I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15996

Determinants of Heat Risk in an Aging Population: A Machine Learning Approach

Hannah Klauber
Nicolas Koch

MARCH 2023

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15996

Determinants of Heat Risk in an Aging Population: A Machine Learning Approach

Hannah Klauber
MCC

Nicolas Koch
MCC and IZA

MARCH 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Determinants of Heat Risk in an Aging Population: A Machine Learning Approach*

This paper identifies individual and regional risk factors for hospitalizations caused by heat within the German population over 65 years of age. Using administrative insurance claims data and a machine-learning-based regression model, we causally estimate heterogeneous heat effects and explore the geographic, morbidity, and socioeconomic correlates of heat vulnerability. Our results indicate that health effects distribute highly unevenly across the population. The most vulnerable are more likely to suffer from chronic diseases such as dementia and Alzheimer's disease and live in rural areas with more old-age poverty and less nursing care. We project that unabated climate change might bring heat to areas with particularly vulnerable populations, which could lead to a five-fold increase in heat-related hospitalization by 2100.

JEL Classification: I14, I18, Q51, Q54, Q58

Keywords: heat, climate change, hospitalization, risk factors, adaptation, machine learning

Corresponding author:
Nicolas Koch
Mercator Research Center on Global Commons and Climate Change (MCC)
Torgauer Straße 12-15
10829 Berlin
Germany
E-mail: koch@mcc-berlin.net

* We thank Christian Günster, and Thomas Ruhnke at AOK Research Institute (WIdO) for making our analysis possible and providing advice on the insurance data. Additionally, we thank Kelton Minor, Jisung Park, Nico Pestel, Nicholas Sanders, as well as the participants at the conferences of the European Association of Environmental and Resource Economists (EAERE) and the Association of Environmental and Resource Economists (AERE) for valuable feedback and suggestions.

1 Introduction

In the exceptionally hot summer of 2018, Europe recorded 104,000 heat-related deaths, more than any other WHO region (Watts et al. 2020). Health risks from heat stress distribute unevenly across the population and evidence suggests that the elderly and those with disabilities and pre-existing conditions are most affected (Campbell et al. 2018). In Germany, a country with a growing elderly population, the absolute number of heat-related mortalities during the summer of 2018 was globally only topped by China and India (Watts et al. 2020). Almost a quarter of the German population is older than 65 years. With population aging, the public healthcare and long-term care system will need to accommodate an increasing number of elderly over the next decades. In the light of finite budgets and finite contributory capacity of the workforce, this raises the understudied question of how to efficiently design localized and targeted protection of those most in need of it (Deryugina et al. 2021 are a notable exception).[1] Adaptation and resilience measures often focus on areas with high local heat exposure. However, if current heat exposure is imperfectly correlated with the heat vulnerability of a population, this approach may not be the most cost-effective targeting of local protection. An efficient adaptation design must consider that the heat-health nexus is a product of both local *heat exposure* and *heat vulnerability* of the local population.

This paper aims to disentangle the two dimensions of exposure and vulnerability for the elderly population in Germany on a much smaller scale than previous studies have been able to map. We combine hospitalization data from the largest public health insurance provider (AOK), covering a third of the population over 65 years of age, with high-resolution weather measure-

[1] Deryugina et al. 2021 investigate how to effectively target US air quality regulation to regions with the most to gain.

ments from 2008 to 2018. Using Machine Learning (ML) based inference, we identify individual risk factors (e.g. pre-existing health conditions), regional risk factors (e.g. characteristics of local healthcare infrastructure), and socio-economic risk factors (e.g. old-age poverty) for heat vulnerability among a host of more than 300 potential determinants. We generate maps that show how heat exposure, heat vulnerability, and heat damage distribute across Germany and project the number and geographic distribution of heat-related hospitalizations under different climate scenarios to the end of the century.

The identification of causal effects rests on random variation in the temporal and spatial occurrence of heat days. To analyze heterogeneity in the heat response ex-post, we use a machine-learning-based method proposed by Chernozhukov et al. (2018) that avoids overfitting and does not require pre-existing guiding principles. The underlying idea is simple. We predict the hospitalization probability for all elderly individuals twice. While the first prediction gives the likelihood of hospitalization based on empirical observations for similar individuals on heat days, the second prediction gives the respective likelihood on non-heat days. The average difference in the two predictions represents the heat-related probability of hospitalization and serves as a proxy for the vulnerability of each elderly person. Including this proxy in our regression analysis, we determine three objects: (i) the Best Linear Predictor (BLP) which returns the average heat effect and indicates the existence of heterogeneity, (ii) Sorted Group Average Treatment Effects (GATES) which represent the effects of heat on groups of varying predicted vulnerability, and (iii) Classification Analysis estimates (CLAN) which reveal the average characteristics of the elderly predicted to be most vulnerable and their place of residence.

Our results show that about a quarter of the elderly population experiences a heat-related hospitalization risk above the peer group's average. Individual health status and characteristics of the residential environment are clearly related to their increased vulnerability. On average, the most vulnerable are older, more often men and they suffer more frequently from diseases such as dementia and Alzheimer's disease, renal insufficiency, diabetes, and depression. Areas where a higher proportion of vulnerable individuals live tend to be more rural, suffer more from old-age poverty, exhibit less capacity and provision of professional nursing care, but feature a higher density of medical practices. We identify a negative correlation between the vulnerability of the elderly population and exposure to heat. Thus, areas with relatively moderate historic exposure might suffer disproportionally from increasing temperatures due to climate change. We project that climate policies that comply with the Paris Agreement could maintain the *status quo* of heat-related health damage. However, under unabated climate change heat-related hospitalization could increase five-fold by 2100.

The paper makes two important contributions. First, it bridges the gap between two complementary strands of literature. Small scale-studies with detailed information on the population and residential environment often cover only specific places such that statistical power and external validity are limited (e.g. Eisenman et al. 2016, Hondula et al. 2012, Schwartz 2005). At the same time, large-scale studies with great temporal and geographical coverage often describe heat effects on broader population subgroups but lack the necessary level of detail to identify more complex heterogeneous relationships (e.g. Karlsson and Ziebarth 2018, Gronlund et al. 2014, Bobb et al. 2014). Our study provides first and detailed insights into the distribution of heat effects within a specific population group for the whole of Germany over eleven years.

Second, we are not aware of other studies that incorporate extensive information on individual level morbidity when estimating the effects of heat on health. Prior studies rely on mortality or hospitalization rates as the only health information and use fixed-effects or Poisson models to examine the dose-response relationship between heat and health (Karlsson and Ziebarth 2018, Deschenes 2014). Our ML-based heterogeneity approaches instead focuses on a simplifying heat event definition (days with at least 30 °C) but fully accounts for individual health histories with respect to 76 different chronic diseases and other diseases that commonly require hospitalization and medication of the elderly. In addition to this rich set of individual-level morbidity variables, we incorporate 231 variables on the demographic, socioeconomic, and infrastructural characteristics of the individuals' zip code area of residence. We process this extensive dataset at the person-day level with a flexible ML technique that applied researchers have only started to use (see Breda et al. 2021, Deryugina et al. 2021, Deryugina et al. 2019, Beaman et al. 2014).

Third, we provide localized estimates of the heat-related health risks in different climate scenarios. In coupling future heat exposure with the heat vulnerability of the local population, we account for the fact that health damage is unlikely to scale linearly with increasing heat exposure. While this scenario analysis does not suffice to draw general conclusions on heat adaptation that may be required in the coming decades, it does provide *ceteris paribus* evidence on the regional vulnerability to climate change and highlights the role of mitigation policy.

The remainder of the paper is organized as follows. In Section 2 we describe the machine learning approach and introduce our empirical model. We provide an overview on the health, air pollution, weather, and residential data in Section 3. Our regression results and visual analysis are presented in Section 4. We conclude after a discussion of our results in Section 5.

2 Empirical Strategy

We are interested in estimating the causal effect of heat exposure on hospitalization in elderly individuals of varying vulnerability. To identify heterogeneity, we implement a recently developed method by Chernozhukov et al. (2018) that rests on a machine-learning approach. The technique requires a binary treatment. Therefore, we define heat events as days with at least 30 °C and implement the following procedure.

First, we randomly divide our person-day observations into two data sets, the train and the main data set, which cover about 50% of the insured individuals each.[2] In the train data set we assign all person-day observations with exposure to at least 30 °C to the treatment group and the remaining observations to the control group. Based on the train data, we then train two prediction models using a gradient tree boosting algorithm.[3] One model rests only on the observations of the treatment group, the other only on the observations of the control group. Both models forecast hospitalization H_{it} for individual i on day t based on the covariates Z_{it}. The covariates include information on the insured individuals (e.g. age, sex, and health status), weather conditions (e.g. precipitation, wind, and humidity), air quality conditions (e.g. NO_2 and $PM_{2.5}$), and characteristics of the individuals' zip code area of residence (e.g. socio-economic status, distance to the nearest hospital, and nursing capacity). Although these variables come

[2] All observations for an individual are exclusively in either the train or the main data set.

[3] More precisely, we use the open-source distributed gradient-boosting framework LightGBM developed by Microsoft.

with varying temporal and spatial resolution[4], we collectively index them at the person-day level, it, for ease of notation. Moreover, Z_{it} includes month, year, and county fixed-effects.[5] In addition, we train a third prediction model that rests on all person-day observations in the train data set and determines heat exposure T_{it} as a function of the covariates Z_{it}.

Second, we predict the hospitalization probabilities for each person-day in the main data set using the models trained for the treatment and the control group. This yields two predicted probabilities: one pretends that person i is exposed to heat on day t and the other pretends that she is not. The difference between both predictions, $\hat{S}(Z_{it})$, corresponds to the hospitalization probability attributable to heat. It serves as a proxy for the individual's daily heat vulnerability, and although potentially biased, it suffices for inference in heterogeneity (Chernozhukov et al. 2018). Additionally, we predict the probability of heat exposure for each person-day, i.e. the predicted propensity score $\hat{p}(Z_{it})$. Based on the main data we then estimate the following weighted regression to obtain the Best Linear Predictor (BLP) of the conditional average heat effect (CATE):

$$H_{it} = \alpha + \beta_1(T_{it} - \hat{p}(Z_{it})) + \beta_2(T_{it} - \hat{p}(Z_{it}))(\hat{S}(Z_{it}) - \bar{\hat{S}}) + \theta \hat{H}^C(Z_{it}) + \epsilon_{it} \tag{1}$$

where the weights are equal to

$$w(Z_{it}) = \frac{1}{\hat{p}(Z_{it})(1 - \hat{p}(Z_{it}))}.$$

[4]Individual information are registered quarterly, weather variables come at the zip area-day level, air quality variables at the zip area-year level, characteristics of the place of residence are time-invariant and available at the zip area level (see Section 3).

[5]The machine-learning algorithm allows for interactions of variables and fixed-effects and, accordingly, is less restrictive than a linear regression with the same variable selection (Deryugina et al. 2019).

The dependent variable H_{it} equals one if individual i is hospitalized on day t and zero otherwise. If the individual is exposed to heat on the same day the indicator T_{it} takes on a value of one. From this variable, we subtract the propensity score $\hat{p}(Z_{it})$. The vulnerability proxy $\hat{S}(Z_{it})$ is centered to its mean $\bar{\hat{S}}$. For greater precision, the model also includes the predicted hospitalization probability for days without heat $\hat{H^C}(Z_{it})$. The model returns two parameters of interest: $\hat{\beta}_1$ is the unbiased estimator of the average heat effect on hospitalization (ATE) and $\hat{\beta}_2 = 0$ is the heterogeneity loading parameter (HET). Rejecting the null hypothesis $\hat{\beta}_2 = 0$ means that heat does not affect individuals homogeneously and that the predicted vulnerability proxy captures the heterogeneity.

To identify how much the heat effect varies across individuals, we estimate Sorted Group Average Treatment Effects (GATES) based on the following equation

$$H_{it} = \alpha + \sum_{l=1}^{7} \gamma_l (T_{it} - \hat{p}(Z_{it})) \cdot 1(G_l) + \theta \hat{H^C}(Z_{it}) + \epsilon_{it} \qquad (2)$$

where indicator $1(G_l)$ is equal to one if the vulnerability proxy of individual i on day t belongs to group l. To define groups, we split the bottom 75 percent of the predicted vulnerability distribution into 25-percentile intervals and consider the 75th, 85th, 95th, and 99th percentiles as lower interval bounds for the top 25 percent. We weight the regression with the weights $w(Z_{it})$ as defined above. The coefficients γ_l represent the average heat effects for these groups of varying predicted vulnerability.

Finally, we describe typical characteristics of the elderly individuals most at risk, i.e. the 1% predicted to be most vulnerable, based on Classification Analysis (CLAN). To this end, we simply test how their mean characteristics and the characteristics of their places of residence compare to those of the group predicted to be less vulnerable than average.

Implementing the ML-based heterogeneity analysis by Chernozhukov et al. (2018) involves practical implementation challenges because of the size of the data and the relatively low probability of hospitalization on any given day. Therefore, we follow Deryugina et al. (2019) and use a downsampling method (see Einav et al. 2018) and run the regressions on equally sized partitions of the data to aggregate the coefficients and standard errors subsequently.[6] In all analyses, we include only observations with an estimated propensity score between 0.1 to 0.9 to reduce noise. Moreover, we estimate equations 1 and 2 with an alternative approach using the Horvitz-Thompson transformation to test the robustness of our results (see Chernozhukov et al. (2018)).

3 Data

Health data

We analyze the settlement data for all inpatient and outpatient hospital treatments of insured individuals over 65 years of age in the years 2008 through 2018.[7] The data are from AOK which is the largest public health insurance provider in Germany and covers about a third of the population. On average, our sample comprises about 5.8 million elderly individuals each year. The data include information on their age, sex, place of residence (5-

[6]We split the data randomly into ten partitions that contain 50,696,668 observations each. To aggregate the results, we calculate the mean of the estimated coefficients and divide the mean of the estimated standard errors by the square root of ten. Due to long computing times, we do not repeat our estimation 100 times as recommended by Chernozhukov et al. (2018) and argue that splitting variation is a minor concern given our large sample sizes (see Deryugina et al. 2019).

[7]The legal basis of these health records is determined in §301 (1) SGB V and §295 SGB V.

digit zip area code), as well as the admission date of each hospitalization. Based on these records, we create a dataset where the unit of observation is the person-day. For every person and every day, we determine whether a hospital treatment occurred or not. Next to the individuals' age, sex, and place of residence, we link additional information on the individuals' health status to this dataset. We use 76 dichotomous variables that capture the presence of pre-existing diseases (classified according to the ICD-10 system[8]) and the therapy with pharmaceuticals (classified according to the ATC-system[9]). The variables detect mostly chronic diseases (e.g. chronic cardiovascular and respiratory diseases, diabetes, dementia, and Alzheimer's disease) and diseases that commonly require hospitalization of the elderly in Germany (Destatis 2017). We obtain them from the AOK Research Institute (WIdO) which generates them for every insured person and every quarter based on the individuals' demand for health care services in the eight preceding quarters. Individuals who are not insured continuously in the eight previous quarters do not enter the analysis. Because we are interested in heat effects, we restrict our sample to the months May through September.

Weather and air pollution data

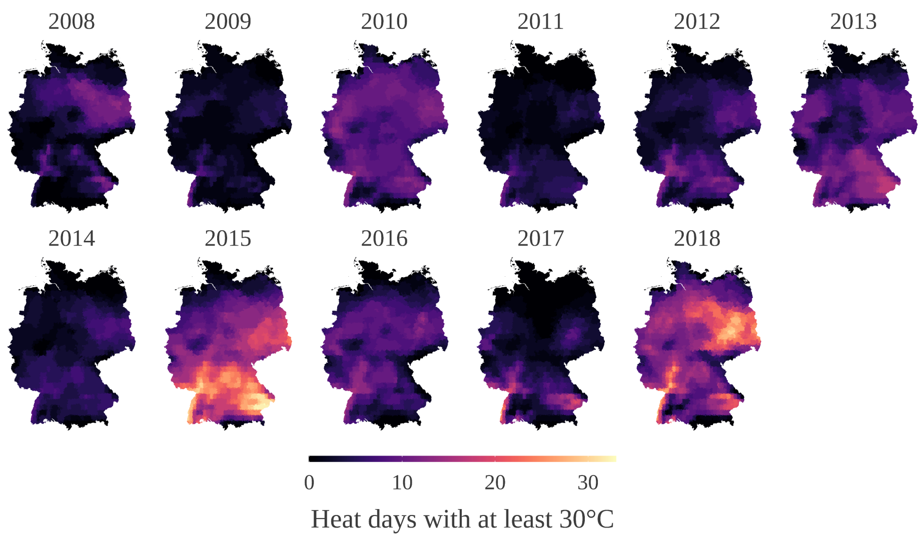
To construct our weather and air quality variables, we use the ERA5 product from the European Center for Medium-Range Weather Forecasts (ECMWF). ERA5 divides the Earth into a grid with a horizontal resolution of 31 km and provides hourly data for each grid cell. We use measurements

[8] The ICD-10-Code is an international system for the statistical classification of diseases and related health problems provided by the WHO. Germany uses the extended version ICD-10-GM (*Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme [website]*).

[9] The Anatomical Therapeutic Chemical (ATC) classification system categorizes drugs based on their active ingredients according to the organ or the system on which they act as well as their therapeutic, pharmacological, and chemical properties.

on temperature, precipitation, relative humidity, wind, cloud cover, surface pressure, and ozone concentrations and aggregate them to the day and zip area level. Based on the temperature data, we determine all areas and days subject to heat, i.e. they exhibit a maximum record of at least 30°C. Figure 1 illustrates how heat days distribute across Germany over the years in our sample. We feed the remaining weather variables as covariates into our ML-model, considering their daily mean, minimum, and maximum value. In addition, we use annual satellite-based particulate matter ($PM_{2.5}$) data from van Donkelaar et al. (2019) and nitrogen dioxide (NO_2) data that are interpolated from local measuring stations and provided by Germany's Federal Environment Agency (UBA) to account for air pollution exposure in each zip area. For future climate projections, we use bias-corrected daily maximum temperatures from the Geophysical Fluid Dynamics Laboratory's Princeton Earth System Model (GFDL-ESM4), which is one of the global climate models in the sixth phase of the Coupled Model Intercomparison Project (CMIP-6) (ISIMIP 2020, Lange 2019).

Figure 1: Heat days in the years 2008 to 2018



Socioeconomics, demographics, and infrastructure

We compile a host of 231 time-invariant variables that describe the places of residence of the elderly individuals in our sample. The variables capture demographic characteristics (e.g. age distribution, household size, and living space), socioeconomic characteristics (e.g. status class and old-age poverty), and infrastructural characteristics (e.g. traffic connections, distance to the nearest hospital, and nursing capacity). We construct variables at the zip area level that feed into our ML-model based on raster data from the 2011 Census from the Federal and State Statistical Offices, publicly available geodata from OpenStreetMap, and commercial data from the company Acxiom[10]. Moreover, we exploit Google Maps' programming interface to calculate driving distances from the zip areas' centroids to the nearest hospital. We use additional county data from the Federal Office for Building and Regional Planning (BBSR) in our descriptive classification analysis (CLAN). A detailed description of all variables is available in Appendix Table 2.

4 Results

4.1 Heterogeneous heat effects on hospitalization

First, we estimate the average treatment effect of heat days on hospitalization rates and determine the presence of heterogeneity based on Equation 1. Coefficient $\hat{\beta}_1$ in Table 1 shows that a heat day increases hospitalization on average by 39.79 admissions per million insured elderly individuals. The

[10]Acxiom collects, analyzes, and sells customer and business information for targeted advertising.

Table 1: The conditional average heat effect

	Hospitalization per million insured individuals
$\hat{\beta}_1$ (ATE)	39.79***
	(5.23)
$\hat{\beta}_2$ (HET)	48,854.26***
	(10,609.56)
Sample size	506,966,676

The table reports the BLP of CATE using the predicted vulnerability proxy $\hat{S}(Z_{it})$. The dependent variable is the daily hospitalization rate per million insured individuals. Parameter β_1 measures the average effect of a heat day on the hospitalization rate. Rejecting the null hypothesis $\beta_2 = 0$ implies that the heat effect is heterogeneous and that the vulnerability proxy captures components of heterogeneity. Standard errors are clustered at the zip area level and are in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$.

effect falls within the range of heat-related hospitalization rates estimated in similar studies.[11]

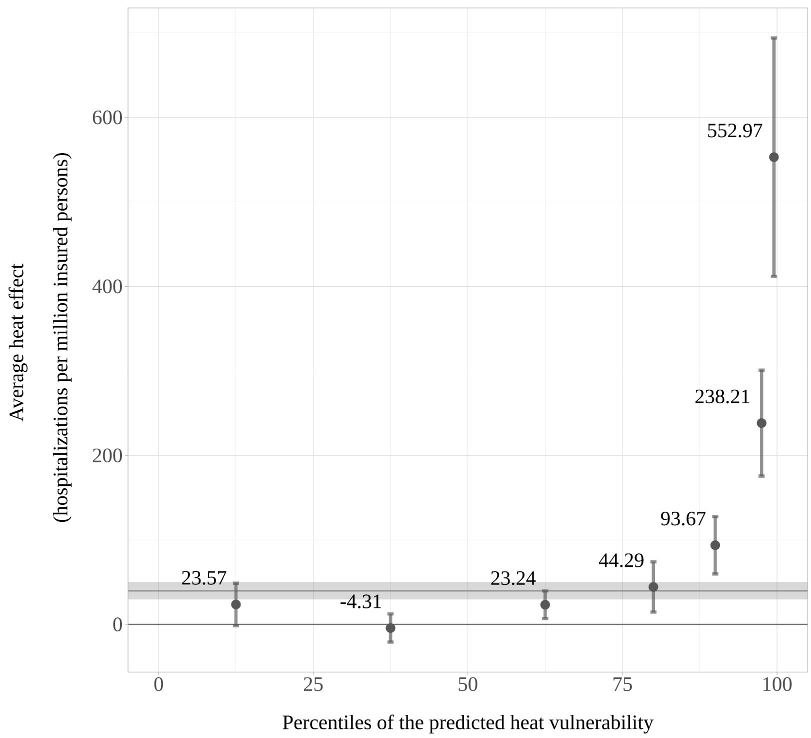
With a p-value well below 0.001 for $\hat{\beta}_2$, we can reject the null hypothesis that there are no differences in vulnerability to heat among the insured. The statistical significance of the heterogeneity loading parameter also indicates that the covariates included in our ML-model are relevant determinants of heat vulnerability. We robustify this finding by using the alternative Horvitz-Thompson estimation strategy proposed by Chernozhukov et al. (2018) and running regressions on data aggregated to the zip area level in Appendix Tables 3 and 4.

Next, we identify the extent to which heat effects differ across groups of varying vulnerability based on Equation 2. Figure 2 illustrates the average

[11] In their analysis of people over 65 years of age in the United States, Bobb et al. (2014) identify statistically significant effects for five out of 214 diagnoses, which add up to an absolute heat-related hospitalization rate of about 12 admissions per million insured individuals and heatwave day. Karlsson and Ziebarth (2018) obtain somewhat larger effects for Germany. For the age group 65-75, admissions per million people are close to 70, for the age group 75+, they are between 90 and 100. However, a direct comparison is not feasible due to different definitions of heat events and health outcomes.

heat effects and their 95% confidence intervals by percentiles of the predicted vulnerability distribution. In this figure, the predicted heat-related hospitalization probability increases from left to right of the x-axis. The purple horizontal line marks the average treatment effect of 39.79 from Table 1. The figure reveals that heat has little or no effect on the hospitalization probability for a large proportion of our elderly population. Roughly three quarters exhibit a below-average hospitalization risk. However, the risk increases significantly and rapidly for the top percentiles. For the most vulnerable 1%, the hospitalization rate on heat days increases to 552.96 per million individuals. This effect is nearly 14 times higher than the average effect.

Figure 2: The average heat effects by groups with varying vulnerability



The figure illustrates GATES estimated based on Equation 2, i.e. the mean heat effects for insured individuals with a predicted hospitalization probability in the $[0, 25)$, $[25, 50)$, $[50, 75)$, $[75, 85)$, $[85, 95)$, $[95, 99)$, and $[99, 100]$ percentile-interval. We report coefficients along with 95% confidence intervals. The horizontal purple line represents the mean heat effect (39.79) from Table 1. Standard errors are clustered at the zip area level. The sample size is 506,966,676.

4.2 Risk factors for heat-related hospitalization

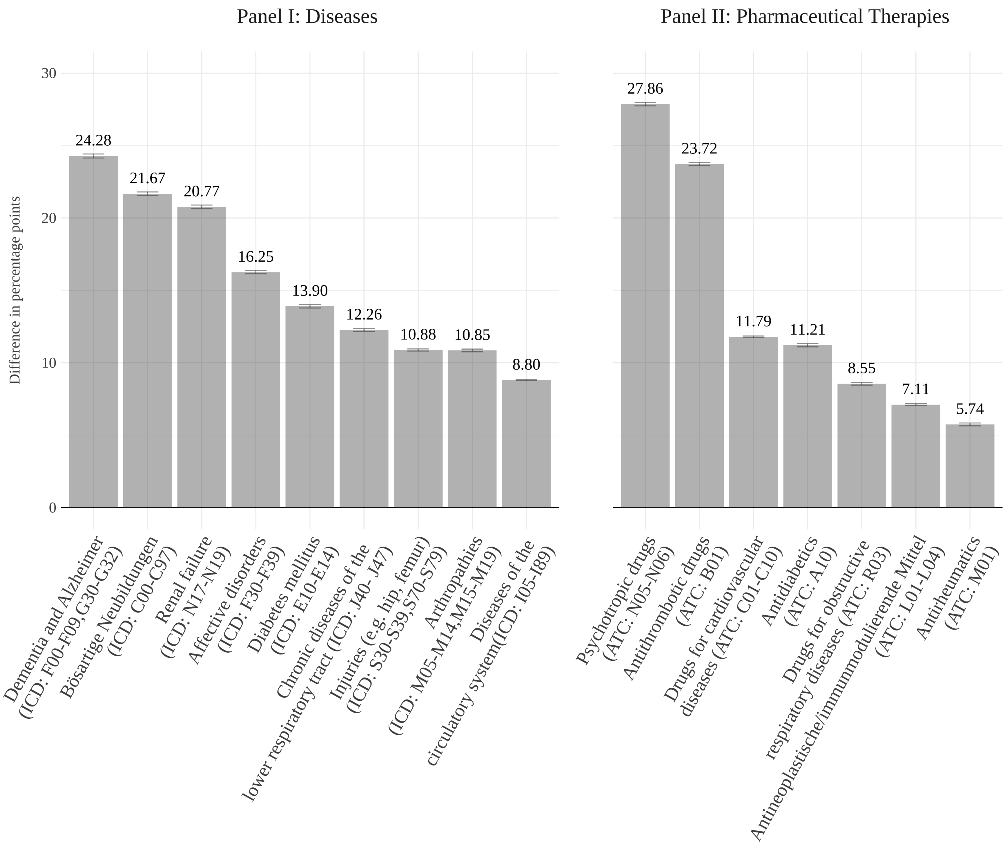
4.2.1 Individual risk factors

To identify individual-related risk factors of heat vulnerability, we compare the most vulnerable 1% of the elderly population with the 75% that exhibit below-average heat effects according to Figure 2. We find that the most vulnerable are, on average, about 3.27 years older, and their proportion of male individuals is 2.97 percentage points higher. Moreover, Figure 3 reveals that the most vulnerable suffer more often from various pre-existing diseases (Panel I) and depend more often on certain pharmaceuticals (II).[12] Concerning diseases, we observe the greatest difference in the prevalence of dementia and Alzheimer's disease. The share of individuals with this condition is about 24 percentage points higher among the most vulnerable than in the comparison group. This observation might be related to the great group differences that we observe for affective disorders such as depression and the intake of psychotropic drugs. In German nursing homes, dementia patients are (too) frequently given psychotropic drugs to manage their behavioral issues (Thürmann 2017). A greater dehydration risk of these patients (Easterling and Robbins 2008, Mentes 2006) could contribute to their vulnerability to heat. At the same time, numerous studies find that heat impacts diseases caused by dehydration in particular (Li et al. 2015, Bobb et al. 2014). These diseases include renal insufficiency and diabetes, amongst others, for which we also observe a significantly greater prevalence among the most vulnerable. Panel II shows that of the most vulnerable, a 23 percentage point higher proportion takes anti-thrombotic drugs, which mainly prevent heart attacks, strokes, embolisms, or leg vein thrombosis. In comparison, the difference in the share of individuals with a cardiovascular condition reported in Panel I, has a significantly smaller

[12] Note, that the identified group differences in Figure 3 are robust to holding age and sex constant.

magnitude. This may be because we consider a broadly defined group of diseases (ICD-10 I05-I89). Diverging evidence in other studies (e.g. Phung et al. 2016, Li et al. 2015) suggests that heat could affect particular diseases within this group very differently (Bobb et al. 2014). In addition, we do not observe whether individuals suffer from multiple diseases within the broad group or just a single one.

Figure 3: Morbidity differences between the most and the least vulnerable



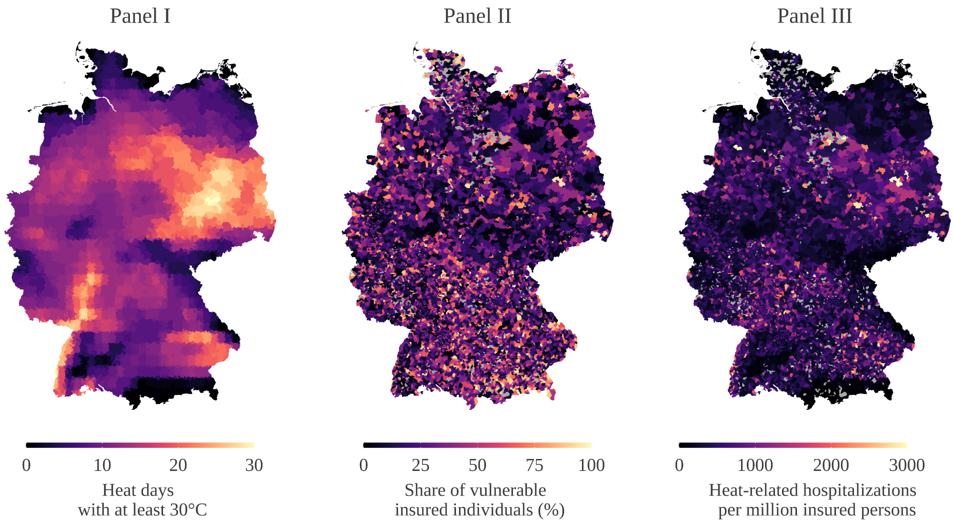
The figures illustrate our CLAN results. The estimated coefficients indicate the average difference between insured individuals in the top 1% percentile and the bottom 75% percentile of the predicted heat vulnerability distribution concerning various diseases (Panel I) and pharmaceutical therapies (Panel II). Standard errors are clustered at the zip area level and coefficients, are plotted with 95% confidence intervals. The sample size is 385,294,673.

4.2.2 Geographic risk factors

Next, we examine the geographic distribution of the vulnerable population and identify regional risk factors. For this analysis, we focus on the extremely hot year 2018.[13] With an average temperature of 2.2 degrees above the value of the internationally valid reference period 1961 to 1990, it was the warmest year since 1881 (DWD 2018). Panel I in Figure 4 illustrates the number of heat days and shows that in particular Berlin and Brandenburg and the Rhine-Main area were strongly affected. We compare this geographic distribution of heat exposure to the distribution of the vulnerable population. Panel II shows the proportion of individuals per zip area that belong to the most vulnerable quarter with above-average heat risk according to Figure 2. Accordingly, proportions below 25% mean that the local population is, on average, less vulnerable than the average elderly individual in Germany. Proportions above 25% indicate a disproportionately vulnerable population. The maps in Panel I and II tentatively indicate a negative relationship between heat exposure and vulnerability. However, Panel II highlights a large dispersion in the heat vulnerability. In some zip code areas all individuals are classified as most vulnerable, while other zip code areas show no individuals in this group. Moreover, the populations of directly neighboring areas often differ substantially in their heat vulnerability.

[13] Focusing on other years in our sample, changes our results only with regard to heat exposure (Panel I, Figure 4). The predicted vulnerability of the population in each zip area (Panel II, Figure 4) does not change markedly over time.

Figure 4: Geographic distribution of heat days, heat vulnerability, and heat-related hospitalizations in 2018



This figure compares the distribution of heat exposure, heat vulnerability, and heat damage across Germany in 2018. Panel I shows the distribution of heat days with temperatures of at least 30 °C, Panel II the proportion of insured individuals in each ZIP area who have a predicted heat vulnerability in the top 25% percentile, and Panel III the number of heat-related hospitalizations per million individuals. We calculate the numbers in Panel III based on the regression coefficients in Figure 2. Areas with fewer than 100 individuals in Panels II and III are colored gray.

Finally, we approximate the total heat-related hospitalizations that occurred in 2018, which are a product of local heat exposure (Figure 4, Panel I) and local heat vulnerability (Panel II). To this end, we multiply the estimated coefficients in Figure 2 with the number of AOK-insured individuals per percentile group and the number of heat days in 2018 for each zip area and divide the quantities by a million. Panel III in Figure 4 shows how the heat-related hospitalizations distribute across the zip areas. The map indicates that heat damage concentrates along two parallel bands: one running from the northern-west (Weser-Ems area) to the southern-east (Lower Lusatia) and one from the southern-west (Rhine-Main area) to the south (Lower Bavaria).

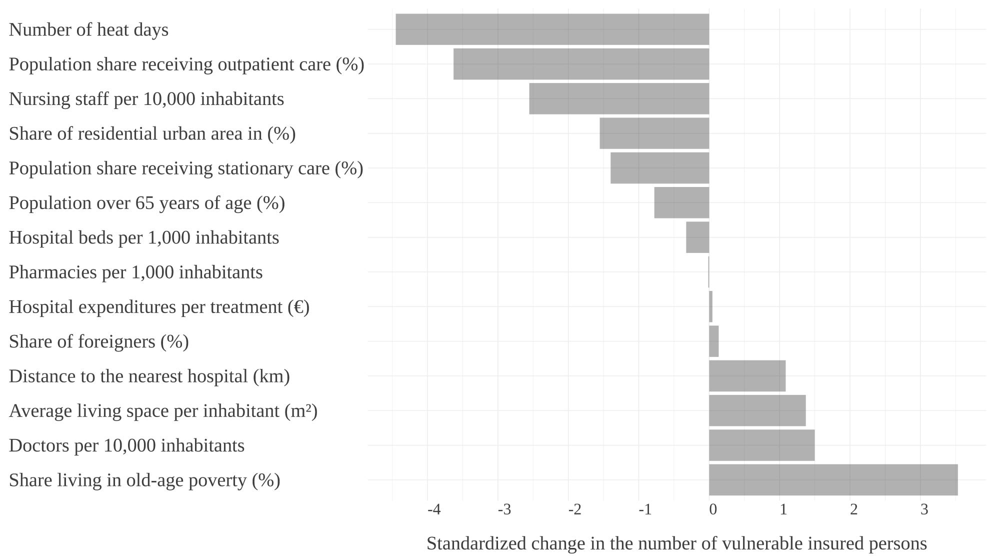
4.2.3 Demographic, socioeconomic, and infrastructural risk factors

We subsequently examine whether the remarkable variation in the share of vulnerable individuals across zip areas is related to local socioeconomic and demographic factors, as well as characteristics of local healthcare infrastructure. To this end, we estimate a LASSO regression where the dependent variable is the number of individuals in each zip area that belong to the most vulnerable quarter.[14] In the regressions, we control for the total number of insured individuals in each zip area and for structural differences in population vulnerability across regions (eastern, southern, western, and northern Germany). Figure 5 lists the most relevant explanatory variables of our model on the left-hand side. The plotted bars represent the standardized correlations between these variables and the number of vulnerable insured individuals. They have no causal interpretation. We again find that vulnerable individuals live more often in areas with fewer heat days. We also observe that variables related to the provision of nursing care are negatively associated with heat vulnerability: in areas where a greater share of care-dependent elderly receives assistance from professional outpatient care services or in nursing homes, there is a lower number of vulnerable individuals. At the same time, the number of nursing care staff per inhabitant is higher in these areas. A possible explanation could be that heat damage is easier to prevent if professional nurses are around and patients are not living alone or cared for at home by relatives only. Additionally, our model identifies that individuals living in an area that is urban, has a greater number of hospital beds, and where the share of the elderly in the total population is higher tend to be less at risk of a hospital admission due to heat. On the other hand, we find a positive correlation between heat

[14] Figure 7 in the Appendix demonstrates that the results remain similar when using the share of vulnerable individuals as the dependent variable instead.

vulnerability and old-age poverty. This is in line with other studies showing that a lower socioeconomic status can contribute to heat vulnerability (**li2015heat**, Campbell et al. 2018). Likewise, increased expenditures on hospital treatments in the previous year, which reflect the morbidity of the local population, predict an increased risk. Variables typically linked to rural areas, such as a longer driving distance to the nearest hospital and more living space per inhabitant, positively correlate with heat vulnerability. This could be linked to various reasons, such as a less dense medical infrastructure or less experience with the management of heat and use of warning systems and heat action plans (Jagai et al. 2017). Somewhat surprisingly, we find that heat vulnerability is positively associated with the number of outpatient physicians. One explanation could be that the presence of physicians increases the likelihood of detecting emergencies and sending patients to the hospital to prevent worse harm.

Figure 5: Demographic, socioeconomic, and infrastructural risk factors of heat-related hospitalization



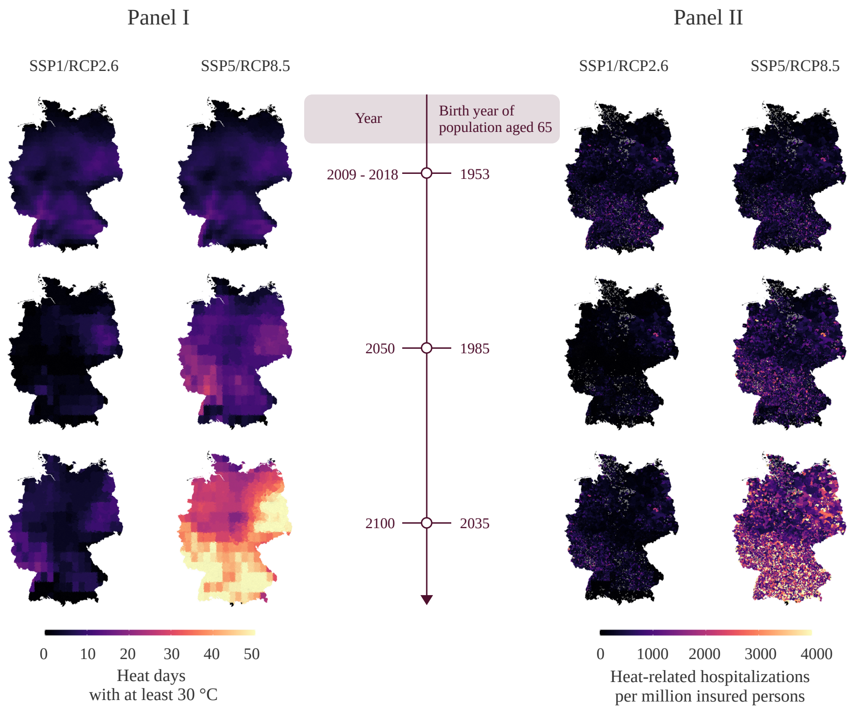
The figure shows demographic, socioeconomic, and infrastructural characteristics of zip areas that predict the number of insured individuals with a heat vulnerability in the top 25% percentile. The bars represent the estimated coefficients from a LASSO regression which assigns zero-coefficients to irrelevant variables. Hospital expenditures refer only to the AOK-insured population of the zip area and the previous year. We exclude zip areas with fewer than 100 insured individuals from the regression. The sample size is 7,322.

4.3 Heat-related hospitalizations in a changing climate

Our analysis shows that more vulnerable populations tend to live in areas with less heat exposure. However, unabated climate change is causing temperatures to rise and will affect these people increasingly. In the following, we assess the potential of climate change to affect heat-related hospitalizations.

In Figure 6, we consider two possible climate change scenarios from the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). The "SSP1/RCP2.6" scenario corresponds to a best-case scenario in which strict climate policy ensures that the global temperature increase stays well below 2 °C by 2100. The "SSP5/RCP8.5" scenario represents a worst-case scenario in which the absence of any actions to reduce CO_2 emissions leads to a temperature rise of 4.7 to 5.1 °C. Panel I shows the average number of heat days in the years 2009 through 2018 and the projections of heat days for the years 2050 and 2100 in both scenarios. Panel II illustrates the projected heat-related hospitalizations per million individuals, respectively. It shows the distribution of hospitalizations that would result, *ceteris paribus*, if the elderly population from 2018 lived in a climate of the future.

Figure 6: Projections of future heat days and heat-related hospitalizations



The figure projects heat exposure and heat-related hospitalizations into the two climate scenarios SSP1/RCP2.6 and SSP5/RCP8.5. Panel I shows the number of heat days with at least 30 °C per year and zip area. Panel II shows the number of heat-related hospitalizations per million insured individuals, year, and zip area. We calculate these numbers based on the predicted heat vulnerability of the insured individuals in 2018 and the regression coefficients that Figure 2 illustrates. Areas with fewer than 100 insured individuals in Panel II are colored in gray.

Overall, Figure 6 illustrates that strict climate policies would maintain the *status quo* of heat-related hospitalization. In contrast, health damages from heat increase sharply in the scenario with high CO_2 emissions. We estimate that compared to the average in 2009 through 2018, heat-related hospitalizations would increase by 87% in 2050 and 486% in 2100. Therefore, missing the 2 °C target can have severe health consequences not only for future generations. Today's 35-year-old will spend their entire lifetime from the age of 65 under a heat exposure suggested by the scenarios for 2050 and 2100. While this back-of-the-envelope projection does not suffice to map future health damages accurately, it does serve as an indicator of the potential of efficient climate policy.

5 Conclusion

The present study shows that heat damages distribute highly unevenly across the elderly population. About a quarter of those over 65 years of age exhibit an above-average risk of hospitalization due to heat. For the most vulnerable among them, the heat effect on hospitalization is nearly 14 times higher than for the average individual. We identify several individual risk factors. On average, the most vulnerable are older, more often male, and they have worse health conditions. In particular, we find that dementia, Alzheimer's disease, and other conditions linked to an increased risk of dehydration are much more prevalent. We also identify several regional risk factors. It appears that vulnerable individuals are more likely to reside in rural areas, with less nursing care, and with more old-age poverty, for instance. Remarkably, our analysis also reveals that they tend to live in places with relatively little heat exposure under current climate conditions. Global warming could significantly increase the exposure of these vulnerable populations in the future. We estimate that unabated climate change could lead up to a 486%-increase in heat-related hospitalization by 2100.

Our analysis contributes evidence to an increasingly relevant area of tension. Due to demographic change, the share of the elderly population in Germany is rising sharply. At the same time, there are ongoing debates about how bottlenecks in the already overburdened care sector can be alleviated. With progressing climate change, the efficient allocation of resources for health protection will only become more relevant. In order to implement measures in a targeted manner, knowledge about the differences in the risk profiles within the population is crucial. Our analysis provides insights on a much smaller scale than previous studies have been able to map that may support the design of local adaptation measures.

However, it also makes clear that climate change mitigation will be critical in determining the magnitude of future health damage.

In favor of a comprehensive heterogeneity analysis, our study focuses on a binary treatment and makes several simplifying assumptions. For instance, it does not account for the fact that an average heat day is unlikely to have the same detrimental health impact as a day of a heat wave with persistently high daytime and nighttime temperatures (Bobb et al. 2014). Moreover, it neglects the dose-response relationship between temperature and hospitalization above and below 30 °C and assumes that health effects occur only on the day of heat exposure without any delay. We acknowledge that these assumptions simplify complex relationships that exist in reality (Hsiang 2016, Li et al. 2015, Deschênes and Greenstone 2011) but are not the subject of our analysis. Future research could better integrate these determinants of heat exposure. Combining the insurance data with information on individual behavior, such as exercising, nutrition, and time use data could provide further valuable insights but requires access to additional data. Finally, while the elderly represent a large and important fraction of the German population, we hope that future research can make progress regarding heterogeneity in the heat vulnerability of younger age groups, such as children or the working population.

References

Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry (2014). Self-selection into credit markets: Evidence from agriculture in mali. Tech. rep. National Bureau of Economic Research.

Bobb, J. F., Z. Obermeyer, Y. Wang, and F. Dominici (2014). Cause-specific risk of hospital admission related to extreme heat in older adults. *Jama* 312(24). Publisher: American Medical Association, pp. 2659–2667.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2021). Do female role models reduce the gender gap in science? Evidence from French high schools.

Campbell, S., T. A. Remenyi, C. J. White, and F. H. Johnston (2018). Heatwave and health impact research: A global review. *Health & place* 53. Publisher: Elsevier, pp. 210–218.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Tech. rep. National Bureau of Economic Research.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. en. *American Economic Review* 109(12), pp. 4178–4219. ISSN: 0002-8282. DOI: 10.1257/aer.20180279. URL: https://pubs.aeaweb.org/doi/10.1257/aer.20180279 (visited on 03/02/2023).

Deryugina, T., N. Miller, D. Molitor, and J. Reif (2021). Geographic and socioeconomic heterogeneity in the benefits of reducing air pollution in the United States. *Environmental and energy policy and the economy* 2(1). Publisher: The University of Chicago Press Chicago, IL, pp. 157–189.

Deschenes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics* 46. Publisher: Elsevier, pp. 606–619.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics* 3(4), pp. 152–85.

Destatis (2017). *Diagnosedaten der Patienten und Patientinnen in Krankenhäusern (einschl. Sterbe- und Stundenfälle) - Fachserie 12 Reihe 6.2.1 - 2016 (letzte Ausgabe - berichtsweise eingestellt)*. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Krankenhaeuser/Publikationen/Downloads-Krankenhaeuser/diagnosedaten-krankenhaus-2120621167004.html.

DIMDI. *Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme [website]*. Published: German Institute for Medical Documentation and Information (DIMDI) [distributor]. https://www.dimdi.de/static/de/klassifikatione 10-gm/kode-suche/htmlgm2022/.

DWD (2018). *Deutschlandwetter im Jahr 2018)*. URL: https://www.dwd.de/DE/presse/pressemitteilungen/DE/2018/20181228_deutschlandwetter_jahr2018_news.html.

Easterling, C. S. and E. Robbins (2008). Dementia and dysphagia. *Geriatric Nursing* 29(4). Publisher: Elsevier, pp. 275–285.

Einav, L., A. Finkelstein, S. Mullainathan, and Z. Obermeyer (2018). Predictive modeling of US health care spending in late life. *Science* 360(6396). Publisher: American Association for the Advancement of Science, pp. 1462–1465.

Eisenman, D. P., H. Wilhalme, C.-H. Tseng, M. Chester, P. English, S. Pincetl, A. Fraser, S. Vangala, and S. K. Dhaliwal (2016). Heat Death Associations with the built environment, social vulnerability and their interactions with rising temperature. *Health & Place* 41. Publisher: Elsevier, pp. 89–99.

Gronlund, C. J., A. Zanobetti, J. D. Schwartz, G. A. Wellenius, and M. S. O'Neill (2014). Heat, heat waves, and hospital admissions among the elderly in the United States, 1992–2006. *Environmental Health Perspectives* 122(11). Publisher: NLM-Export, pp. 1187–1192.

Hondula, D. M., R. E. Davis, M. J. Leisten, M. V. Saha, L. M. Veazey, and C. R. Wegner (2012). Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983-2008: a case-series analysis. *Environmental Health* 11(1). Publisher: BioMed Central, pp. 1–11.

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics* 8. Publisher: Annual Reviews, pp. 43–75.

ISIMIP (2020). *ISIMIP3b simulation round simulation protocol - all sectors combined*. URL: https://protocol.isimip.org/protocol/ISIMIP3b/index.html#input-data.

Jagai, J. S., E. Grossman, L. Navon, A. Sambanis, and S. Dorevitch (2017). Hospitalizations for heat-stress illness varies between rural and urban areas: an analysis of Illinois data, 1987–2014. *Environmental Health* 16(1). Publisher: Springer, p. 38.

Karlsson, M. and N. R. Ziebarth (2018). Population health effects and health-related costs of extreme temperatures: Comprehensive evidence from Germany. *Journal of Environmental Economics and Management* 91. Publisher: Elsevier, pp. 93–117.

Lange, S. (2019). Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1. 0). *Geoscientific Model Development* 12(7).

Li, M., S. Gu, P. Bi, J. Yang, and Q. Liu (2015). Heat waves and morbidity: current knowledge and further direction-a comprehensive literature review. *International Journal of Environmental Research and Public Health* 12(5). Publisher: Multidisciplinary Digital Publishing Institute, pp. 5256–5283.

Mentes, J. (2006). Oral hydration in older adults: Greater awareness is needed in preventing, recognizing, and treating dehydration. *AJN The American Journal of Nursing* 106(6). Publisher: LWW, pp. 40–49.

Phung, D., P. K. Thai, Y. Guo, L. Morawska, S. Rutherford, and C. Chu (2016). Ambient temperature and risk of cardiovascular hospitalization: An updated systematic review and meta-analysis. *Science of The Total Environment* 550. Publisher: Elsevier, pp. 1084–1102.

Schwartz, J. (2005). Who is sensitive to extremes of temperature?: A case-only analysis. *Epidemiology* 16(1). Publisher: LWW, pp. 67–72.

Thürmann, P. A. (2017). 11 Einsatz von Psychopharmaka bei Pflegebedürftigen. *Pflege-Report 2017. Schwerpunkt: Die Versorgung der Pflegebedürftigen. Jacobs K, Kuhlmey A, Greß S, Klauber J, Schwinger A (eds.), Schattauer, Stuttgart*, pp. 119–151.

Van Donkelaar, A., R. V. Martin, C. Li, and R. T. Burnett (2019). *Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors: European Regional Estimates (V4.EU.03) [dataset]*. Published: Environmental Science & Technology, 53(5): 2595-2611. https://sites.wustl.edu/acag/datasets pm2-5/.

Watts, N., M. Amann, N. Arnell, S. Ayeb-Karlsson, J Beagley, K. Belesova, and A Costello (2020). The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises. *Lancet*.

Appendix

Appendix Tables

Table 2: Variable overview

No.	Variable	Description	Type	Temporal resolution	Unit of observation [1]	Source
Data on AOK-insured elderly individuals						
1	Sex	Sex of the insured individual	binary	time-invariant	individual	AOK Research Institute (WIdO)
2	Age	Age of the insured individual	continuous	day	individual	WIdO
3	Place of residence	5-digit zip code area of place of residence of the insured individual	factor	quarter	individual	WIdO
4	Hospital treatment expenditures	Expenditures per hospital treatment in year 2017	continuous	2017	zip area	WIdO
Diseases (ICD-10 Code)						
5	C00-C14	Malignant neoplasms of lip, oral cavity and pharynx	binary	quarter	individual	WIdO
6	C15-C26	Malignant neoplasms of digestive organs	binary	quarter	individual	WIdO
7	C30-C39	Malignant neoplasms of respiratory and intrathoracic organs	binary	quarter	individual	WIdO
8	C40-C41	Malignant neoplasms of bone and articular cartilage	binary	quarter	individual	WIdO
9	C43-C44	Melanoma and other malignant neoplasms of skin	binary	quarter	individual	WIdO
10	C45-C49	Malignant neoplasms of mesothelial and soft tissue	binary	quarter	individual	WIdO
11	C50	Malignant neoplasm of breast	binary	quarter	individual	WIdO
12	C51-C58	Malignant neoplasms of female genital organs	binary	quarter	individual	WIdO
13	C60-C63	Malignant neoplasms of male genital organs	binary	quarter	individual	WIdO
14	C64-C68	Malignant neoplasms of urinary tract	binary	quarter	individual	WIdO
15	C69-C72	Malignant neoplasms of eye, brain and other parts of central nervous system	binary	quarter	individual	WIdO
16	C73-C75	Malignant neoplasms of thyroid and other endocrine glands	binary	quarter	individual	WIdO
17	C76-C80	Malignant neoplasms of ill-defined, secondary and unspecified sites	binary	quarter	individual	WIdO
18	C81-C96	Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue	binary	quarter	individual	WIdO
19	C97	Malignant neoplasms of independent (primary) multiple sites	binary	quarter	individual	WIdO
20	D50-D53	Nutritional anaemias	binary	quarter	individual	WIdO
21	D55-D59	Haemolytic anaemias	binary	quarter	individual	WIdO
22	D60-D64	Aplastic and other anaemias	binary	quarter	individual	WIdO
23	E00-E07	Disorders of thyroid gland	binary	quarter	individual	WIdO
24	E10-E14	Diabetes mellitus	binary	quarter	individual	WIdO

25	E70-E90	Metabolic disorders	binary	quarter	individual	WIdO
26	F00-F09	Organic, including symptomatic, mental disorders	binary	quarter	individual	WIdO
27	F30-F39	Mood [affective] disorders	binary	quarter	individual	WIdO
28	G30-G32	Other degenerative diseases of the nervous system	binary	quarter	individual	WIdO
29	G40-G47	Episodic and paroxysmal disorders	binary	quarter	individual	WIdO
30	H25-H28	Disorders of lens	binary	quarter	individual	WIdO
31	H40-H42	Glaucoma	binary	quarter	individual	WIdO
32	I05-I09	Chronic rheumatic heart diseases	binary	quarter	individual	WIdO
33	I10-I15	Hypertensive diseases	binary	quarter	individual	WIdO
34	I20-I25	Ischaemic heart diseases	binary	quarter	individual	WIdO
35	I26-I28	Pulmonary heart disease and diseases of pulmonary circulation	binary	quarter	individual	WIdO
36	I30-I52	Other forms of heart disease	binary	quarter	individual	WIdO
37	I60-I69	Cerebrovascular diseases	binary	quarter	individual	WIdO
38	I70-I79	Diseases of arteries, arterioles and capillaries	binary	quarter	individual	WIdO
39	I80-I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified	binary	quarter	individual	WIdO
40	J00-J06	Acute upper respiratory infections	binary	quarter	individual	WIdO
41	J40- J47	Chronic lower respiratory diseases	binary	quarter	individual	WIdO
42	K55-K64	Other diseases of intestines	binary	quarter	individual	WIdO
43	M05-M14	Inflammatory polyarthropathies	binary	quarter	individual	WIdO
44	M15-M19	Arthrosis	binary	quarter	individual	WIdO
45	M40-M43	Deforming dorsopathies	binary	quarter	individual	WIdO
46	M45-M49	Spondylopathies	binary	quarter	individual	WIdO
47	M50-M54	Other dorsopathies	binary	quarter	individual	WIdO
48	M80-M85	Disorders of bone density and structure	binary	quarter	individual	WIdO
49	N17-N19	Renal failure	binary	quarter	individual	WIdO
50	N40-N51	Diseases of male genital organs	binary	quarter	individual	WIdO
51	S30-S39	Injuries to the abdomen, lower back, lumbar spine and pelvis	binary	quarter	individual	WIdO
52	S70-S79	Injuries to the hip and thigh	binary	quarter	individual	WIdO
Prescriptions of pharmaceuticals (ATC-Code)						
53	A10	Drugs used in diabetes	binary	quarter	individual	WIdO
54	B01	Antithrombotic agents	binary	quarter	individual	WIdO
55	B03	Antianemic preparations	binary	quarter	individual	WIdO
56	C01	Cardiac therapy	binary	quarter	individual	WIdO
57	C02	Antihypertensives	binary	quarter	individual	WIdO
58	C03	Diuretics	binary	quarter	individual	WIdO
59	C04	Peripheral vasodilators	binary	quarter	individual	WIdO
60	C05	Vasoprotectives	binary	quarter	individual	WIdO
61	C06	Other agents acting on the caridovascular system	binary	quarter	individual	WIdO
62	C07	Beta blocking agents	binary	quarter	individual	WIdO
63	C08	Calcium channel blockers	binary	quarter	individual	WIdO
64	C09	Agents acting on the Renin-Angiotensin system	binary	quarter	individual	WIdO
65	C10	Lipid modifying agents	binary	quarter	individual	WIdO
66	G04	Urologicals	binary	quarter	individual	WIdO
67	H02	Corticosteroids for systemic use	binary	quarter	individual	WIdO
68	H03	Thyroid therapy	binary	quarter	individual	WIdO
69	H05	Calcium homeostasis	binary	quarter	individual	WIdO
70	L01	Antineoplastic agents	binary	quarter	individual	WIdO
71	L02	Endocrine therapy	binary	quarter	individual	WIdO
72	L03	Immunostimulants	binary	quarter	individual	WIdO
73	L04	Immunosuppressants	binary	quarter	individual	WIdO
74	M01	Antiinflammatory and antirheumatic products	binary	quarter	individual	WIdO
75	M05	Drugs fortreatment of bone diseases	binary	quarter	individual	WIdO
76	N03	Antiepileptics	binary	quarter	individual	WIdO
77	N05	Psycholeptics	binary	quarter	individual	WIdO
78	N06	Psychoanaleptics	binary	quarter	individual	WIdO
79	R03	Drugs for obsructive airway diseases	binary	quarter	individual	WIdO
80	S01	Ophtalmologicals	binary	quarter	individual	WIdO

Weather conditions							
81	Cloud coverage	Min.	Minimum of the share of the area covered by clouds	continuous	day	zip area	European Centre for Medium-Term Weather Forecasting (ECMWF)
82		Mean	Average of the share of the area covered by clouds	continuous	day	zip area	ECMWF
83		Max.	Maximum of the share of the area covered by clouds	continuous	day	zip area	ECMWF
84	Relative humidity	Min.	Minimum water vapor saturation of the air (%)	continuous	day	zip area	ECMWF
85		Mean	Average water vapor saturation of the air (%)	continuous	day	zip area	ECMWF
86		Max.	Maximum water vapor saturation of the air (%)	continuous	day	zip area	ECMWF
87	Surface pressure	Min.	Minimum pressure (force per unit area) of the atmosphere on the land surface (Pa)	continuous	day	zip area	ECMWF
88		Mean	Average pressure (force per unit area) of the atmosphere on the land surface (Pa)	continuous	day	zip area	ECMWF
89		Max.	Maximum pressure (force per unit area) of the atmosphere on the land surface (Pa)	continuous	day	zip area	ECMWF
90	Precipitation	Min.	Minimum precipitation (liquid and frozen water, including rain and snow) falling on the earth's surface (m)	continuous	day	zip area	ECMWF
91		Mean	Average precipitation (liquid and frozen water, including rain and snow) falling on the earth's surface (m)	continuous	day	zip area	ECMWF
92		Max.	Maximum precipitation (liquid and frozen water, including rain and snow) falling on the earth's surface (m)	continuous	day	zip area	ECMWF
93	East wind	Min.	Minimum horizontal speed of air moving eastward at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
94		Mean	Average horizontal speed of air moving eastward at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
95		Max.	Maximum horizontal speed of air moving eastward at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
96	North wind	Min.	Minimum horizontal speed of air moving in the north direction at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
97		Mean	Average horizontal speed of air moving in the north direction at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
98		Max.	Maximum horizontal speed of air moving in the north direction at a height of ten meters above the earth's surface (m s-1)	continuous	day	zip area	ECMWF
99	Vertical velocity	Min.	Minimum velocity of vertical air movement (Pa s-1)	continuous	day	zip area	ECMWF
100		Mean	Average velocity of vertical air movement (Pa s-1)	continuous	day	zip area	ECMWF
101		Max.	Maximum velocity of vertical air movement (Pa s-1)	continuous	day	zip area	ECMWF
Air quality							
102	Ozone	Min.	Minimum mass of ozone per kilogram of air (kg kg-1)	continuous	day	zip area	ECMWF
103		Mean	Average mass of ozone per kilogram of air (kg kg-1)	continuous	day	zip area	ECMWF
104		Max.	Maximum mass of ozone per kilogram of air (kg kg-1)	continuous	day	zip area	ECMWF
105	PM2.5		Mean PM2.5 concentration (µg/m³)	continuous	year	zip area	Van Donkelaar et al. 2019
106	NO2		Mean NO2 concentration (µg/m³)	continuous	year	zip area	German Environment Agency (UBA)
Socioeconomic characteristics							
107	Socioeconomic status class very low		Proportion of households in zip area with very low socioeconomic status	continuous	2019	zip area	Acxiom [2]
108	Socioeconomic status class low		Proportion of households in zip area with low socioeconomic status	continuous	2019	zip area	Acxiom [2]
109	Socioeconomic status class rather low		Proportion of households in zip area with rather low socioeconomic status	continuous	2019	zip area	Acxiom [2]
110	Socioeconomic status class intermediate		Proportion of households in zip area with intermediate socioeconomic status	continuous	2019	zip area	Acxiom [2]
111	Socioeconomic status class rather high		Proportion of households in zip area with rather high socioeconomic status	continuous	2019	zip area	Acxiom [2]
112	Socioeconomic status class high		Proportion of households in zip area with high socioeconomic status	continuous	2019	zip area	Acxiom [2]
113	Socioeconomic status class very high		Proportion of households in zip area with very high socioeconomic status	continuous	2019	zip area	Acxiom [2]
114	Social service facilities		Number of nonresidential facilities that provide social services, such as advocacy, counseling, employment services, veterans services, housing services, wellness programs, and recreational activities, per 100,000 residents	continuous	2020	zip area	OpenStreetMap (OSM) [3]
115	Food serving facilities		Number of establishments that distribute packaged food, usually to the poor, for free or below market price, per 100,000 residents	continuous	2020	zip area	OSM [3]

116	Old-age poverty	Share of the population receiving basic subsistence income among residents 65 years of age and older in %	continuous	2017	county	German Federal Office for Building and Regional Planning (BBSR) [4]
117	Golf courses	Number of golf courses per 10,000 residents	continuous	2020	zip area	OSM [3]
118	Tennis courts	Number of tennis courts per 10,000 residents	continuous	2020	zip area	OSM [3]
			Infrastructure			
119	Commercial area	Proportion of the area used for commercial enterprises and their offices	continuous	2020	zip area	OSM [3]
120	Residential area	Proportion of the area used for residential purposes	continuous	2020	zip area	OSM [3]
121	Industrial area	Proportion of the area used for industrial uses, e.g. factories or warehouses	continuous	2020	zip area	OSM [3]
122	Airports	Number of airport passenger buildings	continuous	2020	zip area	OSM [3]
123	Highway	Indicator that is equal to one if a highway route passes through the area	binary	2020	zip area	OSM [3]
124	Public transport	Number of public transport stops	continuous	2020	zip area	OSM [3]
125	Helicopter pads	Euclidean distance to the nearest landing area designated for helicopters in emergency situations (km)	continuous	2020	zip area	OSM [3]
126	Emergency supply	Number of buildings and areas, for the storage of ambulance vehicles, medical equipment, protective equipment, and other medical supplies per 10,000 inhabitants	continuous	2020	zip area	OSM [3]
127	Clinic distance	Driving distance to the nearest medium-sized medical facility or health center from the centroid of the area (km)	continuous	2020	zip area	OSM [3]
128	Hospitl distance	Driving distance to the nearest hospital offering inpatient medical treatment from the centroid of the area (km)	continuous	2020	zip area	OSM [3]
129	Hospital beds	Hospital beds per 1,000 inhabitants	continuous	2016	county	BBSR [4]
130	Medical practices	Number of outpatient medical practices per 1,000 inhabitants	continuous	2020	zip area	OSM [3]
131	Doctors	Number of family physicians, general practitioners, and internists per 10,000 inhabitants	continuous	2017	county	BBSR [4]
132	Pharmacies	Pharmacies per 1,000 inhabitants	continuous	2020	zip area	OSM [3]
133	Defibrillators	Defibrillators per 1,000 inhabitants	continuous	2020	zip area	OSM [3]
134	Nursing homes	Number of nursing homes that provide 24-hour permanent care and assist the residents with the basic activities of daily living per 10,000 inhabitants. Nursing home residents include elderly or seriously ill people and those with physical and mental disabilities.	continuous	2020	zip area	OSM [3]
135	Senior citizen facilities	Number of facilities for the elderly and pensioners, e.g. retirement homes, per 10,000 inhabitants	continuous	2020	zip area	OSM [3]
136	Group residences	Number of group residential homes that provide social and medical services, (e.g recreational activities, meals, supervision) for specific population groups per 10,000 inhabitants	continuous	2020	zip area	OSM [3]
137	Assisted living	Number of facilities for assisted living, e.g., for people with disabilities or the elderly, often with caregiver present and emergency button installed, per 10,000 inhabitants	continuous	2020	zip area	OSM [3]
138	Inpatient care	Percentage share of the population receiving permanent inpatient care of the total care-dependent population	continuous	2017	county	BBSR [4]
139	Outpatient care	Percentage share of the population receiving outpatient care of the total care-dependent population	continuous	2017	county	BBSR [4]
140	Nursing staff	Nursing staff employed in nursing homes and in outpatient care services per 10,000 inhabitants	continuous	2017	county	BBSR [4]
			Demographic characteristics			
141	Population size	Number of inhabitant	continuous	2011	zip area	Statistic departments of the federation and the federal states [5]
142	Territory size	Area in square meters	continuous	2020	zip area	OSM [3]
143	Population density	Number of inhabitants per square meter	continuous	2011	zip area	Statistic departments of the federation and the federal states[5] OSM [3]
144	Average age	Average age of the population in years	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
145	Young population	Percentage share of the population younger than 18 years of age in the total population	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]

146	Old population		Percentage share of the population older than 65 years of age in the total population	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
147	Share of foreigners		Percentage share of foreigners in the total population	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
148	Average household size		Average number of individuals belonging to one household	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
149	Flat vacancy		Proportion of flats that are vacant	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
150	Living space per inhabitant		Average living space per resident in square meters	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
151	Wohnraum je Wohnung		Durchschnittliche Wohnfläche je Wohnung in m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[5]
152	Distribution of household size	1 person	Percentage share of households including one person	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
153		2 persons	Percentage share of households including two persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
154		3 persons	Percentage share of households including three persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
155		4 persons	Percentage share of households including four persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
156		5 persons	Percentage share of households including five persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
157		6 and more persons	Percentage share of households including six and more persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
158	Household type	Couples without child(ren)	Percentage share of households consisting of couples without child(ren)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
159		Couples with child(ren)	Percentage share of households consisting of couples with child(ren)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
160		Single parents	Percentage share of households with single parent(s)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
161		Multi-person households without nuclear family	Percentage share of multi-person households without nuclear family	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
162	Marital status	Married couples	Percentage share of households with married couples	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
163		Registered partnerships	Percentage share of households with registered partnerships	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
164		Non-marital partnerships	Percentage share of households with non-marital partnerships	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
165		Single mothers	Percentage share of households with single mothers	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

166		Single fathers	Percentage share of households with single fathers	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
167	Size of nuclear family	2 persons	Percentage share of households with a nuclear family with 2 persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
168		3 persons	Percentage share of households with a nuclear family with 3 persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
169		4 persons	Percentage share of households with a nuclear family with 4 persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
170		5 persons	Percentage share of households with a nuclear family with 5 persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
171		6 and more persons	Percentage share of households with a nuclear family with 6 and more persons	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
172	Family type	Married couples without child	Percentag share of households consisting of married couples without children	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
173		Married couples, at least 1 child younger than 18	Percentag share of households consisting of married couples with at least one child under 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
174		Married couples, all children older than 18	Percentag share households consisting of married couples with children over 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
175		Registered partnerships without children	Percentage share of households consisting of registered partnerships without children	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
176		Registered partnerships, at least one child younger than 18	Percentage share of households consisting of registered partnerships with at least one child under 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
177		Registered civil partnerships, all children older than 18	Percentage share of households consisting of registered civil partnerships with children over 18	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
178		Non-marital cohabitation without children	Percentage share of households consisting of non-marital partnerships without children	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
179		Non-marital cohabitation, at least 1 child younger than 18	Percentage share of households consisting of non-marital cohabiting couples with at least one child under 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
180		Non-marital cohabitation, all children older than 18	Percentage share of households consisting of non-marital cohabiting couples with children over 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
181		Single fathers, at least 1 child younger than 18	Percentage share of households consisting of single fathers with at least one child under 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
182		Single fathers, all children older than 18	Percentage share of households consisting of single fathers with children over 18	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
183		Single mothers, at least 1 child younger than 18	Percentage share of households consisting of single mothers with at least one child under 18 years of age	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
184		Single mothers, all children older than 18	Percentage share of households consisting of single mothers with children over 18	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

185	Senior citizen status	Households with exclusively senior citizens	Percentage share of households with only senior citizens	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
186		Households with seniors and younger people	Percentage share of households with senior citizens and younger people	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
187		Households without senior citizens	Percentage share of households without senior citizens	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
188	Construction year	Before 1919 (buildings)	Percentage share of buildings built before 1919	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
189		Before 1919 (flats)	Percentage share of flats in buildings built before 1919	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
190		1919 – 1948 (buildings)	Percentage share of buildings built in 1919 – 1948	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
191		1919 – 1948 (flats)	Percentage share of flats in buildings built in 1919 – 1948	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
192		1949 – 1978 (buildings)	Percentage share of buildings built in 1949 – 1978	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
193		1949 – 1978 (flats)	Percentage share of flats in buildings built in 1949 – 1978	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
194		1979 – 1986 (buildings)	Percentage share of buildings built in 1979 – 1986	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
195		1979 – 1986 (flats)	Percentage share of flats in buildings built in 1979 – 1986	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
196		1987 – 1990 (buildings)	Percentage share of buildings built in 1987 – 1990	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
197		1987 – 1990 (flats)	Percentage share of flats in buildings built in 1987 – 1990	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
198		1991 – 1995 (buildings)	Percentage share of buildings built in 1991 – 1995	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
199		1991 – 1995 (flats)	Percentage share of flats in buildings built in 1991 – 1995	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
200		1996 – 2000 (buildings)	Percentage share of buildings built in 1996 – 2000	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
201		1996 – 2000 (flats)	Percentage share of flats in buildings built in 1996 – 2000	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
202		2001 – 2004 (buildings)	Percentage share of buildings built in 2001 – 2004	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
203		2001 – 2004 (flats)	Percentage share of flats in buildings built in 2001 – 2004	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
204		2005 – 2008 (buildings)	Percentage share of buildings built in 2005 – 2008	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

205		2005 – 2008 (flats)	Percentage share of flats in buildings built in 2005 – 2008	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
206		2009 and later (buildings)	Percentage share of buildings built in 2009 or later	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
207		2009 and later (flats)	Percentage share of flats in buildings built in 2009 or later	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
208	Ownership type	Community of apartment owners (buildings)	Percentage share of buildings owned by a community of apartment owners	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
209		Community of apartment owners (flats)	Percentage share of flats in buildings owned by a community of apartment owners	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
210		Private person/s (buildings)	Percentage share of buildings owned by private individuals	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
211		Private person/s (flats)	Percentage share of flats in buildings owned by private individuals	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
212		Housing cooperative (buildings)	Percentage share of buildings owned by housing cooperative	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
213		Housing cooperative (flats)	Percentage share of flats in buildings owned by housing cooperative	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
214		Municipality or municipal housing company (buildings)	Percentage share of buildings owned by municipalities or municipal housing companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
215		Municipality or municipal housing company (flatss)	Percentage share of flats in buildings owned by municipalities or municipal housing companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
216		Private housing company (buildings)	Percentage share of buildings owned by private-sector housing companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
217		Private housing company (flats)	Percentage share of flats in buildings owned by private-sector housing companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
218		Other private company (buildings)	Percentage share of buildings owned by other private-sector companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
219		Other private company (flats)	Percentage share of flats in buildings owned by other private-sector companies	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
220		Federal or state (buildings)	Percentage share of buildings owned by the federal or state government	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
221		Federal or state (flats)	Percentage share of ats in buildings owned by the federal or state government	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
222		Non-profit organization (e.g. church) (buildings)	Percentage share of buildings owned by non-profit organizations (e.g. churches)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
223		Non-profit organization (e.g. church) (flats)	Percentage share of flats in buildings owned by non-profit organizations (e.g. churches)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
224	Type of building	Residential buildings (excluding dormitories) (buildings)	Percentage share of residential buildings (excluding dormitories)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

225		Residential buildings (excluding dormitories) (flats)	Percentage share of flats in residential buildings (excluding dormitories)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
226		Dormitory (buildings)	Percentage share of dormitories	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
227		Dormitory (flats)	Percentage share of flats in dormitories	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
228		Other building with living space (buildings)	Percentage share of other buildings with residential space in %.	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
229		Other building with living space (flats)	Percentage share of flats in in other buildings with living space	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
230	Building type size	Detached single family house (buildings)	Percentage share of buildings that are detached single-family homes	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
231		Detached single family house (flats)	Percentage share of flats in buildings that are detached single-family homes	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
232		Detached house: semi-detached house (buildings)	Percentage share of buildings that are detached single-family houses and semi-detached houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
233		Detached house: semi-detached house (flats)	Percentage share of flats in buildings that are detached single-family houses and semi-detached houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
234		Single family house: terraced house (buildings)	Percentage share of buildings that are detached single-family houses and terraced houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
235		Single family house: terraced house (flats)	Percentage share of flats in buildings that are detached single-family houses and terraced houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
236		Detached two-family house (buildings)	Percentage share of buildings that are detached two-family houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
237		Detached two-family house (flats)	Percentage share of flats in buildings that are detached two-family houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
238		Two-family house: semi-detached house (buildings)	Percentage share of buildings that are detached two-family houses and semi-detached houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
239		Two-family house: semi-detached house (flats)	Percentage share of flats in buildings that are detached two-family houses and semi-detached houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
240		Two-family house: terraced house (buildings)	Percentage share of buildings that are detached two-family houses and terraced houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
241		Two-family house: terraced house (flats)	Percentage share of flats in buildings that are detached two-family houses and terraced houses	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
242		Multi-family house: 3-6 flats (buildings)	Percentage share of buildings that are multi-family houses with 3-6 flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
243		Multi-family house: 3-6 flats (flats)	Percentage share of flats in buildings that are multi-family houses with 3-6 flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
244		Multi-family house: 7-12 flats (buildings)	Percentage share of buildings that are multi-family houses with 7-12 flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

245		Multi-family house: 7-12 flats (flats)	Percentage share of flats in buildings that are multi-family houses with 7-12 flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
246		Multi-family house: 13 and more flats (buildings)	Percentage share of buildings that are multi-family houses with 13 and more flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
247		Multi-family house: 13 and more flats (flats)	Percentage share of flats in buildings that are multi-family houses with 13 and more flats	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
248	Type of heating	District heating (buildings)	Percentage share of buildings with district heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
249		District heating (flats)	Percentage share of flats in buildings with district heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
250		Floor heating (buildings)	Percentage share of buildings with floor heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
251		Floor heating (flats)	Percentage share of flats in buildings with floor heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
252		Block heating (buildings)	Percentage share of buildings with block heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
253		Block heating (flats)	Percentage share of flats in buildings with block heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
254		Central heating (buildings)	Percentage share of buildings with central heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
255		Central heating (flats)	Percentage share of flats in buildings with central heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
256		Single/multi-room furnaces (also night storage heating) (buildings)	Percentage share of buildings with single/multi-room furnaces (also night storage heating)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
257		Single/multi-room furnaces (also night storage heating) (flats)	Percentage share of flats in buildings with single/multi-room furnaces (also night storage heating)	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
258		No heating in the building or apartments (buildings)	Percentage share of buildings without heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
259		No heating in the building or apartments (flats)	Percentage share of flats in buildings without heating	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
260	Use according to occupancy by household	Ownership: with currently headed household	Percentage share of apartments occupied by the owner with the currently headed household	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
261		Ownership: no currently headed household	Percentage share of apartments occupied by the owner without the current household	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
262		Rented: with currently headed household	Percentage share of apartments rented out for residential purposes with the currently managed household	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
263		Rented: no currently headed household	Percentage share of apartments rented out for residential purposes excluding the currently managed household	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

264		Vacation and leisure apartment	Percentage share of apartments that are vacation and leisure apartments	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
265		Vacant	Percentage share of apartments that are vacant	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
266	Number of rooms	1 room	Percentage share of apartments with one room	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
267		2 rooms	Percentage share of apartments with two rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
268		3 rooms	Percentage share of apartments with three rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
269		4 rooms	Percentage share of apartments with four rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
270		5 rooms	Percentage share of apartments with five rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
271		6 rooms	Percentage share of apartments with six rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
272		7 and more rooms	Percentage share of apartments with seven and more rooms	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
273	Apartment size	Under 30	Percentage share of apartments with less than 30 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
274		30 - 39	Percentage share of apartments with 30 - 39 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
275		40 - 49	Percentage share of apartments with 40 - 49 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
276		50 - 59	Percentage share of apartments with 50 - 59 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
277		60 - 69	Percentage share of apartments with 60 - 69 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
278		70 - 79	Percentage share of apartments with 70 - 79 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
279		80 - 89	Percentage share of apartments with 80 - 89 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
280		90 - 99	Percentage share of apartments with 90 - 99 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
281		100 - 109	Percentage share of apartments with 100 - 109 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
282		110 - 119	Percentage share of apartments with 110 - 119 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
283		120 - 129	Percentage share of apartments with 120 - 129 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

284		130 - 139	Percentage share of apartments with 130 - 139 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
285		140 - 149	Percentage share of apartments with 140 - 149 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
286		150 - 159	Percentage share of apartments with 150 - 159 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
287		160 -169	Percentage share of apartments with 160 -169 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
288		170 - 179	Percentage share of apartments with 170 - 179 m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
289		180 and more	Percentage share of apartments with 180 and more m²	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
290	Age (groups of 10)	Under 10	Percentage share of people younger than 10 years old	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
291		10 - 19	Percentage share of people aged 10-19	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
292		20 - 29	Percentage share of people aged 20 - 29	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
293		30 - 39	Percentage share of people aged 30 - 39	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
294		40 - 49	Percentage share of people aged 40 - 49	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
295		50 - 59	Percentage share of people aged 50 - 59	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
296		60 - 69	Percentage share of people aged 60 - 69	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
297		70 - 79	Percentage share of people aged 70 - 79	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
298		80 and older	Percentage share of people aged 80 and older	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
299	Marital status (detailed)	Single	Percentage share of people who are single	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
300		Married	Percentage share of people who are married	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
301		Widowed	Percentage share of people who are widowed	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
302		Divorced	Percentage share of people who are divorced	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
303		Registered partnership	Percentage share of people who are in a registered partnership	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

304		Registered partner deceased	Percentage share of people whose registered partner deceased	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
305		Registered partnership annulled	Percentage share of people who annulled their registered partnership	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
306		Without specification	Percentage share of persons for whom there is no information on marital status	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
307	Country of birth	Germany	Percentage share of persons born in Germany	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
308		EU27-country	Percentage share of persons born in EU27-country	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
309		Other Europe	Percentage share of persons born in another European country	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
310		Other world	Percentage share of persons born in a country outside Europe	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
311		Other	Percentage share of persons for whom there is no indication of country of birth	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
312	Sex	Male	Percentage share of persons who are male	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
313		Female	Percentage share of persons who are female	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
314	Religion	Roman Catholic Church (public)	Percentage share of people who are Roman Catholic	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
5		Protestant church (public)	Percentage share of people who are Protestant	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
6		Not specified	Percentage share of persons for whom there is no indication of religion	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
7	Citizenship	Germany	Percentage share of persons with German citizenship	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
8		EU27-country	Percentage share of persons with citizenship from a EU27-country	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
9		Other Europe	Percentage share of persons with citizenship from another European country	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
10		Other world	Percentage share of persons with citizenship from another country outside of Europe	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
11		Other	Percentage share of persons for whom there is no information on nationality	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
12	Citizenship by selected countries	Bosnia and Herzegovina	Percentage share of persons who are nationals of Bosnia and Herzegovina	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
13		Greece	Percentage share of persons who are nationals of Greece	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

14		Italy	Percentage share of persons who are nationals of Italy	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
15		Kazakhstan	Percentage share of persons who are nationals of Kazakhstan	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
16		Croatia	Percentage share of persons who are nationals of Croatia	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
17		Netherlands	Percentage share of persons who are nationals of the Netherlands	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
18		Austria	Percentage share of persons who are nationals of Austria	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
19		Poland	Percentage share of persons who are nationals of Poland	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
20		Romania	Percentage share of persons who are nationals of Romania	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
21		Russ. Federation	Percentage share of persons who are nationals of the Russ. Federation	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
22		Turkey	Percentage share of persons who are nationals of Turkey	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
23		Ukraine	Percentage share of persons who are nationals of Ukraine	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
24		Other	Percentage share of persons who are nationals of other countries	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
25	Number of citizenships	One citizenship	Percentage share persons with one citizenship	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
26		Several, German and foreign	Percentage share of persons with multiple citizenship, including German citizenship	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
27		Several, only foreign	Percentage share of persons with multiple but only foreign citizenships	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]
28		Not known	Percentage share of persons for whom the number of citizenships is not known	continuous	2011	zip area	Statistic departments of the federation and the federal states[6]

[1] All data is linked by zip area code. Since the geographical coverage of some zip code areas changes over time, we combine areas that are merged or split partly or completely within the observation period as one total area that remains constant over time.

[2] Households are categorized along several dimensions, including income, education level, type and number of owned cars, type of housing, and the share of of self-employed and business decision-makers in the neighborhood.

[3] We retrieved data from OpenStreetMap in September 2020. Data quality and coverage may vary across regions and variables.

[4] BBSR data are only available at the county level. For the zip area regions, we calculate a weighted mean from the county values where each county value enters with a weight proportional to the share of zip area residents living in the county.

[5] The variable is based on raster data in a 1-kilometer grid. More information: Gitterbasierte Auswertungen des Zensus 2011

[6] The variable is based on raster data in a 100-meter grid. More information: Gitterbasierte Auswertungen des Zensus 2011

Table 3: Horvitz-Thompson Transformation

	Hospitalizations per million insured persons
$\hat{\beta}_1$ (ATE)	29.4334***
	(3.9232)
$\hat{\beta}_2$ (HET)	36,096.2100***
	(7,946.9910)
Observations	506,966,676

The table reports the same effects as Table , however using the alternative estimation strategy based on the Horvitz-Thompson transformation (Chernozhukov et al. 2018). Standard errors are clustered at the ZIP code level and are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.

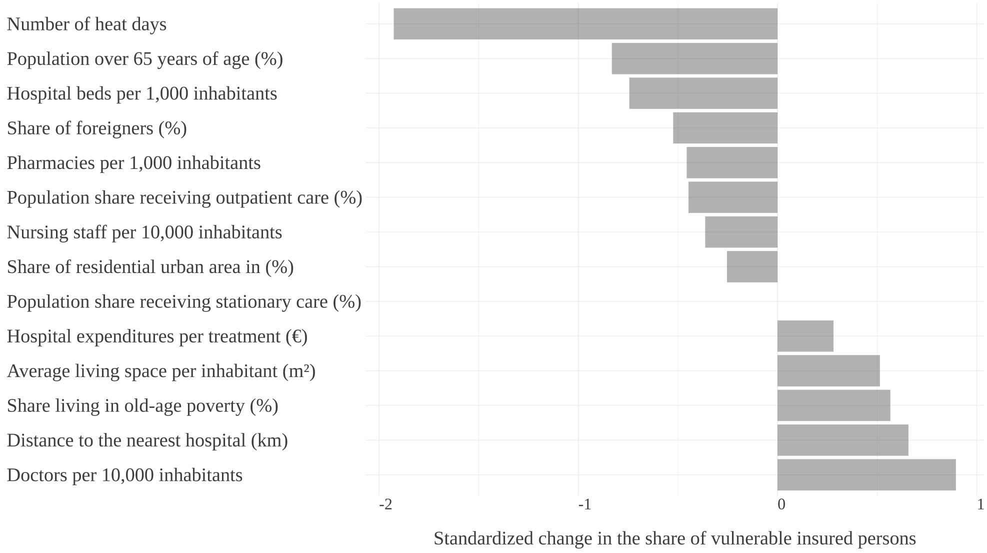
Table 4: Data aggregated to the ZIP code level

	Hospitalizations per million insured persons	
	(1)	*(2)*
$\hat{\beta}_1$ (ATE)	50.8254***	38.5783***
	(8.9604)	(6.7933)
$\hat{\beta}_2$ (HET)	121,870.7000***	93,253.1600***
	(23,040.7600)	(18,181.0700)
Horvitz-Thompson transformation	no	yes
Observations	1,411,347	1,411,347

The table reports the same effects as Table and Table , however with the data aggregated to the ZIP code level. Standard errors are clustered at the ZIP code level and are in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$.

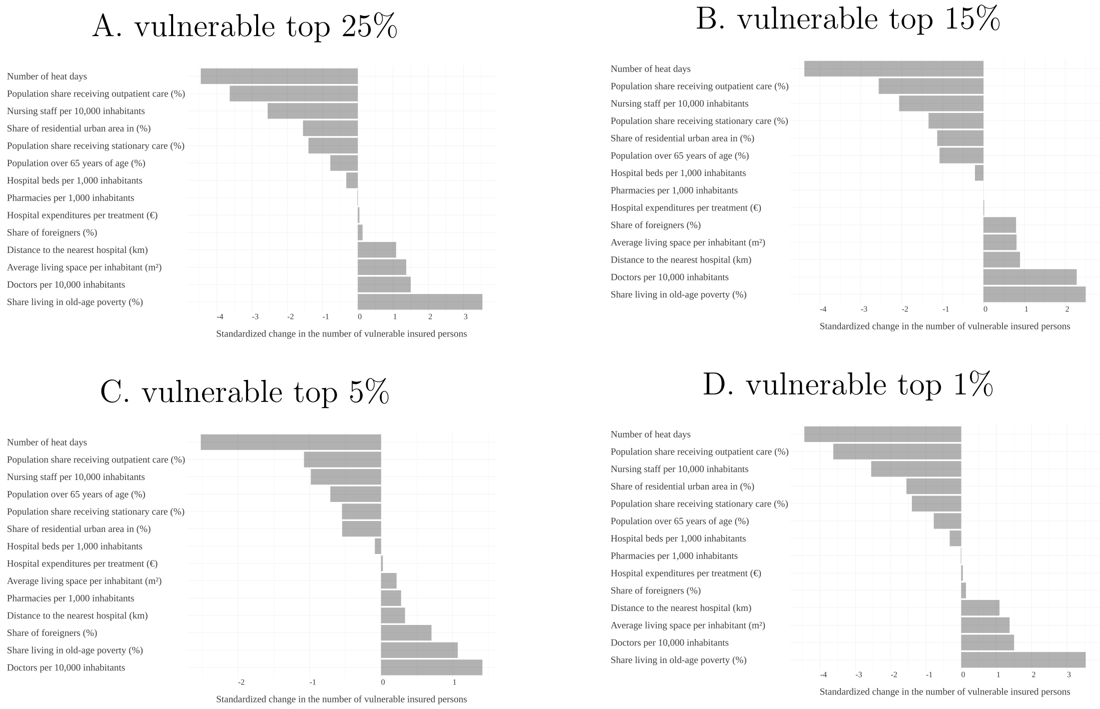
Appendix Figures

Figure 7: Risk factors that predict the share of the vulnerable elderly



The figure shows demographic, socioeconomic, and infrastructural characteristics of zip areas that predict the share of insured individuals with a heat vulnerability in the top 25% percentile. The bars represent the estimated coefficients from a LASSO regression which assigns zero-coefficients to irrelevant variables. Hospital expenditures refer only to the AOK-insured population of the zip area and the previous year. We exclude zip areas with fewer than 100 insured individuals from the regression. The sample size is 7,322.

Figure 8: Risk factors that predict the number of elderly with varying vulnerbaility



The figure shows demographic, socioeconomic, and infrastructural characteristics of zip areas that predict the number of insured individuals with a heat vulnerability in the top 25% (Panel A), the top 15% (Panel B), the top 5% (Panel C), and the top 1% (Panel D) percentile. The bars represent the estimated coefficients from a LASSO regression which assigns zero-coefficients to irrelevant variables. Hospital expenditures refer only to the AOK-insured population of the zip area and the previous year. We exclude zip areas with fewer than 100 insured individuals from the regression. The sample size is 7,322.