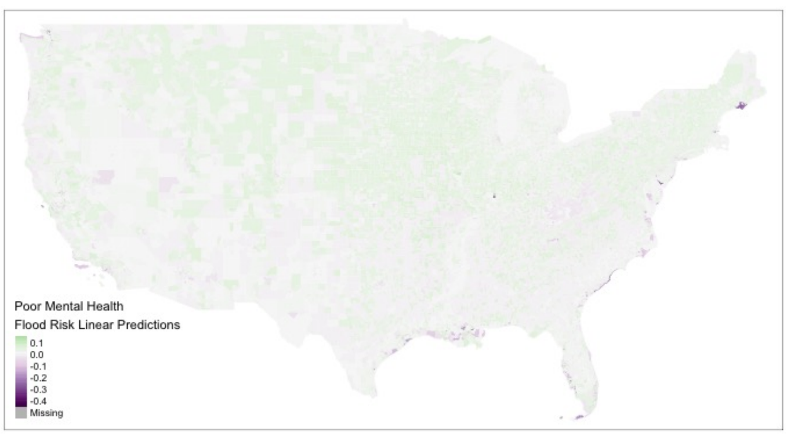
(a) Coronary Heart Disease



(b) High Blood Pressure

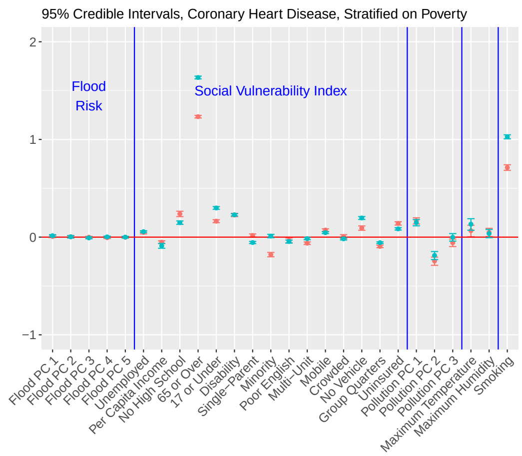


(c) Asthma

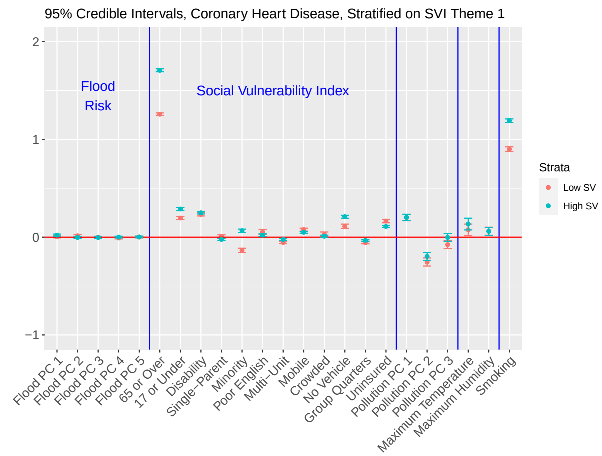


(d) Poor Mental Health

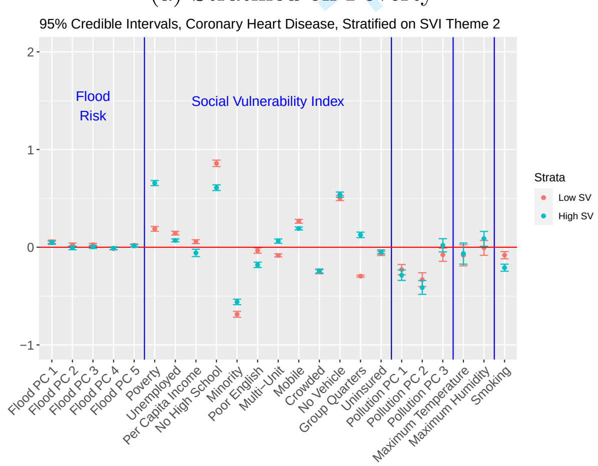
Figure S4: Choropleth maps for the flood risk linear predictions for the four health outcomes, derived from BHMs stratified on all SVI Themes.



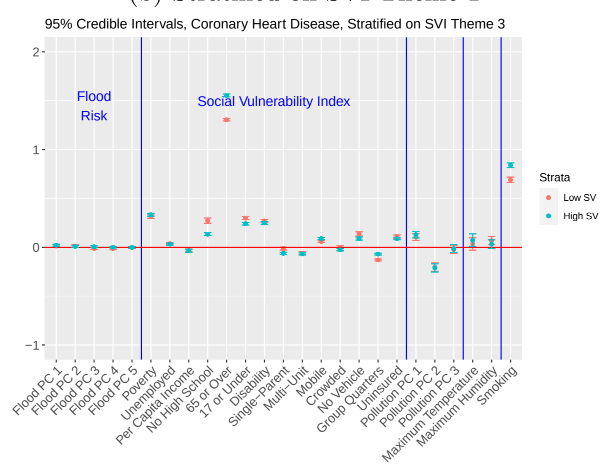
(a) Stratified on Poverty



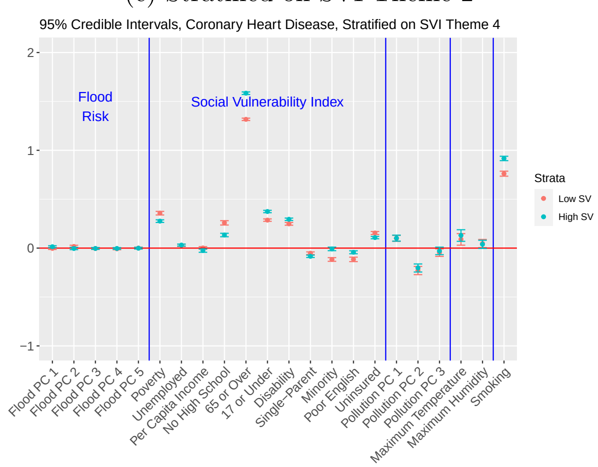
(b) Stratified on SVI Theme 1



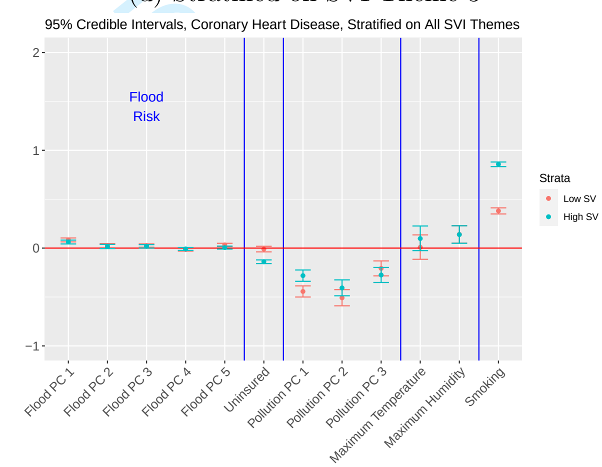
(c) Stratified on SVI Theme 2



(d) Stratified on SVI Theme 3

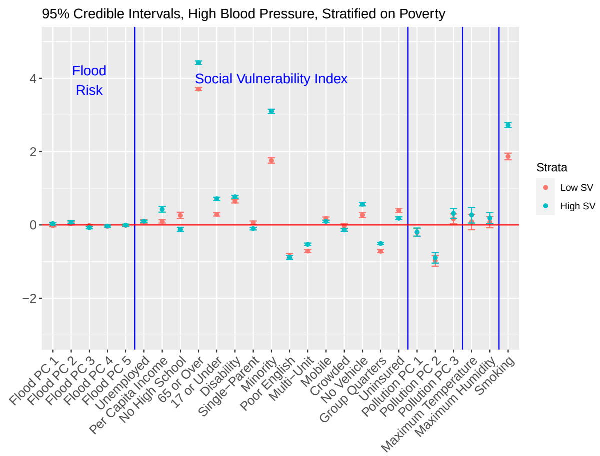


(e) Stratified on SVI Theme 4

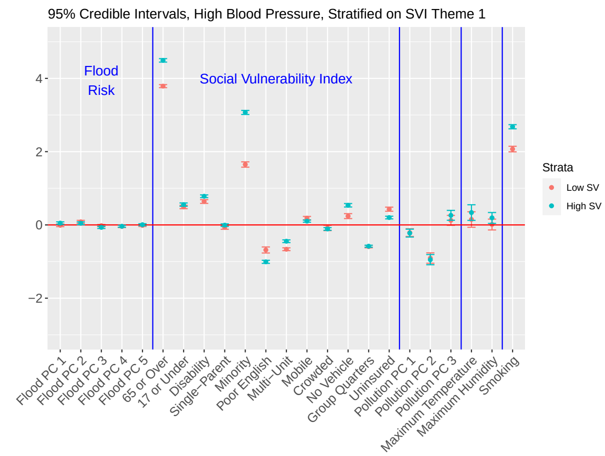


(f) Stratified on all SVI Themes

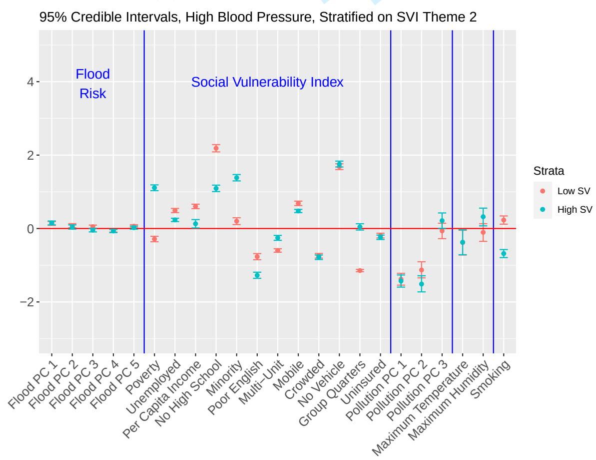
Figure S5: Credible intervals (95%) for the Coronary Heart Disease model stratified on poverty, theme 1: socioeconomic status, theme 2: household composition, theme 3: minority status, theme 4: housing type, and all SVI themes. These figures correspond to Figure 2a in the main article.



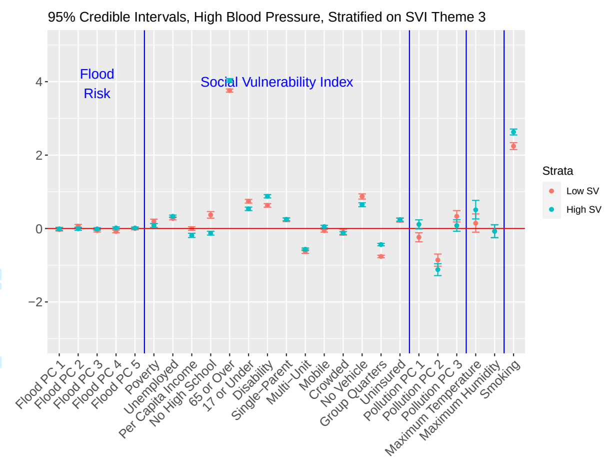
(a) Stratified on Poverty



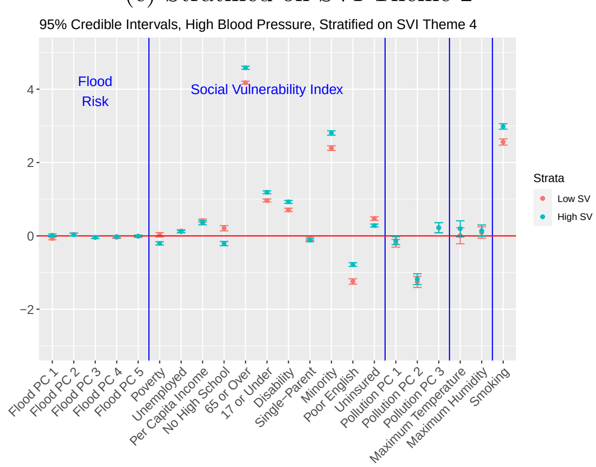
(b) Stratified on SVI Theme 1



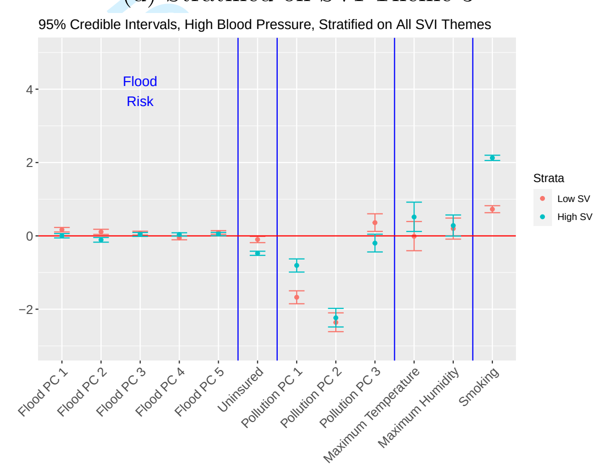
(c) Stratified on SVI Theme 2



(d) Stratified on SVI Theme 3

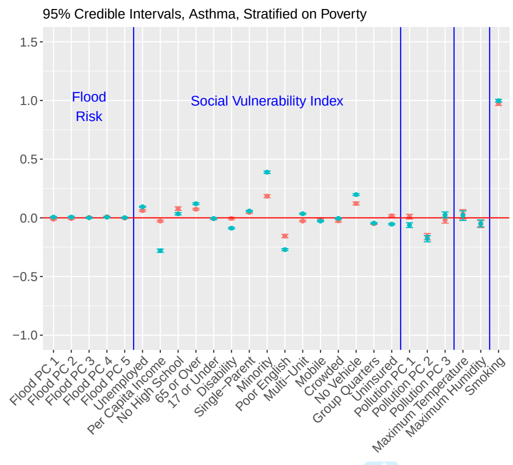


(e) Stratified on SVI Theme 4

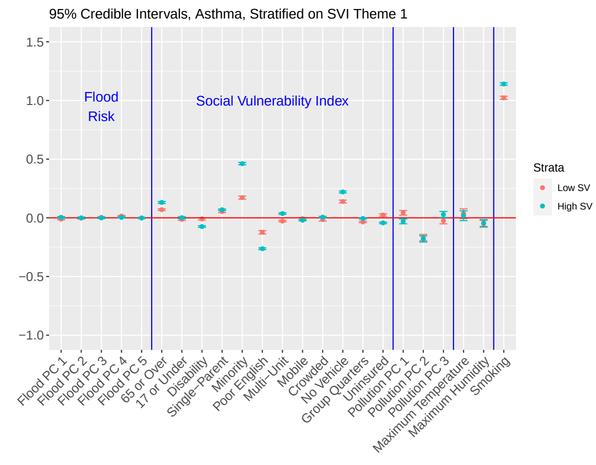


(f) Stratified on all SVI Themes

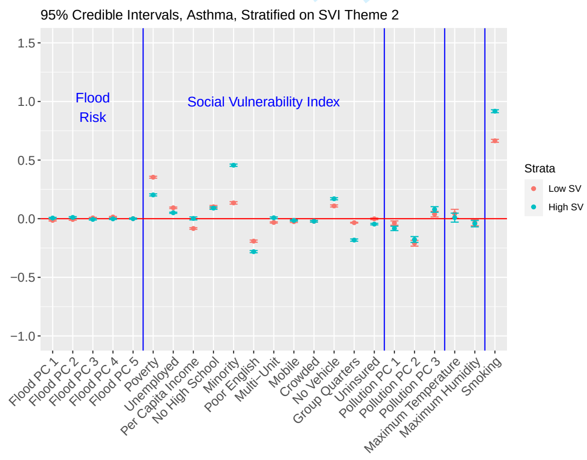
Figure S6: Credible intervals (95%) for the High Blood Pressure model stratified on poverty, theme 1: socioeconomic status, theme 2: household composition, theme 3: minority status, theme 4: housing type, and all SVI themes. These figures correspond to Figure 2b in the main article.



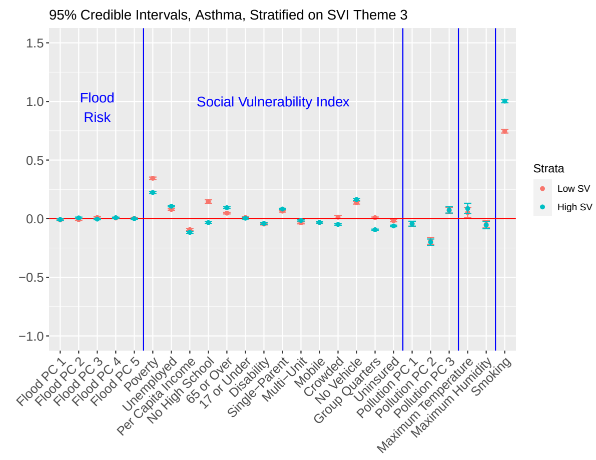
(a) Stratified on Poverty



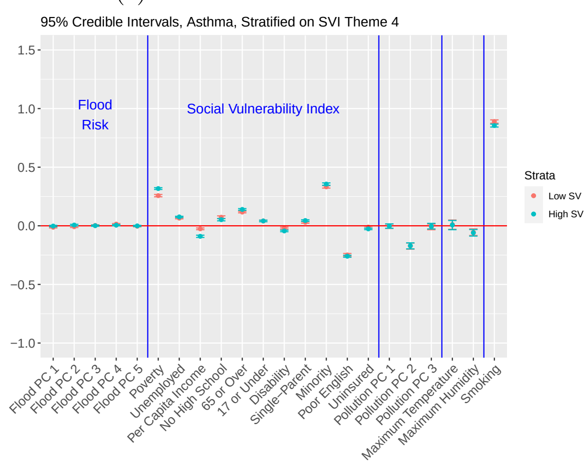
(b) Stratified on SVI Theme 1



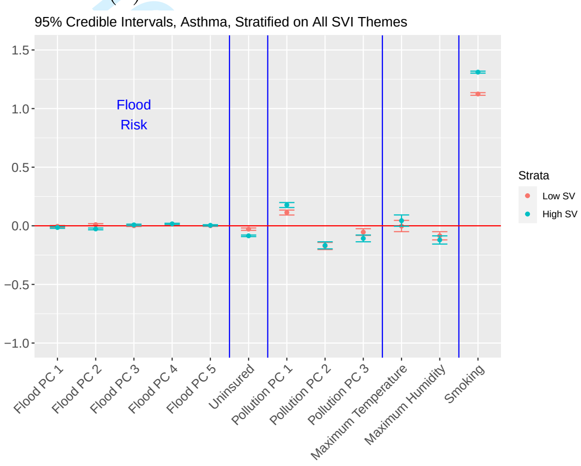
(c) Stratified on SVI Theme 2



(d) Stratified on SVI Theme 3

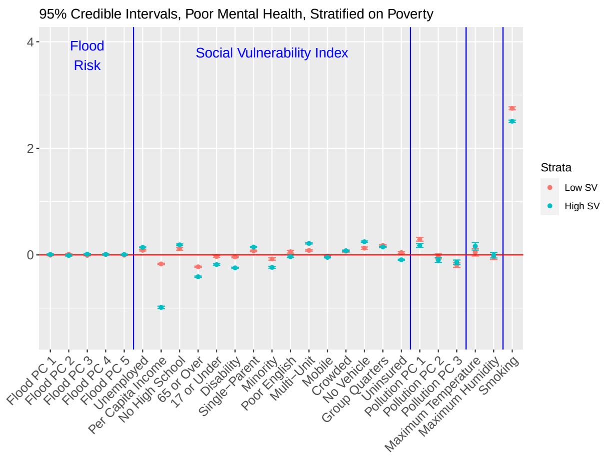


(e) Stratified on SVI Theme 4

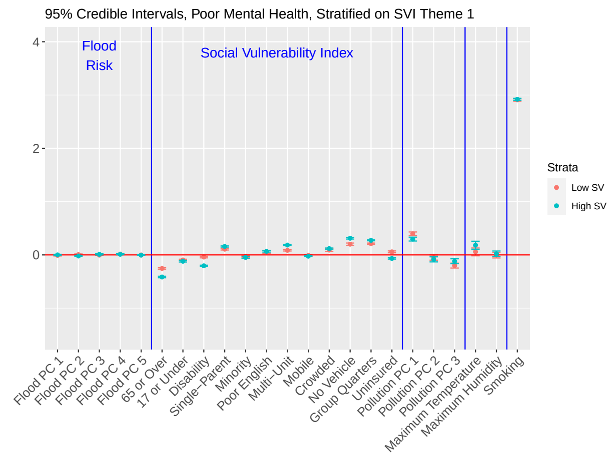


(f) Stratified on all SVI Themes

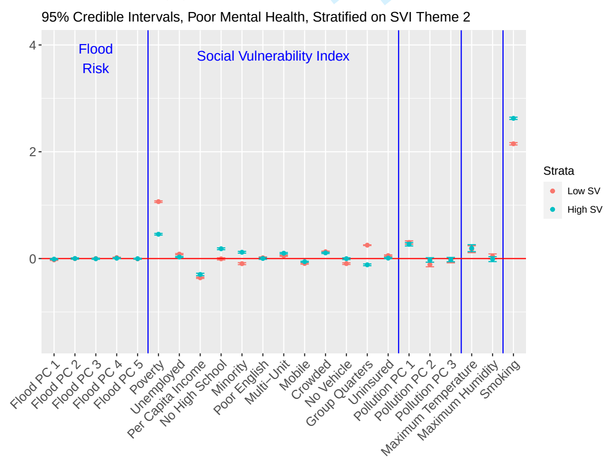
Figure S7: Credible intervals (95%) for the Asthma model stratified on poverty, theme 1: socioeconomic status, theme 2: household composition, theme 3: minority status, theme 4: housing type, and all SVI themes. These figures correspond to Figure 2c in the main article.



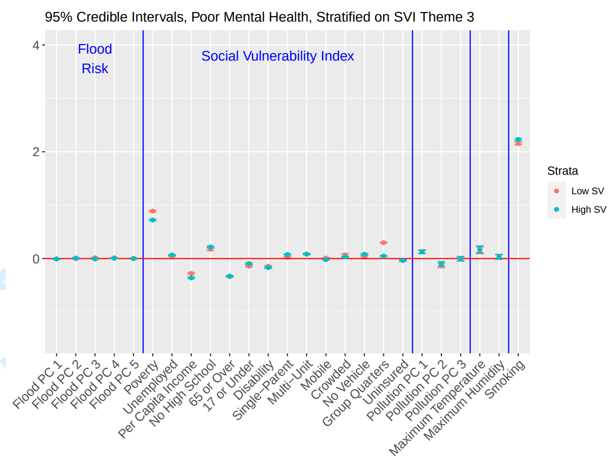
(a) Stratified on Poverty



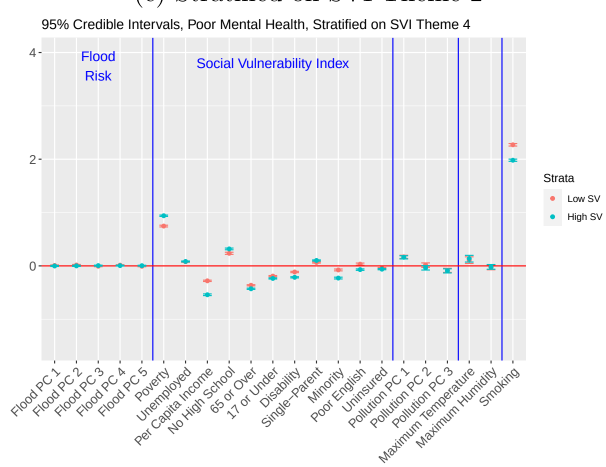
(b) Stratified on SVI Theme 1



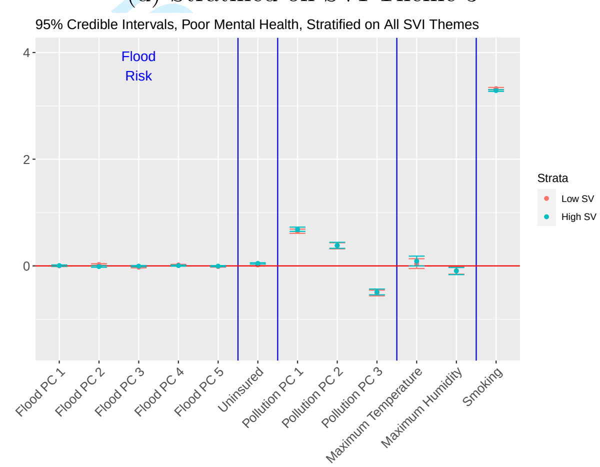
(c) Stratified on SVI Theme 2



(d) Stratified on SVI Theme 3



(e) Stratified on SVI Theme 4

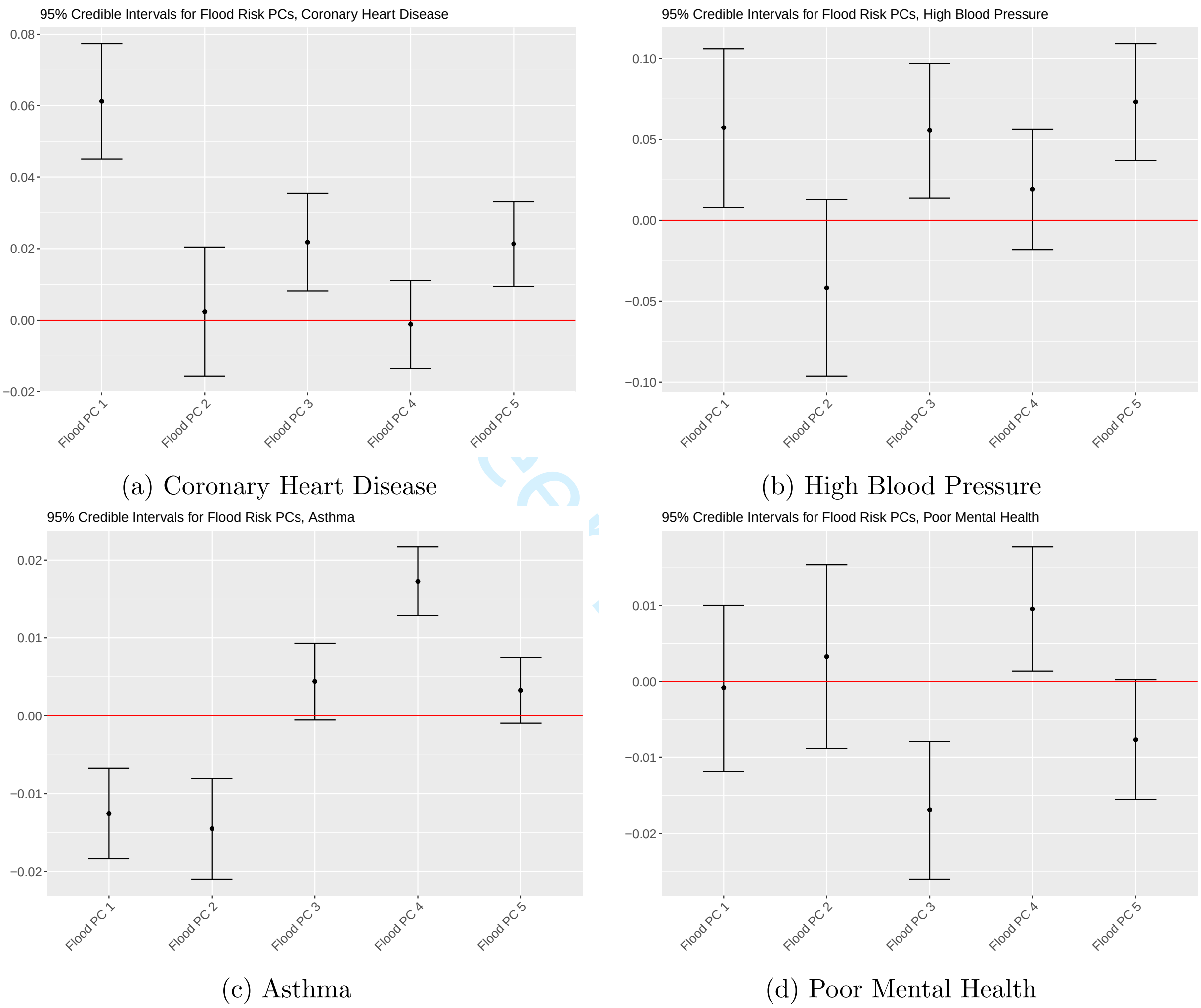


(f) Stratified on all SVI Themes

Figure S8: Credible intervals (95%) for the Poor Mental Health model stratified on poverty, theme 1: socioeconomic status, theme 2: household composition, theme 3: minority status, theme 4: housing type, and all SVI themes. These figures correspond to Figure 2d in the main article.

Supplementary Materials Figure S10 shows the flood risk coefficient estimates for models stratified on all SVI Themes, but without spatially structured random effects. That is, the models used only the independent error term with variance ν^2, as shown in Equation 1 in the main article, to account for the residuals. Simplifying the models in this way led to exaggerated coefficient estimates. However, Moran's I test showed significant spatial autocorrelation among census tracts (P-value < 0.0001 for each of the four health outcomes), so the larger effects were likely spurious.

Supplementary Materials Figure S11 shows the results for models stratified on all SVI Themes where flood factor groupings were used instead of flood risk PCs. FF Group 1 is the percentage of properties with FF score 2, 3, or 4; FF Group 2 is the percentage of properties with FF score 5, 6, or 7; FF Group 3 is the percentage of properties with FF score 8, 9, or 10. For coronary heart disease, there was a clear trend of higher prevalence with higher flood factors. However, the trends were less clear with the other three health outcomes. These results show that flood risk and prevalences of health conditions do not necessarily have a straightforward relationship, especially when social vulnerability is taken into account.



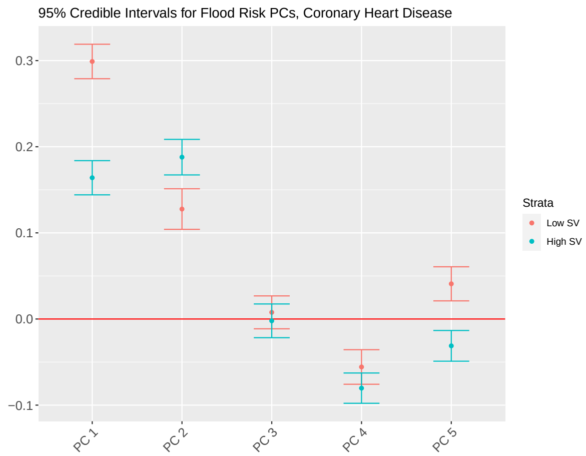
(a) Coronary Heart Disease

(b) High Blood Pressure

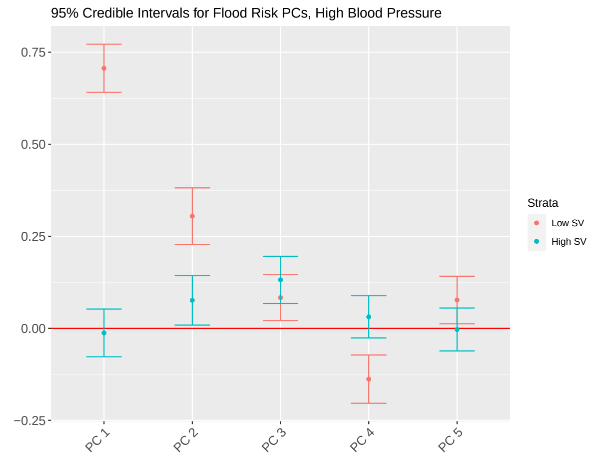
(c) Asthma

(d) Poor Mental Health

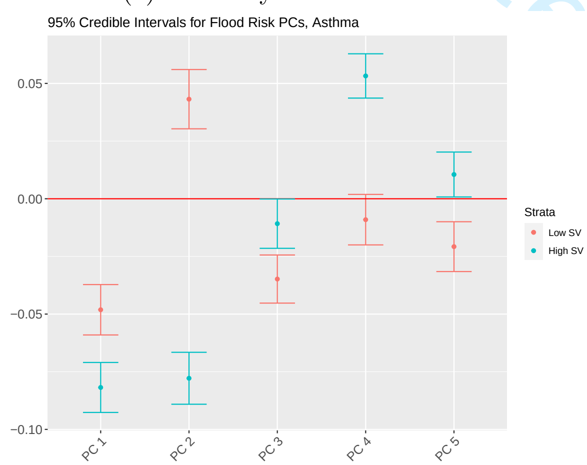
Figure S9: Credible intervals (95%) of flood risk PCs from unstratified models with all SVI variables omitted for each health outcome. Each of the four models contains all secondary exposures as listed in Table 1, but only the coefficients for the flood risk variables are shown.



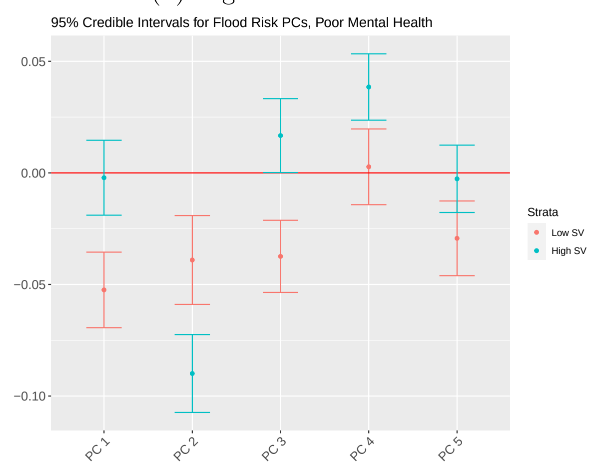
(a) Coronary Heart Disease



(b) High Blood Pressure

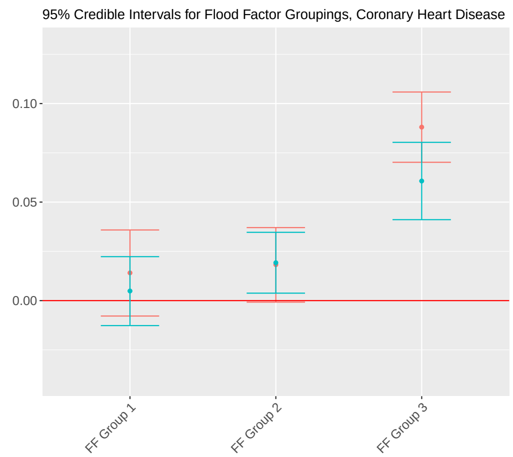


(c) Asthma

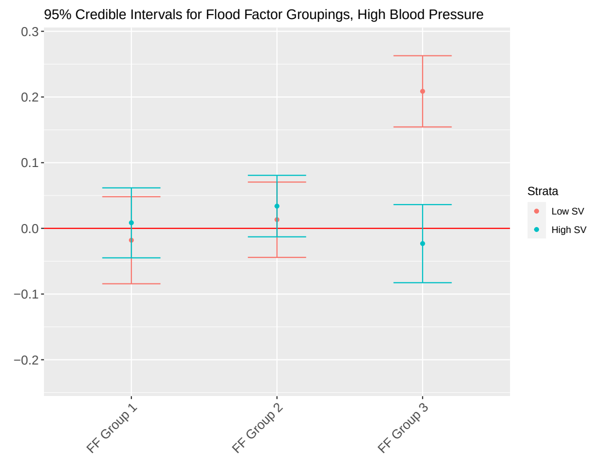


(d) Poor Mental Health

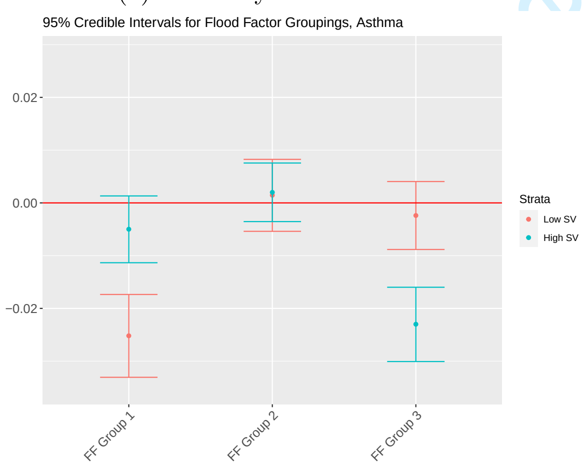
Figure S10: Credible intervals (95%) of flood risk PCs from linear regression models without spatially structured random effects for each health outcome, stratified on all SVI Themes. Each of the four models contains all secondary exposures listed in Table 1, but only the coefficients for the flood risk variables are shown.



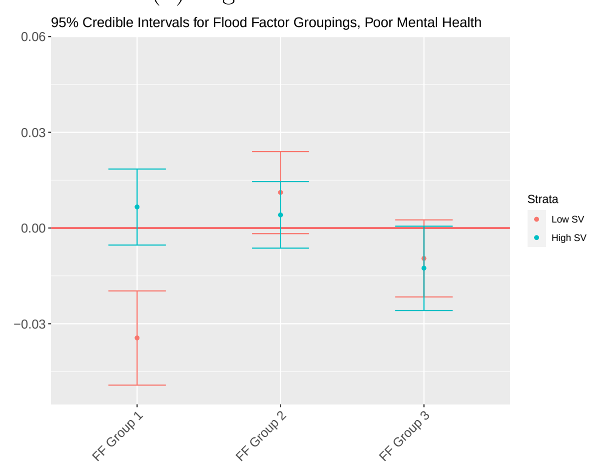
(a) Coronary Heart Disease



(b) High Blood Pressure



(c) Asthma



(d) Poor Mental Health

Figure S11: Credible intervals (95%) of flood factor groupings for each health outcome, stratified on all SVI Themes. Each of the four models contains all secondary exposures as listed in Table 1, but only the coefficients for the flood risk variables are shown.

3 Software

Below is a list of R packages used in this analysis.

1. `tidyverse`: collection of packages used for data cleaning and plotting [8].

2. `sf`: support for simple features, i.e., a standardized way to encode spatial vector data [6].

3. `spdep`: weighting schemes and statistics for spatial dependence. In particular, I use the function `moran.test` from this package which conducts Moran's I test for spatial autocorrelation [2].

4. `usdm`: this package includes `vifstep`, which calculates the variance inflation factor for a set of variables and excludes the highly correlated variables through a stepwise procedure [5].

5. `Matrix`: implements a hierarchy of dense and sparse matrices that may have a specific structure. The adjacency matrix of all census tracts in the US is a large, sparse binary matrix that corresponds to the 'ngCMatrix' class in the package `Matrix` [1].

6. `CARBayes`: implements the Bayesian CAR model [4]. We have adjusted the `CARBayes` code to accept a sparse matrix of class 'ngCMatrix.'

7. `coda`: output analysis and diagnostics for Markov Chain Monte Carlo simulations [7].

References

[1] D. Bates and M. Maechler. *Matrix: Sparse and Dense Matrix Classes and Methods*, 2021. URL https://CRAN.R-project.org/package=Matrix. R package version 1.4-0.

[2] R. Bivand and D. W. S. Wong. Comparing implementations of global and local indicators of spatial association. *TEST*, 27(3):716–748, 2018. URL https://doi.org/10.1007/s11749-018-0599-x.

[3] Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, and Division of Population Health. PLACES: Local data for better health, census tract data 2020 release, January 2021. URL https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/4ai3-zynv.

[4] D. Lee. *CARBayes version 5.2.3: An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors*. University of Glasgow.

[5] B. Naimi, N. a.s. Hamm, T. A. Groen, A. K. Skidmore, and A. G. Toxopeus. Where is positional uncertainty a problem for species distribution modelling. *Ecography*, 37: 191–203, 2014. URL https://doi.org/10.1111/j.1600-0587.2013.00205.x.

[6] E. Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446, 2018. URL https://doi.org/10.32614/RJ-2018-009.

[7] M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006. URL https://journal.r-project.org/archive/.

[8] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. URL https://doi.org/10.21105/joss.01686.