Population Health Effects and Health-Related Costs of Extreme Temperatures: Comprehensive Evidence from Germany

Martin Karlsson
CINCH, University of Duisburg-Essen[††]
Lund University

Nicolas R. Ziebarth
Cornell University[*]

May 14, 2018

Abstract

This study assesses the short and medium-term impact of extreme temperatures on population health and health-related costs in Germany. For 1999 to 2008, we link the universe of 170 million hospital admissions and all 8 million deaths with weather and pollution data at the day-county level. Extreme heat significantly and immediately increases hospitalizations and deaths. This finding holds irrespective of whether we employ econometric models that are standard in economics or models that are standard in epidemiology; we compare and discuss both approaches. We find evidence for partial "harvesting." At the end of a 30-day window, the immediate health effects are, on average, one quarter lower, but this reduction is primarily evident for cardiovascular and neoplastic diseases. Moreover, aggregating at the yearly level reduces the effect size by more than 90 percent. The health-related economic costs accumulate up to €5 million per 10 million population per hot day with maximum temperatures above 30 °C (86 °F).

Keywords: population health effects, extreme temperatures, hot day, cold day, weather, pollution, hospital admissions, mortality, climate change
JEL classification: I12, I18, Q54, Q58

[‡]We thank the German Federal Statistical Office (*Statistisches Bundesamt (destatis)*), the German Meteorological Service (*Deutscher Wetterdienst (DWD)*) and the German Federal Environmental Office (*Umweltbundesamt (UBA)*) that provided the data basis for the study as well as Joerg Blankenback for his great support in the interpolation of the geodesic data. In particular, we thank Evelyn Forget, Silviya Nikolova, Seiro Ito, and Reed Walker for outstanding discussions of this paper. Moreover, we thank Daniel Baumgarten, Antonio Bento, Damon Clark, John Cawley, Peter Eibich, Maria Fitzpatrick, Rick Geddes, Albrecht Glitz, Dan Grossman, Don Kenkel, Hyuncheol Kim, Ilyana Kuziemko, Michael Kvasnicka, Dan Lichter, Dean Lillard, Sean Lyons, Alan Mathios, Jordan Matsudaira, Vincent Pohl, Emily Owens, Sharon Sassler, Steve Stillmann, Hanna Wielandt, Robert Williams III, Will White, Martina Zweimüller and participants at the 2013 meeting of the American Economic Association (AEA), the 2nd Workshop on Energy Policy and Environmental Economics at Cornell, the 2013 Conference of the European Society for Population Economics (ESPE) in Aarhus, the 2013 UK Health Economists' Study Group (HESG) Meeting at Warwick, the 2013 Canadian Health Economists' Study Group (CHESG) Meeting in Winnipeg, the Economics of Disease Conference in Darmstadt 2013 as well as seminar participants of the Population Center (CPC) and the Institute on Health Economics, Health Behaviors and Disparities (IHEHBD) at Cornell University, the German Institute for Economic Research (DIW Berlin), and the Berlin Network of Labour Market Researchers (BeNA) for their helpful comments and discussions. We also thank Maike Schmitt (TU Darmstadt), Felix Heinemann (TU Darmstadt), Peter Eibich (DIW Berlin), Lauren Jones (former PAM PhD student, now OSU), and Katherine Wen (Cornell University) for excellent research assistance. All remaining errors or shortcomings of the article are our own. The research reported in this paper is not the result of a for-pay consulting relationship. Our employers do not have a financial interest in the topic of the paper which might constitute a conflict of interest. Funding from the Cornell Institute for Social Science (ISS) *Small Grant Program* as well as the Cornell Population Center (CPC) *Seed Grant Program* are gratefully acknowledged.

[*]Corresponding author: Nicolas R. Ziebarth, Cornell University, Department of Policy Analysis and Management (PAM), 106 Martha Van Rensselaer Hall, Ithaca, NY 14850, USA, phone: +1-(607)255-1180, fax: +1-(607)255-4071, e-mail: nrz2@cornell.edu

[††]University of Duisburg-Essen, Chair of Health Economics, Schützenbahn 70, 45117 Essen, Germany e-mail: martin.karlsson@uni-due.de

1 Introduction

Climate change is one of the great challenges of modern society. The Stern (2006) report states that the world's average temperature has risen by 0.7 °C (1.3 °F) over the past 100 years and projects that this trend will continue into the future. For the US, the predicted temperature increase ranges between 2 and 6 °C (4 and 11 °F) by the end of the 21st century (United States Global Change Research Program, 2009). Moreover, climate scientists project a significant increase in inclement weather conditions, such as the number of hot days with temperatures above 30 °C (86 °F) or the number of heat waves. More precisely, the Intergovernmental Panel on Climate Change (IPCC) projects: *"It is very likely that hot extremes, heat waves and heavy precipitation events will continue to become more frequent."* (IPCC (2007), p. 46, 53).

Studies in health economics as well as environmental epidemiology and bio-statistics empirically assess the impact of extreme temperatures on human health. A recent literature review article in economics summarizes: *"[...] with global temperatures expected to rise substantially over the next century, understanding these relationships [between climatic factors and economically relevant outcomes] is increasingly important for assessing the "damage function" that is central to estimating the potential economic implications of future climate change* (Dell et al. (2014), p. 2)." Some economic studies define a hot day as a day where the daily maximum temperature exceeds 30 °C.[1] whereas others use the threshold of 90 °F for the *mean* daily temperature.

Panel A of Table 1 lists and categorizes select economic studies on the impact of extreme temperatures on human health. Methodologically, these studies typically regress the mortality rate of (US) counties or states on temporal and spatial fixed effects, e.g., state and month-year fixed effects. Assuming that the remaining temperature variation is exogenous to humans, this approach allows one to identify the causal effects of extreme temperatures on mortality.

[Insert Table 1 about here]

As Panel B of Table 1 shows, the epidemiological and bio-statistical literature on this topic is older and richer. The majority of studies in this field are also based on US data but typically exploit daily mortality counts of cities over longer time horizons. Instead of employing parametric OLS fixed effects models using mortality or hospitalization rates as outcomes, these studies mostly employ log-linear poisson models of death or hospital counts and model seasonal effects as smooth spline functions. Basu and Samet (2002), Åström et al. (2011), Deschênes (2014) and Hondula et al. (2015) provide literature reviews on the relationship between heat events and human health.

[1] This is identical to the official definitions of the public meteorological services in the German speaking countries. (Deutscher Wetterdienst (DWD), 2017).

One unifying theme of both scientific fields is the examination of the so called "harvesting hypothesis." While it has been clearly documented that deaths and hospital admissions spike in the short run, there is no conclusive consensus on the medium- to long-run effects of extreme temperatures on human health. Several studies provide support of the harvesting hypothesis that mostly older people die during heat waves—that is, people in weak health who would have died in the near future, regardless of whether there was a heat wave (Braga et al., 2001, 2002; Hajat et al., 2005; Stafoggia et al., 2009; Deschênes and Moretti, 2009). If extreme temperatures cause premature death for older people by only a couple of weeks, then the overall population health effects (and predicted health losses due to climate change) are naturally overestimated in static models.

Other unifying research questions with inconclusive evidence surround the questions of (i) the negative impact of cold weather on human health (Braga et al., 2001; Anderson and Bell, 2002; Goldberg et al., 2011; Son et al., 2014; Gasparrini et al., 2015; Chung et al., 2015; Son et al., 2016; White, 2017), and (ii) whether and how fast humans are able to adapt to changes in extreme ambient temperatures (Braga et al., 2001; Anderson and Bell, 2002; Deschênes and Moretti, 2009; Deschênes and Greenstone, 2011; Bobb et al., 2014; Deschênes, 2014; Nordio et al., 2015; Barreca et al., 2016).

The main objective of this paper is to comprehensively assess the short- and medium-run population health effects of extreme temperatures, particularly heat events, for an entire industrialized nation over one decade. As illustrated in Table 1, the existing evidence is either based on mortality *or* hospital data and typically does not use both data sources jointly. Additionally, due to data limitations, most existing studies cannot exploit complete censuses of either deaths or hospitalizations. Rather, they typically rely on data from single cities or population sub-groups. This paper uses the universe of hospital admissions and deaths from 1999 to 2008 for Germany, the most populous European country and fourth largest industrialized nation in the world. We aggregate these data to the county level and merge them with 11 weather and 11 pollution measures on the daily county level. Weather and pollution measures were recorded daily by a network of over 2,300 governmental ambient monitors, distributed over an area the size of Montana. The comprehensiveness of our data allow us to draw a very complete picture of all severe health shocks that are triggered by extreme temperatures. However, we do not observe health shocks that do not lead to deaths or hospitalizations.

Second, we see this paper as an attempt to bridge the two different literature strands in health economics and epidemiology, which essentially empirically analyze the same research questions. It is very difficult, if not impossible, to compare results from studies across disciplines as they are

presented in very different ways; the metrics that would be needed to make a comparison with results from other approaches are typically not provided. In order to fill this gap, we carry out specifications that are closest to the standard specifications in each literature strand and compare the effects sizes of both modeling approaches. We also compare the effects sizes of our findings with those in the two literature strands.

Third, thanks to the richness of our data, we carry out extensive tests to check whether the heat-health effect size varies when we control for other weather and pollution-related confounding factors or "effect modifiers." The confounding impact of multiple pollutants is an important question of inquiry, particularly in the epidemiological literature (Katsouyanni et al., 2001; Dominici et al., 2010; Bobb et al., 2013; Deryugina et al., 2016).

Another main contribution of this paper is to investigate the harvesting hypothesis in detail. Whereas several definitions of the harvesting hypothesis exist, the definition used in this paper hypothesizes that a significant share of the negative short-term effects of heat events are not present in the medium-run because many hospitalized or dead people would have been hospitalized or dead even in the absence of the heat event shortly after. This definition of "harvesting" would predict short-term excess mortality and hospitalizations, followed by an under proportional development of mortality and hospitalizations. In other words, we contrast the short and the medium-run health effects of heat. To provide evidence for or against the validity of the harvesting hypothesis, we (a) investigate how mortality develops in the 30 days after a heat event. Because the relevance of the harvesting hypothesis may depend on the disease type, we (b) investigate these time trends separately for five disease categories. In addition, we (c) profile the age structure of those who are hospitalized or die during heat events, again separately by five disease categories. Finally, we (d) aggregate the data up to the monthly and annual level. By aggregating up, we solely exploit the remaining monthly and annual variation in heat events as a source of exogenous variation.

As a final contribution, we provide an assessment to better understand the health-related costs associated with extreme temperatures. The most concrete climate change prediction of the IPCC (2007) is an increase in the number of extreme heat events. Thus, we attempt to monetize the health losses associated with one additional hot day for an entire nation. Two factors drive the estimates: first, the choice to consider other climatic effect modifiers like air pollution or not and second, whether short- or medium-run effects are the basis of the calculation. We also decompose the total health costs into direct effects due to deaths and hospitalizations as well as indirect effects for lost labor and a loss of quality of life while being hospitalized. When applying standard values for a statistical life, the mortality effects are responsible for at least half of the total health costs.

Our findings show that extreme heat has a highly significant and large short-term impact on

both hospitalizations and deaths, whereas the results for extreme cold are less consistent. With regard to heat events, the standard empirical models in economics and epidemiology provide the same qualitative findings and are consistent, though the estimated effect on mortality is larger when we use standard approaches from economics. We are able to rule out that this difference is driven by differences in the definition of a hot day. Instead, it appears that the difference is attributable to unobserved heterogeneity between counties. When comprehensively considering other contemporaneous weather and pollution conditions—effect modifiers—the net impact of extreme temperature on health shrinks significantly in both modeling approaches, but the decrease is larger in the economic models.

At the population level, there is strong evidence for "harvesting", but only for heart and respiratory diseases. As expected given their medical peculiarities, infectious or metabolic diseases do not show the characteristic short-term spike followed by a decrease in hospitalizations or deaths.

We also find a clear age gradient in the hospitalization and mortality pattern of extreme heat: mostly older people are hospitalized or die during heat events. However, maybe surprisingly, we do not find such age gradients when considering *relative* increases in percent, which tend to be roughly similar for all age groups.

The last part of the paper estimates that—using the economics definition—one additional hot day with temperatures above 30 °C (86 °F) causes monetized health losses of between €750 thousand to €5 million per 10 million residents, depending on the underlying assumptions.

2 Datasets, Main Variables, and Identifying Variation

2.1 *Mortality Census*: The Universe of all German Deaths 1999-2008

The first dataset is the *Mortality Census* which is provided by the German Federal Statistical Office. The *Mortality Census* includes every death that occurred on German territory. Per year, one observes approximately 800,000 deaths, i.e, about 8 million deaths from 1999 to 2008. To obtain the working dataset, we aggregate the individual-level data at the day-county level and generate the mortality rate per 100,000 population.

Appendix A shows all raw measures included in the *Mortality Census*. It contains information on age, gender, day of death, county of residence as well as the primary cause of death in ICD-10 (10th revision of the *International Statistical Classification of Diseases and Related Health Problems*) classification.

Construction of Main Dependent Variables

Using information on the primary cause of death, we generate a series of dependent variables. To do so, we extract the letter and digits of the ICD-10 code, e.g., J00-J99 refers to "the respiratory system." In some cases, the second and third ICD-10 digits are helpful in identifying more specific conditions. In addition to the *all-cause mortality rate*, which is simply the sum of all deaths, we examine five specific subgroups: the (i) *cardiovascular mortality rate*, (ii) *respiratory mortality rate*, (iii) *infectious mortality rate*, (iv) *metabolic mortality rate*, and (v) *neoplastic mortality rate*.

The total daily mortality rate is 3 deaths per 100,000 population—1.4 or almost 50% of which are caused by cardiovascular health issues. The summary statistics of the all-cause and cause-specific mortality rates are displayed in Appendix A.

2.2 *Hospital Admission Census*: The Universe of all German Hospital Admissions 1999-2008

The second dataset is the *Hospital Admission Census*. Access is again provided by the German Federal Statistical Office. It contains data on all German hospital admissions from 1999 to 2008. Germany has about 82 million inhabitants and registers about 17 million hospital admissions per year. We observe every single hospital admission from 1999 to 2008, i.e., a total of more than 170 million hospitalizations.[2] To obtain our working dataset, we aggregate the individual-level data at the day-county level and normalize admissions per 100,000 people using official population counts (see Appendix E).

As seen in Appendix B, along with other admission characteristics, the *Hospital Admission Census* provides information on the age and gender of the patient, the day of admission, the length of stay, the county of residence as well as the primary ICD-10 diagnosis.

Construction of Main Dependent Variables

Analogous to the *Mortality Census*, using information on the primary diagnosis, we generate a series of dependent variables. Again, the dependent variables indicate different diagnoses, generated by extracting the letter and digits of the ICD-10 code. In addition to the *all-cause hospitalization rate*, we examine five subgroups: (i) *cardiovascular hospitalizations* (I00-I99), (ii) *respiratory hospitalizations* (J00-J99), (iii) *infectious hospitalizations* (A00-B99), (iv) *metabolic hospitalizations*

[2] By law, German hospitals are required to submit depersonalized information on every single hospital admission. This excludes military hospitals and hospitals in prisons. The 16 German states collect the information and the German Federal Statistical Office (*Statistische Ämter des Bundes und der Länder*) provides restricted data access for researchers.

(E00-E89), and (v) *neoplastic hospitalizations* (C00-D49).

For the last section, to assess total health care costs, we also exploit hospital information on deaths and the length of stay. For example, *cardiovascular death* identifies people who died after they were admitted to a hospital due to a cardiovascular disease. *Cardiovascular hospital days* includes the number of nights that a patient spent in a hospital after a cardiovascular admission.

After having summed over county-level daily admissions, as above, we normalize the dependent variables per 100,000 population using official population data at the year-county level (Federal Institute for Research on Building, Urban Affairs and Spatial Development, 2012). Appendix B displays the summary statistics of all dependent variables.[3] For example, on a given day, we observe 58 hospital admissions per 100,000 population. On average, a day triggers 489 hospital days, i.e., the 58 admissions have an average length of stay of 8.4 days. The largest single group of diseases is *cardiovascular hospitalizations*. Nine cardiovascular admissions per 100,000 population make up 16% of all admissions.

2.3 Official Daily Weather Data from 1,044 stations 1999-2008

The weather data are provided by the German Meteorological Service (*Deutscher Wetterdienst (DWD)*), a publicly funded federal institution. Weather measures were collected from 1999 to 2008 from up to 1,044 meteorological monitors which were distributed all over Germany. Figure 1 shows the distribution of all ambient monitors along with county borders. Figure 2a shows a boxplot of the mean temperature over the twelve months of the year (averaged over all ten years). The graph illustrates the large cross-county as well as cross-seasonal variation in temperatures. One observes a clear increase in average temperatures during the summer months. Figure 2b shows the daily cross-county temperature variation over all ten years. One observes the typical seasonal trends along with many spikes in the high-frequency data. The empirical models will exploit the rich positive and negative weather shocks across space and over time.

[Insert Figure 1 and 2 about here]

The paper uses official data from all existing weather stations in a given year. As described in Appendix D1, we interpolate the point measures into county space on a daily basis using Inverse Distance Weighting (IDW).

[3] Note that the German data protection laws prohibit us from reporting min. and max. values.

Construction of Extreme Temperature Indicators & Identifying Variation

Indicators in Economics. In our main specification, we employ semiparametric variants of a standard "economic model" to net out seasonal and geographic effects and let a series of temperature regressors float flexibly while allowing for precise enough estimates. However, we also employ threshold models to measure extreme heat and cold for the following reasons:

(i) Although there exist no international definitions of hot or cold days, our threshold measures follow the official definitions by the German Meteorological Service. They define a *Hot Day* as a day with a maximum temperature above 30 $^\circ$C (86 $^\circ$F).

(ii) The economics literature has employed threshold indicators which facilitates a comparison of results (cf. Deschênes and Moretti, 2009; Barreca et al., 2016).[4]

(iii) Defining a binary indicator to measure *Hot* and *Cold Days* simplifies the empirical analysis, provides the reader with a better intuition, and makes it easier to follow the thought experiment wherein we ask, "What are the health effects of one additional *Hot Day*?"

(iv) As we will demonstrate in the Results section, there is empirical evidence that most adverse health effects kick in when temperatures exceed 30 $^\circ$C (86 $^\circ$F). Thus we define the first pair of binary measures, following the economics literature, as:

- **Hot Day** = 1 if the max. temperature >30 $^\circ$C (86 $^\circ$F), 0 else.
- **Cold Day**= 1 if the min. temperature < -10 $^\circ$C (14 $^\circ$F), 0 else.

Indicators in Epidemiology. The second set of extreme heat and cold indicators follows the standard definitions in epidemiology. Here, the binary measure for *Hot Day II* equals 1 when the average daily temperature exceeds percentile 97.5 of the county-level temperature distribution over all years.[5] The definition for *Cold Day II* is analogous but refers to percentile 2.5 of the county-level temperatures.

Panel B of Table C1 in Appendix C shows the descriptive statistics for the two sets of extreme temperature indicators. Let us start with heat events and the epidemiological definition. As expected, 2.5% of all county-day observations are hot days. The maximum daily temperature during a *Hot Day II* is on average 31.2 $^\circ$C (88.1 $^\circ$F) but varies between 23 and 39 $^\circ$C (73 $^\circ$F and 102 $^\circ$F).

In contrast, according to the standard definition in economics, 2% of all days are *Hot Days* with maximum temperatures above 30 $^\circ$C (86 $^\circ$F). The maximum daily temperature during a *Hot Day*

[4]To be precise, the US studies by Deschênes and Moretti (2009) and Barreca et al. (2016) define a *Hot Day* as a day with the *mean* temperature >90 $^\circ$F (32 $^\circ$C).

[5]Because no official definition exists, many epidemiological papers provide robustness checks varying the thresholds between the 95th and 99th percentile.

is on average 31.9 $^\circ$C (89.4 $^\circ$F) and varies between 30 and 39 $^\circ$C (86 $^\circ$F and 102 $^\circ$F). This means that the epidemiological definition leads to lower average maximum temperatures on hot days, but also a longer left temperature distribution tail with temperatures below 30 $^\circ$C (86 $^\circ$F). The two indicators are identical in 99.2 percent of all days.

However, the definition in economics results in a quarter fewer hot days over our sample period. On average, it translates into seven *Hot Days* per year but this figure varies between 4 (1999, 2004, 2007) and 18 (2003) over the years. The variation in *Hot Days* between counties is even larger and varies between 0 and 40 per year (Figure 3b). In our empirical harvesting tests, we aggregate the data at the month-county and year-county level and exploit this variation in the monthly and annual number of *Hot Days*.

[Insert Figure 3 about here]

Figure 3 plots the distributions of *(i)* the maximum daily county-level temperatures, and *(ii)* the annual number of *Hot Days* per county according to the definition in economics. The figure shows that the annual maximum temperatures follow a normal distribution with the mass point around 14 $^\circ$C (57 $^\circ$F). Moreover, the annual number of *Hot Days* is skewed to the right and exhibits substantial variation with many counties showing more than 10 *Hot Days* per year. Figure 3 illustrates that the identifying variation stems from the majority of counties and not just a small subset of "hot" counties. Thus extrapolation and out-of-sample predictions are largely avoided.

Turning to cold events, the epidemiological *Cold Day II* definition is symmetric to the one for *Hot Day II*, which is why we observe 2.5% of all county-day observations as cold days (Panel B of Table C1). The average minimum temperature during a *Cold Day II* is -10 $^\circ$C (14 $^\circ$F) and varies between -3.5 and -25 $^\circ$C (25.7 $^\circ$F and -13 $^\circ$F).

According to the definition in economics, slightly over 1% or about 20,000 of all county-day observations are *Cold Days* with minimum temperatures below -10 $^\circ$C (14 $^\circ$F). Similar to the hot day case, the average minimum temperature on cold days is lower according to the economics definition (-12.5 $^\circ$C or 9.5 $^\circ$F); and the economics definition results in a truncated temperature distribution with minimum temperatures strictly below the threshold of -10 $^\circ$C (14 $^\circ$F). This translates into about 4.5 *Cold Days* per year, ranging from an average of 0.4 *Cold Day* in 2008 to 9 *Cold Days* in 2003. The annual county-level variation in *Cold Days* lies between 0 and 41 (see Figure 3b).

2.4 Official Daily Pollution Data from 1,314 stations 1999-2008

In extended models, we use daily pollution data from five different pollutants (NO_2, SO_2, NO_10, O_3, CO) to study effect size variation when controlling for pollution. The epidemiological literature

discusses extensively relevant aspects of controlling for multiple pollutants and their roles as effect modifiers (Katsouyanni et al., 2001; Dominici et al., 2010; Bobb et al., 2013). The pollution data are provided by the German Federal Environmental Office (*Umweltbundesamt (UBA)*), a publicly funded federal agency. From 1999 to 2008, pollution measures are collected from up to 1,314 ambient monitors (Figure 1). As with the weather measures and as described in Section F1, we interpolate the monitor point measures into the county space. Table D1 in Appendix D shows all raw pollution measures on a daily county-level basis.

Appendix D describes and graphically illustrates the tempo-spatial variation of the pollutants and their association with weather conditions: All pollutants have in common that they *(i)* exhibit some seasonal pattern, *(ii)* exhibit strong (non-)linear associations with the weather indicators—in particular the temperature.

3 Empirical Approaches and Identification

3.1 Empirical Approach in Health Economics

One standard approach in the economics literature would estimate the following model by OLS:

$$Y_{cd} = \alpha + \sum_{h\in\{<10,10-20\ldots>80\}} \beta_h \, MeanTemp^h_{cd} + \sum_{j=2}^{468} \nu_j \, county_j + \sum_{k=2}^{52} \zeta_k \, Week_k \\ + \sum_{m=\text{Feb1999}}^{\text{Dec 2008}} \sigma_m \, month_m + \theta \, X_{ct} + \sum_{i=1}^{30} \sum_{h\in\{<10,\ldots,>80\}} \gamma_{hi} \, MeanTemp^h_{c,d-i} + \epsilon_{cd} \tag{1}$$

where, depending on the specification, Y_{cd} would either denote the mortality rate or the hospital admission rate per 100,000 population in county c on day d. $MeanTemp^h_{cd}$ are a series of temperature regressors that equal 1 if the average daily temperature in the county falls into a bin of 10 $^\circ$ F and equal zero otherwise. To make our findings comparable to existing studies (mostly from the US), we employ eight temperature dummies (<10, 10-20, 20-30, 30-40, 50-60, 60-70,70-80, above 80 $^\circ$ F) and evaluate their health impact relative to temperatures between 40-50 $^\circ$ F (4.4-10 $^\circ$ C). The temperature coefficients then semiparametrically describe the temperature-health relationship, net of seasonal influences. We call this the "temperature bin model" (as compared to the "threshold model" which solely uses the *Hot Day* and *Cold Day* dummies described in Section 2).

In some specifications, in order to account for the serial correlation in temperature, we add 30 lags of $MeanTemp^h_{cd}$. Moreover, we routinely net out county fixed effects, $\sum_{j=2}^{468} \nu_j \, county_j$, week fixed effects, $\sum_{k=2}^{52} \zeta_k \, Week_k$, and year-month fixed effects, $\sum_{m=\text{Feb 1999}}^{\text{Dec 2008}} \sigma_m \, month_m$, in order to

adjust for trends and permanent differences in health across counties. In our main specification, we thus assume that regional differences in seasonality are captured by the county fixed effects (as well as the county-year controls described below). Given the relatively small variation in climatic conditions across counties in Germany, this assumption is likely to hold.

We also include a set of county-level covariates, X_{ct}. This vector contains the share of private hospitals, the bed density and the county-level GDP per capita (see Appendix E). We routinely cluster standard errors at the county level, but show that clustering at the state level or two-way clustering at the county and day level does not affect the main results. All econometric models are weighted by the total county population in a given year.[6]

Our baseline specification does not consider any other contemporaneous weather and pollution conditions. In addition to the fixed effects, only the eight temperature bins are added to the model. One can think of this approach as a reduced form "intention-to-treat" approach where the main regressor of interest absorbs all weather and pollution conditions that are correlated with the exogenous weather or pollution indicator.

In subsequent specifications, by contrast, we progressively add more variables that control for various environmental factors. These models include—in addition to the temperature bins—eight continuous weather (W_{cd}) measures (humidity, precipitation, cloud coverage, wind speed, storm force, air and vapor pressure, hours of sunshine) as well as five continuous pollution (P_{cd}) measures (NO_2, SO_2, NO_10, O_3, CO) in addition to their own and cross interactions (see also Tables C1 and D1).

3.2 Empirical Approach in Environmental Epidemiology

The standard approach in environmental epidemiology would estimate a log-linear hierarchical random-coefficient poisson model of the form:

$$\begin{aligned} log(\mathbb{E}\left[Y_{cd} \mid \cdot\right]) = \alpha_c + \beta_c\, HotDay_{cd} + \sum_{i=1}^{30} \gamma_{c,i}\, HotDay_{cd,d-i} \\ + \sum_{k=2}^{7} \zeta_k\, DayWeek_k \; + \theta\, X_{ct} + ns(time_c) \\ + ns(weather_c) + ns(pollution_c) \end{aligned} \tag{2}$$

[6]We weight by total pfor two reasons. First, it makes our estimates representative of the entire German population. Second, when relying on local averages, weighted least squares are more efficient than OLS. The second point holds only in the absence of common shocks at the local level; however, our clustered standard errors consider common group errors (cf. Solon et al., 2015).

where the variable definitions are similar to above. Importantly, instead of including binary fixed effects for each county and month-year, this empirical specification includes (typically cubic) splines of the date, $ns(time_c)$, to capture long-run trends and seasonal pattern. α_c is a set of county-specific intercepts.

Analogous to above, the baseline model would not consider other contemporaneous climatic confounders in addition to the *Hot Day II* and *Cold Day II* indicators. An "effect modifier model", by contrast, would additionally consider a series of separate cubic splines for weather and pollution measures $ns(weather_c)$, and $ns(pollution_c)$. Specifically, these models include separate smooth functions for eight additional weather and five pollution measures (Tables C1 and D1).

3.3 Comparison

One main purpose of this paper is to compare and contrast empirical approaches in health economics and epidemiology. Such a comparison is complicated by the fact that the two literatures (a) tend to define the outcome differently, (b) use different specifications for the main independent variables, and (c) allow for different lag structures in the regressions. Concerning (a), epidemiological models are often based on count data techniques, whereas contributions in economics tend to use mortality or hospitalization rates as outcomes. Effect sizes can nevertheless be compared by reporting relative changes (in percent) of the outcome variables. Concerning (c), we compare different specifications that include different numbers of lags.

Concerning (b), to compare estimates based on binary and more flexible representations of extreme weather events, we translate estimates from flexible specifications with temperature bins into binary impact measures using a linear index of the point estimates and their variances. In particular, we derive the *Hot Day* effect from equation (1) as follows:

$$\begin{aligned} \hat{\beta}_{HotDay} &= \left(\Pi_{HotDay} - \Pi_{NonHotDay}\right)' \hat{\beta} \\ \widehat{\mathrm{Var}}\left(\hat{\beta}_{HotDay}\right) &= \left(\Pi_{HotDay} - \Pi_{NonHotDay}\right)' \widehat{\mathrm{Var}}\left(\hat{\beta}\right) \left(\Pi_{HotDay} - \Pi_{NonHotDay}\right) \end{aligned} \tag{3}$$

where Π_{HotDay} represents the histogram of $MeanTemp_{cd}$ on a hot day, and $\Pi_{NonHotDay}$ the histogram on all other days. Our estimates for $\hat{\beta}_{ColdDay}$ and $\widehat{\mathrm{Var}}\left(\hat{\beta}_{ColdDay}\right)$ are defined analogously. As such, we can compare the changes in mortality attributed to hot and cold days according to different estimation methods.

3.4 Identification

One appealing aspect of using extreme temperature variation to estimate its impact on health is that changes in temperatures are very likely orthogonal to the error term in equations (1) and (2). It is very plausible that short-term weather variation is exogenous to the outcomes of any one individual (cf. Angrist et al., 2000). As Dell et al. (2014) put it: *"By harnessing exogenous variation over time within a given spatial unit, these studies [a growing body of research applying panel methods to examine how climate influences economic outcomes] help credibly identify (i) the breadth of channels linking weather and the economy [...] (p. 1)."* Remember that the econometric models net out a rich array of geographic, seasonal, and time effects and rely on high-frequency within county variation. Positive and negative temperature shocks are then linked to contemporaneous health effects at the day-county level. As Table 1 shows, this econometric approach to identification is carried out by the large majority of the leading and published temperature-health studies.

One could still list the following three identification concerns: (i) based on (un)observables, people may self-select into living in specific regions and (ii) individual-level exposure to weather and pollution conditions is unknown and (iii) adaption behavior may bias the "true" causal effect downward.

With respect to (i): One particular strength of our approach is its reliance on the universe of all deaths and hospital admissions over one decade, from the fourth largest industrialized nation in the world. To the extent that one is interested in the real-world effects of heat on population health in a given geographic area, we take the view that one should consider and *include* sorting into regions; the identified parameters then represent the effects on population health once geographic preferences are accounted for. In the case of Germany, it should be added that (intergenerational) geographic mobility is historically very low. Using the SOEP we find that, in a given year, only about 1% of all SOEP respondents move, which also includes within-county moving (Wagner et al., 2007; SOEP, 2012).

With respect to (ii) and adaptation behavior: Similar to above, we argue that we intentionally want to estimate an "intention-to-treat (ITT)" population health effect, including avoidance behavior and human adaptation to extreme temperatures. This parameter is arguably a relevant parameter for policymakers. The relationship that this paper intends to expose is: Given that humans have the capacity to adjust to extreme temperatures, based on current real-world behavioral data, how would climate change in the form of more heat events most likely affect population health? Without question, this ITT estimate represents a lower bound estimate as compared to a "full exposure" estimate keeping adaptation behavior constant. On the other hand, it can be

expected that adaptation behavior will further increase if extreme temperatures become more frequent in the future. To the extent that adaptation increases, our estimates will instead represent upper bound predictions of the effects of future heat waves. Janke (2014) uses English data and air pollution alerts to show that avoidance behavior exists for asthma but that the lower bound ITT estimates do not statistically differ from estimates modeling avoidance behavior.

Please note that it is beyond the scope of this paper to make projections about human behavioral adaptation and/or technological progress that could facilitate adaption behavior in the future. Such projections are inherently uncertain and notoriously difficult to make. However, recent empirical evidence shows that humans can adapt to adverse climatic conditions and that adaptation has increased over time (cf. Zivin and Neidell, 2013; Deschênes and Greenstone, 2011; Deschênes, 2014; Barreca et al., 2016). Given this recent empirical evidence, an approach that assumes no further adaptation behavior will lead to upper-bound estimates of the potential adverse health effects of climate change. They can thus be interpreted as "business-as-usual" scenarios which are useful to assess the willingness to pay to avoid these consequences.

Finally, the setup of the German health care system is particularly well-suited for our research objective because institutional and geographic access barriers to hospitals are very low. Germany has one of the highest densities of hospital beds worldwide, universal health care coverage, and virtually no access barriers for inpatient care (cf. OECD, 2017). German counties are comparable to US counties but less heterogeneous in terms of population density and area size. Germany's climatic conditions are ideal to empirically study and identify the instantaneous effects of extreme temperatures. Like most countries in the North Temperature Zone, Germany has four seasons, hot summers and cold winters (Figure 2). For example, during the 10 years that we study, daily maximum temperatures range from -14 $^\circ$C (7 $^\circ$F) to 39 $^\circ$C (102 $^\circ$F). As Figures 3 demonstrates, the identification of parameters is based on a broad set of counties and largely avoids out-of-sample predictions. All German counties experienced rich variation in extreme temperature.

4 Short-Term Effects of Heat and Cold on Population Health

4.1 Population Health Effects of Heat and Cold Events

Economic Models. Table 2 shows the results for the standard models in the (health) economics literature as formalized in equation (1). Panel A displays the effects on mortality and Panel B displays the effects on hospitalizations. Each column in each panel represents one model, estimated by OLS. The dependent variable always measures the all-cause mortality or hospitalization rate and does not distinguish by diagnoses. The models in columns (1) and (2) do *not* control for any

contemporaneous weather or pollution conditions. In contrast, column (3) considers five continuous pollutants at the daily level. Column (4) adds several *other* (than temperature) weather conditions such as sunshine or precipitation (the list of variables is in Panel A of Table C1 and Table D1). Column (5) additionally adds 30 lags of *Hot Day* and *Cold Day*. In the lower part of each panel, we report the estimated overall effect of hot and cold days using the approach of Section 3.3. Whereas the first five columns present the results of the temperature bin model based on the mean daily temperature, columns (6) and (7) present the results for the threshold model with *Hot Day* and *Cold Day* as main regressors.

[Insert Table 2 about here]

For mortality in Panel A, the impact of the mean temperature is roughly flat over a wide range of temperature bins: from the lowest measured temperatures up to around 60 $^\circ$F, the point estimates vary within a range of only 0.07, which corresponds to 2.5% of the baseline risk. From that point onward, the impact increases rapidly in each bracket. For the highest mean temperature bracket, 80 $^\circ$F to 90 $^\circ$F, the point estimate at 0.96 corresponds to 33% of the baseline risk. Controlling for pollution levels reduces the temperature gradient somewhat (column (2) vs. column (3), Panel A); including lags of extreme temperatures further reduces it (column (4) vs. column (5), Panel A). However, the largest drop in the temperature-health relationship is observed when we add other contemporaneous weather indicators in column (4).[7]

For hospitalizations in Panel B, the temperature-health relationship is non-monotonic: very cold mean temperatures (below 10 $^\circ$F) are associated an increase in admissions. Moderately cold weather (10-40 $^\circ$F), on the other hand, is associated with a decrease in admissions relative to the reference bin of 40-50 $^\circ$F. For temperatures of between 50-80 $^\circ$F, the relationship between temperature and admissions is positive for the parsimonious model without further weather controls and lags (columns (1) to (3), Panel B). However, as above, adding contemporaneous weather controls reduces the coefficient sizes substantially and turn them insignificant (or even negative for the 60 to 70 $^\circ$F bin). For the highest temperature bin (80-90 $^\circ$F), as above for mortality, we find a robust and highly significant positive gradient between temperatures and hospital admissions. It is noteworthy that adding contemporaneous weather controls (other than temperature) reduces the size of the health-admissions gradient in this highest temperature bin by about half (columns (4) and (5), Panel B).

[7]Because we are conducting multiple tests for multiple endpoints, we checked the robustness of our significance tests by using the procedure suggested by Benjamini and Hochberg (1995). For all conventional false discovery rates, the conclusions regarding significance and insignificance for all of the parameters remain the same.

When we use equation (3) to calculate the total impact of extreme temperatures, our estimates for heat and mortality (Panel A) corresponds to an impact of 0.35 or 12% of the baseline risk; in our most saturated specification, this effect size is reduced to 3.5%. When comparing this effect size to the effect size for the threshold model using solely a *Hot Day* dummy (columns (6) and (7)), the results are very similar. The heat-health estimates for hospital admissions (Panel B) yield effect sizes between 0.7% (column (5)) and 5.9% (column (1)) for the approach in equation (3). Again, the results for the threshold model are very similar.

With respect to cold and mortality, the first three columns of Table 2 suggest that cold is associated with a significant reduction in mortality by around two percent. However, this cold weather effect is not stable; once we control for other weather variables, cold temperatures are positively associated with mortality. In particular vapor pressure, wind speed and cloud coverage seem to be responsible for the association. For hospital admissions, we observe exactly the same pattern, which has also been reported by other studies (Schwartz et al., 2004). White (2017) argues that the decrease in admissions on cold days is likely driven by behavioral factors, such as a decreased willingness to seek treatments on cold days.

Comparison with Literature. As shown in Table 1, Deschênes and Moretti (2009) estimate a similar model at the daily level for the US and the years 1972 to 1988. They find that *average* daily temperatures of more than 80 $^\circ$F (26.7 $^\circ$C) increase the death rate by between 4 and 5%. Thus the results are broadly consistent despite Germany's stronger heat-health relationship. Explanations for the different effect sizes could be the different time periods or differences in geography and climate zones. An alternative explanation could be the much lower diffusion of air conditioners in Germany compared to the US (cf. Barreca et al., 2016).

Epidemiological Models. Table 3 shows the results when estimating models that are standard in Epidemiology, see equation (2). Instead of estimating OLS models with a rich set of temporal and spatial fixed effects, these are random effect poisson models with cubic splines and estimate incidence rate ratios. The effect can then be directly interpreted in percentages. Otherwise, the structures of the tables are similar: Panel A shows the effects on mortality and Panel B the effects on hospitalizations. Each column in each panel represents one model. The models in columns (1) to (2) and (5) to (6) do not control for contemporaneous weather and pollution conditions, whereas columns (3) to (4) and (7) to (8) consider cubic splines of five pollutants and other weather conditions, such as the hours of sunshine or the precipitation level. In columns (4) and (8) we also control for 30 lags of the main independent variables *Hot Day* (*Hot Day II*) and *Cold Day* (*Cold Day II*).

[**Insert Table 3 about here**]

Starting with the heat-health relationship, the epidemiological models find that a hot day increases deaths by about 7% and admissions between 1.2 and 1.4%. These results hold whether we define hot days in terms of temperatures or in terms of percentiles. On the other hand, as above, controlling for effect modifiers such as pollutants reduces the effect sizes. In particular, the inclusion of lags reduces the estimated impact of a hot day to around 4%.

For the cold-health relationship, the least restrictive specifications in columns (1) and (5) suggests an increase in mortality by 8 to 9% on cold days. However, most of this effect dissipates once we control for pollution splines, and vanishes almost completely once we control for lags of the extreme temperature variables. For hospital admissions, a cold day is associated with a significant reduction of between 1 and 3%. This effect becomes more pronounced when controlling for pollution splines.[8]

Comparison with Economic Models. Concerning the heat-mortality relationship, the results are similar to those delivered by economic models, even though the effect is roughly one-third smaller in the least restrictive model. This difference is not attributable to differences in the definition of a hot day, as the comparison in Table 3 shows. The main difference between equations (1) and (2), apart from functional form assumptions, is the absence of county fixed effects in the epidemiological specification. Hence the (small) difference in effect sizes could be driven by adaptation of the human body or endogenous sorting of frail individuals into counties where heat events are rare.

Concerning the cold-mortality relationship, the two approaches are in stark contrast with each other. According to economic models, a cold day is associated with an average reduction in mortality by 2%, whereas the epidemiological models suggest an increase in mortality ranging from 0.6% to 10%. The apparent reason for this discrepancy is that the temperature bin model fails to recover the effect of extreme cold on health. When we use a dummy for extreme cold in the economic model (columns (6)-(7) of Table 2), the estimated effect is actually positive, albeit relatively small (0.7 to 0.9% of the baseline risk)—smaller than most of the epidemiological effect sizes—but of the same magnitude as the most restrictive specifications in Table 3.

For hospitalizations in Panel B of Tables 2 and 3, the baseline models are very consistent for

[8]Note that clustering standard errors in these models using high frequency data and millions of observations is computational very challenging. Many studies in epidemiology provide unclustered standard errors and/or do not discuss clustering. We follow this example for illustrative purposes. However, we strongly advise to interpret the displayed, non-clustered, standard errors with caution. Thanks to the rich databases and as shown for the economic models, clustering at the county or county-day level would inflate standard errors substantially but the coefficient estimates very likely would remain statistically significant at conventional levels.

cold, but the heat effects differ. Again, we can rule out that the difference is driven by the different definitions of a hot day in economics vs. epidemiology. Besides, given that the estimates from the threshold models in columns (6) and (7) are very consistent with those from the temperature bin models in columns (1) to (5) of Table 2, we can also rule out that irregularities in the temperature gradient are responsible for the difference. Instead, as for mortality, unobserved heterogeneity between counties appears to be responsible for the differences.

Comparison with Literature. Table 1 lists several epidemiological studies on the relationship between heat and mortality; almost all of them exploit data from specific geographic regions, mostly cities, but not entire countries. Differences in effects sizes could be due to (i) differences in the underlying population, (ii) differences in the time periods studied, or (iii) differences in the temperature distributions across regions. Nordio et al. (2015), for example, use daily mortality data for US cities—a total of 211 cities over five decades. They categorize cities by eight clusters, depending on geography, and find that average daily temperatures above 30 $^{\circ}$C increase the relative death risk by more than 10%, consistent with our findings. Son et al. (2016) report a 6% higher mortality risk in São Paulo (Brazil) for the 99th as compared to the 90th temperature percentile. Chung et al. (2015) use data from the 1990s and 2000s for three cities in Taiwan and six cities each in Korea and Japan to identify heat-mortality effects of between 5 and 10%. They also identify cold-mortality effects which were, however, much smaller in size.

We find the following for hospital admissions: Bobb et al. (2014) use admission data of Medicare enrollees for 1,943 US counties and the years 1999 to 2010. They find elevated risks of between 5 and 18% for five disease groups but non-significant effects for the majority of the 214 disease groups investigated. Son et al. (2014) investigate the relationship between heat and cold events and hospital admissions in eight cities in South Korea. They show for Seoul that admissions for heart diseases increase almost linearly from zero to ten percent when temperatures increase from 20 to 30 $^{\circ}$C.

4.2 Robustness Checks

Table 4 presents a series of robustness checks. The reference specification is always the effect of one *Hot Day* on the hospital admission rate as in column (6) of Table 2, Panel B. All findings also hold when using the mortality rate as the dependent variable (results available upon request).

Column (1) in Panel A reports results with standard errors clustered at the state instead of the county level; Column (2) applies two-way clustering by county and date (Cameron and Miller, 2011) and column (3) corrects for spatial dependence (Driscoll and Kraay, 1998). As compared to the standard specification, standard errors increase, but the coefficient estimates remain highly

significant at the one percent level.

The next three columns add nation-level (column (4)), state-level (column (5)) as well as county-level (column (6)) time trends to the model. The latter two specifications slightly reduce the magnitude of the estimated *Hot Day* coefficients. Column (7) in Panel A adds county-by-year fixed effects. Again, the coefficients are very robust in size and significance.[9]

[Insert Table 4 about here]

Column (1) in Panel B serves as comparison for the fully saturated model and excludes the year 2000 for which no PM_{10} data are available. Excluding this year does not alter the results. Column (2) simply takes the logarithm of the dependent variable, which is an alternative way of modeling the heat-health relationship.

Column (3) in Panel B shows yet another way of modeling the heat-health relationship. Here, the model includes the maximum daily temperature as a continuous variable, along with the *Hot Day* dummy and an interaction between *Hot Day* and a continuous variable which captures the difference between the average maximum temperature that prevails on *Hot Days*, 32 $^\circ$ C (89 $^\circ$ F), and the county-specific maximum temperature on a given *Hot Day*. In other words: The interaction term indicates the degree to which hospitalizations increase with every temperature degree increase above 32 $^\circ$ C (89 $^\circ$ F). An average *Hot Day* increases admissions by about 1.6 per 100,000 population or 3%. As shown by the interaction term, the effect of a *Hot Day* increases by 1 percentage point for every degree Celsius above 32 $^\circ$ C (89 $^\circ$ F).

Column (6) in Panel B exclude all counties in years when no active weather monitor existed within 60 km (37.5 miles) of the county centroid. In addition to the extensive discussion in Appendix F, this serves as an additional robustness check for whether measurement errors could potentially affect the findings. As seen, this is not the case.[10]

4.3 Testing for Adaptation Behavior

Columns (4) and (5) of Panel B interacts *Hot Day* with a dummy for hot and cold regions. A similar approach has been used by Deschênes and Greenstone (2011) and Barreca et al. (2015) These specifications indirectly test adaptation behavior: Namely, whether the human body adapts to heat and warmer temperatures when humans live in warmer versus colder regions. We define a

[9] The standard estimate for the years 2006 to 2008 is 2.18***. We have to restrict these specifications to three years due to computer memory constraints. The results are robust when restricting the sample to the years 2003-2005 or 2000-2002.

[10] In an additional robustness check, we restricted our sample to 100 counties that contained the same monitors throughout the period. The results, which are available upon request, are also very similar to our baseline results.

"warm region" as a region where the mean annual county-level temperature falls into the highest temperature mean quartile for Germany (>10 °C, 50 °F). A "cold region" is defined as a county with a mean annual temperature in the lowest temperature quartile (<9 °C, 48 °F). Accordingly, we define two dummy variables, *Cold* and *Warm Region*, and add them to the models in levels and in interactions with the *Hot Day* indicator.

As in Barreca et al. (2015), there is clear evidence in support of the human body adaptation hypothesis because, in warm regions, the effect of a *Hot Day* is 20% smaller (corresponding to 0.6 fewer admissions per 100,000) than the average *Hot Day* effect. Likewise, in cold regions, the effect of a *Hot Day* is 15% larger. Note that this finding would also be in line with heat-(in)sensitive individuals sorting into colder (warmer) regions; however, geographic mobility in Germany is traditionally very low. Whatever the exact mechanism, it is unlikely to alter our main findings; in warm regions, *Hot Days* still lead to 4% more hospital admissions and in cold regions, they lead to 6% more admissions.

4.4 Effect Heterogeneity by Age

It is well-known that extreme temperatures particularly affect frail members of society, for example, older people. This has implications for the relevance of "harvesting", which we will address in the following section. To assess the age gradient of the heat-health relationship, this section shows estimates for age-specific mortality and hospitalization rates, which we also stratify by diagnoses.

Figure 4 shows the results for mortality. Figure 4a displays estimates for all-cause mortality by age group. The line connecting circles shows the *absolute* effect sizes per 100,000 population (with 95 percent confidence intervals) and the triangles show the *relative* effect sizes in percent for each of the 17 age groups. The absolute effect is relatively flat for most age groups (even though sometimes statistically significant thanks to the large sample size) and increases progressively from age 50 onward—from less than 0.2 deaths per 100,000 to almost 3 deaths per 100,000 for people above 75 years. However, the relative increase in percent for each age group fluctuates around 10% for all age groups above 30 (and is quite irregular for younger age groups).

Figures 4b to 4f show the corresponding estimates by cause of death. Clearly, cardiovascular diseases (Figure 4b) are responsible for most of the increase in the oldest age group; and neoplasms (Figure 4e) contribute a significant share to the increased risk for individuals above 50. Interestingly, for most death causes and most age groups, the relative increase in mortality fluctuates between 10 and 20%.

[Insert Figure 4 and 5 about here]

Figure 5 shows the results for hospital admissions. There are several differences compared to the findings for mortality in Figure 4. First, the effect is less heterogeneous with regard to age: also the least-affected age group of 10 to 19 year olds experiences a significant increase in admissions on hot days. Second, the relative increase in admissions is more stable across age groups and corresponds to an increase of between 5-10% for all age groups. Third, the effect is less concentrated in specific diagnoses; cardiovascular diseases are largely responsible for the increase in mortality, but they are only responsible for a small share of the increase in admissions on hot days. Individuals with respiratory, infectious or metabolic diseases all show increases in admission rates of around 10% on hot days; for metabolic diseases and people above 75, the increase in admission rates is even close to 30%.

5 Harvesting and the Monetized Health Costs of a *Hot Day*

Who dies during extreme heat events? Obviously, the answer to this question has important implications when quantifying the economic relevance of heat events. The literature discusses a phenomenon called the "harvesting hypothesis" (cf. Rabl, 2005; Fung et al., 2005). According to the harvesting hypothesis, heat events only temporarily lead to higher mortality rates. The hypothesis suggests, for example, that people who die during heat events would have otherwise died a few days later in the counterfactual scenario of the absence of the heat event. If this was true, then the medium-term population health effect of heat events would be substantially smaller than the immediate short-term effect. Empirically, an increase in mortality rates of predominately old people during heat events would support the harvesting hypothesis. Similarly, a decline in mortality rates in the days following a heat event would also be strongly in line with the harvesting hypothesis.

This paper makes several contributions to the harvesting debate. In the previous section, using the entire German population over a decade, we have shown how excess mortality varies by age for different diseases and death causes. We now turn to a systematic analysis of dynamic aspects of the heat-health relationship: First, we analyze the medium-term impact of a hot day on mortality by summing over and plotting mortality rates up to 30 days after a heat event. We do this separately for different causes of death. Second, we analyze the medium- to long-term impact of a hot day by aggregating the data at different levels in time and exploiting hot day variation at these levels, for example, the annual county level.

5.1 How Does Mortality Develop After Heat Events?

After having established a clear age gradient for hospital admissions and mortality on hot days, we examine how mortality rates develop in the days *following* a heat event. Figure 6 plots 30 estimated lags using flexible models like the one in equation (1). Figure 6a shows the coefficients for individual lags (left-hand side) and cumulative effects (right-hand side) for all-cause mortality. Apparently, the heat effect spikes on the day after the heat event and then returns to zero during the following week. The cumulative effect increases from 0.2 on the hot day to 0.7 ten days later, after which a harvesting effect kicks in and reduces the overall effect size. Mortality rates are thus repressed between day 10 and day 20 following a hot day. This "harvesting" reduces the cumulative effect from around 0.7 to around 0.5—after which it remains fairly stable.

As before, a large part of the effect is driven by cardiovascular diseases. Investigating this death cause separately in Figure 6b also reveals clear evidence of harvesting. Moreover, we find evidence for a harvesting effect in neoplasms (Figure 6e) which appears very plausible from a medical perspective. Cancer patients get admitted to hospitals and die during heat events but mortality then develops under proportionally in subsequent days and weeks. In contrast, the effects for respiratory diseases and infections are more persistent over time (Figures 6c and 6d).

[Insert Figure 6 about here]

Summing up, it appears to be the case that the general hypothesis of "harvesting" requires a more nuanced discussion and analysis. While, for some disease groups (such as cardiovascular diseases or cancer patients), the data show pattern that are clearly consistent with the harvesting hypothesis—excess mortality during heat events, followed by sharp decreases and an under-proportional development—no such clear pattern exists for other disease groups (such as respiratory and infectious diseases).

5.2 Aggregating Up: Exploiting Monthly and Annual Variation

In the next step, Table 5 aggregates the daily county-level data at different levels to provide evidence for medium-run effects of heat events. This is an alternative test for the relevance of the harvesting hypothesis. If it were true that heat events triggered persistent adverse health effects that would not have occurred in the counterfactual state, then an additional hot day should also significantly elevate the monthly and annual mortality and hospitalization rates—not just the daily one. However, due to data limitations and power issues, researchers often cannot implement this test because one obviously needs to observe enough years with enough variation in the annual

number of *Hot Days*.[11] In addition, the number of regional units of observations, in this case counties, must be sufficiently large. Our data and setting fulfill all of these conditions.

Panel A of Table 5 shows the results for deaths and Panel B shows the results for hospital admissions. All models are similar to equation (1). In the first column, we simply aggregate at the weekly level and differentiate between the first, second, third, and fourth week of and after a *Hot Day*. When aggregating at the weekly level, the coefficients remain statistically significant and relatively large. As for deaths in Panel A, the coefficient estimate for the first week of the heat event remains as large as the daily estimate in Table 2 and highly significant. In contrast, the estimates for the next two weeks are negative and statistically significant. As for hospital admissions in Panel B, the weekly effect size is only one third of the daily one. Additionally, the three weekly effects for the three following weeks have negative coefficients (but lack statistical precision). These findings suggest that, following spikes during heat events, decreases in admissions kick in earlier than decreases in deaths. However, the overall patterns are very similar and reinforce the harvesting hypothesis.

[Insert Table 5 about here]

Column (2) aggregates the data at the county-month level, and column (3) aggregates the data at the county-year level, solely exploiting the monthly and annual variation in *Hot Days*. As for mortality in Panel A, column (2) yields a highly significant estimate of 0.013 and column (4) a highly significant estimate of just 0.0025. The yearly coefficient is more than 100 times smaller than the daily one. This only translates into minor annual mortality increases of 0.08%, or 2 additional deaths per additional *Hot Day* in Germany. This finding is very similar to the one in Deschênes and Greenstone (2011) for the US, who find an annual age-adjusted mortality rate increase of 0.11% per additional hot day. Because the applied value of a statistical life year is not specific to this study but derived from a rich literature, the monetized health costs estimates are also comparable.

This finding is reinforced for hospital admissions in Panel B: the heat-admission coefficients decrease substantially to 0.03 in columns (2) and (3). These effect sizes translate into admission increases by about 0.05%—reduced by a factor of 100 as compared to the standard estimate in column (1) of Table 2. However, the coefficients in columns (2) and (3) remain statistically significant.

[11] Deschênes and Greenstone (2011) are an exception.

5.3 Monetizing the Health Loss of One Additional *Hot Day*

Finally, we assess and monetize the total health effects caused by extreme temperatures. Such quantitative measures may be useful to design appropriate public health measures, and they may eventually tell us something about the societal costs associated with climate change. However, we are reluctant to embark on a full-scale estimate of the health costs of climate change.

Given the complex nature of climate change, it is not surprising that long-term projections of its future trajectory are relatively vague. Concrete statements are hard to find in the famous Stern (2006) report. According to the IPCC (2007), it is very likely that hot extremes, heat waves and heavy precipitation events will continue to become more frequent (p. 46, 53). The underlying state-of-the art global climate model of the IPCC is the third version of the so-called *Hadley Centre Coupled Model (HadCM3)* (Pope et al., 2000). These climate models are very complex and require many assumptions and scenarios. Deschênes and Greenstone (2011) make use of the *HadCM3 model* and the "business-as-usual" scenario to predict the change in the number of *Hot Days* from 2070 to 2099 relative to 1968 to 2002 for different US regions.[12] For the region whose climate comes closest to Germany's, New England, Deschênes and Greenstone (2011) estimate a 20% increase in the number of *Hot Days*. Hübler et al. (2008) make use of the Regional Climate Model *REMO* and predict "two to five times as many hot days [for Germany from 2071 to 2100 relative to 1971 to 2000]" (p. 383).

Given the difficulty and inherent uncertainty of making such long-term predictions (Heal and Millner, 2014), we focus on the monetized health effects of one additional hot day for the following reasons: (i) Additional hot days are very plausible climate change predictions and are typically referred to in climate change models. (ii) One additional hot day is an intuitively plausible concept. The monetized health effects can be easily scaled up or down to alternative climate change predictions. (iii) One additional hot day represents an increase of about 14% in the total number of hot day in Germany, which is very much in line with the Deschênes and Greenstone (2011) prediction for New England using *HadCM3*. (iv) Finally, we abstain from estimating the impact of fewer cold days because projections concerning cold are not unambiguous. On the one hand, the IPCC (2007) projects that snow cover will contract (globally) in the future. On the other hand, loss of arctic sea ice has been linked to extreme cold weather in North America and Europe (Liu et al., 2012). In fact, climate change could result in both more heat *and* cold events in the mid-latitudes.

Table 6 summarizes the findings from the empirical models and calculates the total health costs of one *Hot Day* under competing assumptions. The basis for these calculations is Table 2 and

[12] Here *Hot Days* are days with a mean daily temperature above 90 °F.

equivalent tables that use the dependent variables *Hospital Days* and *Hospital Death*, see Appendix B. In line with the dependent variables (see Tables A1 and B1), the monetized health effects include (A) total hospital days due to a *Hot Day*, (B) total deaths after a hospital stay due to a *Hot Day*, (C) immediate death due to a *Hot Day*. Table 6 additionally differentiates by three main models: the (a) parsimonious model, the (b) model that includes effect modifiers at the day-county level, and the (c) the model outlined in column (4) of Table 5, which aggregates at the *year*-county level, thereby completely internalizing the harvesting effects.

The first four columns of Table 6 monetize the different costs of hospital stays. Column (1) multiplies the total number of hospital days by the average cost of one hospital day in Germany, which is €500 (German Federal Statistical Office, 2013b). Column (2) shows the loss in labor productivity by multiplying the approximate share of the working population, 50%, by the number of heat-related hospital days and the average daily gross wage in 2012, including employer-mandated benefits: €150 (German Federal Statistical Office, 2013a). This column likely overestimates the true effects, given that mostly the elderly are hospitalized on hot days, see Figure 5.

Columns (3) and (4) convert the number of hospital days into lost Quality-Adjusted Life Years (QALYs) by assuming that 365 hospital days equal a loss of one QALY (column (3)) and half a QALY (column (4)), respectively. We evaluate one QALY with €100,000 ($130,000) (Kniesner et al., 2010; Robinson et al., 2013).

[Insert Table 6 about here]

Column (5) monetizes the total number of deaths, which is the sum of deaths after a hospital stay and immediate deaths. Again, one QALY is evaluated at a value of €100,000. The first two rows use the data at the day-county level and ignore harvesting; here we assume that people who died would have lived another calendar year absent the heat event. The third row aggregates at the year-county level and accounts for harvesting; here we assume that people who died would have lived another 30 years.[13]

As seen in the final two columns: First, the upper and lower bound QALY assumptions barely affect the estimates (and neither do varying assumptions about their value). Second, the parsimonious county-day model yields the largest monetized health loss estimates. The approach that aggregates at the year-county level, internalizing harvesting, yields the lowest monetized health loss. Third, all estimates are relatively close and small in size. The estimated monetized losses range from €6m to €43m per hot day for an entire nation with a GDP of €2.5 trillion and 82

[13] 30 years is roughly the difference between the average current age of Germans and their life expectancy. Alternative assumptions do not alter the main findings.

million residents. The equivalent values for the US would lie between \$30 million and \$212 million or between €0.07 (\$0.10) and €0.52 (\$0.68) per resident.[14] Fourth, assuming that climate change permanently induces one additional hot day per year and taking the largest annual loss estimate of €43 million, the nominal health-related welfare loss over one life cycle of 80 years would accumulate to €3.4 billion for Germany. Applying a discount rate of 2.5% reduces this sum to €470 million or about €6 (\$8) per resident. The equivalent values for the US would be \$16.8 and \$2.3 billion, respectively.

When comparing these back of the envelope calculations with the economics literature on the impact of adverse weather and pollution on labor productivity, one finds that even ozone levels below the US regulatory thresholds would affect the productivity of agricultural workers (Zivin and Neidell, 2012). Accordingly, a reduction in ozone levels by 10ppb could lead to annual labor productivity benefits of \$700 million. Moretti and Neidell (2011) estimate that the annual costs of hospital admissions due to respiratory diseases accumulate to \$44 million per year for Los Angeles. Chang et al. (2016) estimate the US-wide labor cost savings as a result of reductions in PM2.5 levels to be about \$2 billion per year. Finally, Zivin and Neidell (2014) report that a hot day with maximum temperatures above 85 °F (29.4 °C) induces a re-allocation of time spent outdoors to indoor leisure as well as a labor productivity loss in "climate-exposed industries."

6 Conclusion

This paper assesses the adverse population health effects of extreme temperatures. At the day-county level, we link weather and pollution measures from more than 2,300 ambient monitors obtained over 10 years to two administrative datasets: (i) a mortality census comprising all deaths on German territory from 1999 to 2008, and (ii) a hospital census of all admissions from 1999 to 2008. These databases together allow us to comprehensively analyze the short-term and longer-term population health effects of heat and cold events.

First, confirming the findings of the existing literature, we find that extreme heat immediately affects population health negatively and leads to more hospitalizations and deaths. This finding holds irrespective of whether we use flexible temperature bin models or threshold models, rich OLS fixed effects models (as standard in economics) or random effect poisson models (as standard in epidemiology). The standard models in economics yield an increase in mortality of about 12% and an increase in hospital admissions of about 6% during heat events.

[14] Assuming an exchange rate of 1.3 and that the US has 311/82=3.8 times as many residents as Germany.

Second, comparing the standard modeling approaches in health economics and epidemiology, we find that the qualitative findings for the heat-mortality relationship are remarkably consistent, but the economic models yield stronger heat-mortality and heat-hospitalization gradients. Modifying our modelling approaches stepwise, we can rule out that differences in the definition of a hot day are driving effect size differences whenever they exist. Instead, hundreds of additional county fixed effects in the economic models, which net out unobserved time-invariant heterogeneity, appear to be responsible for the differences.

Third, the magnitude of the heat-health relationship significantly decreases when considering "effect modifiers", i.e., when controlling for other climatic and pollution conditions. For example, in the economic models the adverse health effects of a hot day is more than halved when comprehensively controlling for non-temperature related climatic conditions such as sunshine, high ozone, or particular matter concentrations that typically prevail during heat events. Adding lags of extreme temperature conditions further decreases effect sizes.

Fourth, in line with the sparse literature on this topic, the findings for cold are inconclusive. In these models, controlling for other weather conditions such as precipitation or wind-speed makes a crucial difference and negative effect estimates turn positive. This holds both for mortality and hospital admissions where, as expected, the cold-health gradient becomes larger with falling temperatures. For example, in our economic fixed effects models, relative to temperatures between 40 and 50 $^\circ$F, daily mean temperatures of less than 10 $^\circ$F increase deaths by 4% and admissions by 22%. By contrast, the epidemiological poisson models yield smaller but still positive cold-mortality associations. However, the cold-hospitalization gradient turns negative for these models when adding cubic pollution splines as effect modifiers.

Fifth, we apply several methods to test the so-called "harvesting hypothesis." This hypothesis implies that heat would only bring forward adverse population health effects by a couple of days and would not have a substantial long-term impact. In other words: excess mortality or admissions on hot days would be followed by sharp decreases in deaths or admissions—and the annual effects would be substantially smaller than the daily effects. We investigate the death rates in the 30 days following a heat event and differentiate by disease categories. We also investigate the age structure of those who were admitted or died on hot days, again by disease categories. Lastly, we exploit the richness of the data by aggregating them at the month-county and year-county level and solely using monthly and annual variation in the number of hot days as the identifying variation.

Our results clearly show that a nuanced discussion is important when it comes to harvesting because the findings depend on the disease category. We find very characteristic sharp increases, followed by sharp decreases, for cardiovascular and neoplastic mortality. However, we do not find

these patterns for disease categories such as infections or metabolic diseases. When differentiating by age groups, deaths and admissions sharply increase for people above 50 during heat events. However, this is only true for absolute effect sizes per 100,000 population—the relative increase in percent is remarkable consistent at around 10% for all age groups above 30. As for mortality, the main driver are heart diseases during hot days. In contrast, as for hospital admissions, respiratory and infectious diseases as well as admissions of cancer patients and patients with metabolic problems all play an important role.

Finally, we try to monetize the health effects of one additional hot day with temperatures above 30 ° C (86 ° F). We provide results for different approaches using different sets of assumptions. The total estimated health loss of one hot day represents a—highly unequally distributed—monetary welfare loss of up to €50 million ($65 million) per 100 million residents.

As a last point, we would like to note the limitations of this study. This paper solely studies the *health* effects of extreme temperatures. Moreoever, it does not consider health effects that lead to ambulatory doctor visits or no treatments at all. However, our calculations suggest that mild health effects do not seem to matter substantially when calculating total monetized health costs. A very large share of the serious health effects should be captured by this study. One important exception may be fetal health effects which may have long-lasting impacts (cf. van den Berg et al., 2006; Wilde et al., 2017). For example, Currie et al. (2015) estimate that the overall discounted long-term societal costs of being born with low birth weight are at least $100,000. We also acknowledge that we do not explicitly estimate (potential) adverse health effects triggered by avoidance behavior. However, this omission should not significantly impact the central findings of this study, which implicitly consider avoidance behavior in its estimates.[15] We also disregard adverse health effects from weather phenomena such as floods, tornadoes, and hurricanes. Lastly, this study abstains from estimating cumulative long-term health effects of ongoing and slowly evolving temperature changes.

More studies on more regions and outcome measures are instrumental for a better understanding of how climate and human health interact.

References

Anderson, B. G. and M. L. Bell (2002). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Environmental Health Perspectives 110*(9), 859863.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies 67*(3), 499–527.

[15] Obviously, for predictions about the future, assumptions about avoidance behavioral matter. The adverse health effects would be mitigated if avoidance behavior further increased (cf. Barreca et al., 2016) and reinforced if avoidance behavior decreased in the future.

Åström, D. O., F. Bertil, and R. Joacim (2011). Heat wave impact on morbidity and mortality in the elderly population: A review of recent studies. *Maturitas 69*(2), 99–105.

Barnett, A., S. Hajat, A. Gasparrini, and J. Rocklöv (2012). Cold and heat waves in the United States. *Environmental Research 112*, 218 – 224.

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2015). Convergence in adaptation to climate change: Evidence from high temperatures and mortality, 1900-2004. *American Economic Review 105*(5), 247–51.

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2016). Adapting to climate change: The remarkable decline in the U.S. temperature-mortality relationship over the 20th century. *Journal of Political Economy 124*(1), 105–159.

Barreca, A. I. (2012). Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management 63*(1), 19–34.

Basu, R. and J. M. Samet (2002). Relation between elevated ambient temperature and mortality: A review of the epidemiologic evidence. *Epidemiologic Reviews 24*(2), 190–202.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

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

Bobb, J. F., F. Dominici, and R. D. Peng (2013). Reduced hierarchical models with application to estimating health effects of simultaneous exposure to multiple pollutants. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 62*(3), 451–472.

Bobb, J. F., R. D. Peng, M. L. Bell, and F. Dominici (2014). Heat-related mortality and adaptation to heat in the United States. *Environmental Health Perspectives 122*(8), 811–816.

Braga, A. L. F., A. Zanobetti, and J. Schwartz (2001). The time course of weather-related deaths. *Epidemiology 12*(6), 662–667.

Braga, A. L. F., A. Zanobetti, and J. Schwartz (2002). The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. *Environmental Health Perspectives 110*(9), 859863.

Cameron, C. and D. L. Miller (2011). Robust Inference with Clustered Data. In A. Ullah and D. E. Giles (Eds.), *Handbook of Empirical Economics and Finance*, pp. 128. CRC Press.

Chang, T., J. S. G. Zivin, T. Gross, and M. J. Neidell (2016). Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy 8*(3), 141–169.

Chen, X., H. Hong, and D. Nekipelov (2011). Nonlinear models of measurement errors. *Journal of Economic Literature 49*(4), 901–37.

Chung, Y., Y. Lim, Y. Honda, Y. Guo, M. Hashizume, M. Bell, B. Chen, and H. Kim (2015). Mortality related to extreme temperature for 15 cities in northeast Asia. *Epidemiology 26*(2), 255–262.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Do housing prices reflect environmental health risks? Evidence from more than 1600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Currie, J. and M. Neidell (2005). Air pollution and infant health: What can we learn from California's recent experience? *The Quarterly Journal of Economics 120*(3), 1003–1030.

Curriero, F. C., K. S. Heiner, J. M. Samet, S. L. Zeger, L. Strug, and J. A. Patz (2002). Temperature and mortality in 11 cities of the Eastern United States. *American Journal of Epidemiology 155*(1), 80–87.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature 52*(3), 740–798.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2016). The mortality and medical costs of air pollution: Evidence from changes in wind direction. Technical Report 22796, National Bureau of Economic Research.

Deschênes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics 46*, 609–619.

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

Deschênes, O., M. Greenstone, and J. Guryan (2009). Climate change and birth weight. *American Economic Review 99*(2), 211–17.

Deschênes, O. and E. Moretti (2009). Extreme weather events, mortality and migration. *The Review of Economics and Statistics 91*(4), 659–681.

Deutscher Wetterdienst (DWD) (2017). *Wetterlexikon: Heißer Tag.* https://www.dwd.de/DE/service/lexikon/Functions/glossar.html;jsessionid=C7A565D8B76F2E8069C7C3A0AE08C80E.live11052?lv2=101094&lv3=101162, retrieved October 16, 2017.

Dominici, F., R. D. Peng, C. D. Barr, and M. L. Bell (2010). Protecting human health from air pollution: Shifting from a single-pollutant to a multipollutant approach. *Epidemiology 21*(2), 187–194.

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics 80*(4), 549–560.

Federal Institute for Research on Building, Urban Affairs and Spatial Development (2012). *INKAR 2011. Indicators and Maps on Spatial Development.* CD-ROM.

Fung, K., D. Krewski, R. Burnett, and F. Dominici (2005). Testing the harvesting hypothesis by time-domain regression analysis. I: Baseline analysis. *Journal of Toxicology and Environmental Health, Part A 68*(13-14), 1137–1154.

Gasparrini, A., Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, and J. e. a. Schwartz (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet 386*(9991), 369 – 375.

German Federal Statistical Office (2013a). www.destatis.de, last accessed on April 20, 2013.

German Federal Statistical Office (2013b). *Federal Health Monitoring.* www.gbe-bund.de, last accessed on April 20, 2013.

Goldberg, M. S., A. Gasparrini, B. Armstrong, and M.-F. Valois (2011). The short-term influence of temperature on daily mortality in the temperate climate of Montreal, Canada. *Environmental Research 111*(6), 853 – 860.

Hajat, S., B. Armstrong, M. Baccini, A. Biggeri, L. Bisanti, A. Russo, A. Paldy, B. Menne, and T. Kosatsky (2006). Impact of high temperatures on mortality: Is there an added heat wave effect? *Epidemiology 17*(6), 632–638.

Hajat, S., B. G. Armstrong, N. Gouveia, and P. Wilkinson (2005). Mortality displacement of heat-related deaths: A comparison of Delhi, São Paulo, and London. *Epidemiology 16*(5), 613–620.

Hanigan, I., G. Hall, and K. B. Dear (2006). A comparison of methods for calculating population exposure estimates of daily weather for health research. *International Journal of Health Geographics 5*(38), 1–6.

Heal, G. and A. Millner (2014). Reflections: Uncertainty and decision making in climate change economics. *Review of Environmental Economics and Policy 8*(1), 120–137.

Heaton, M. J. and R. D. Peng (2012). Flexible distributed lag models using random functions with application to estimating mortality displacement from heat-related deaths. *Journal of Agricultural, Biological, and Environmental Statistics 17*(3), 313–331.

Hondula, D. M., J. Balling, Robert C., J. K. Vanos, and M. Georgescu (2015). Rising temperatures, human health, and the role of adaptation. *Current Climate Change Reports 1*(3), 144–154.

Hübler, M., G. Klepper, and S. Peterson (2008). Costs of climate change: The effects of rising temperatures on health and productivity in Germany. *Journal of Ecological Economics 68*, 381–393.

Hyslop, D. R. and G. W. Imbens (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics 19*(4), 475–481.

IPCC (2007). *Climate Change 2007: Synthesis Report.* http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_synthesis_report.htm, last accessed on January 25, 2013.

Janke, K. (2014). Air pollution, avoidance behaviour and children's respiratory health: Evidence from England. *Journal of Health Economics 38*, 23–42.

Katsouyanni, K., G. Touloumi, E. Samoli, A. Gryparis, A. Le Tertre, Y. Monopolis, G. Rossi, D. Zmirou, F. Ballester, A. Boumghar, H. Anderson, B. Wojtyniak, A. Paldy, R. Braunstein, J. Pekkanen, C. Schindler, and J. Schwartz (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology 12*(5), 521–531.

Kniesner, T., W. Viscusi, and J. Ziliak (2010). Policy relevant heterogeneity in the value of statistical life: New evidence from panel data quantile regressions. *Journal of Risk and Uncertainty 40*(1), 15–31.

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016). Caution, drivers! Children present: Traffic, pollution, and infant health. *Review of Economics and Statistics 98*(2), 350–366.

Liu, J., J. A. Curry, H. Wang, M. Song, and R. M. Horton (2012). Impact of declining Arctic sea ice on winter snowfall. *Proceedings of the National Academy of Sciences of the United States of America (PNAS) 109*(11), 4074–4079.

Maddala, G. (1977). *Econometrics.* McGraw-Hill Book Co. New York.

Moretti, E. and M. Neidell (2011). Pollution, health, and avoidance behavior: Evidence from the ports of Los Angeles. *Journal of Human Resources 46*(1), 154–175.

Nordio, F., A. Zanobetti, E. Colicino, I. Kloog, and J. Schwartz (2015). Changing patterns of the temperaturemortality association by time and location in the US, and implications for climate change. *Environment International 81*, 80 – 86.

OECD (2017). OECD Health Data: Health status. http://dx.doi.org/10.1787/data-00540-en, last accessed on August 16, 2017.

Peng, R. D., J. F. Bobb, C. Tebaldi, L. McDaniel, M. L. Bell, and F. Dominici (2011). Toward a quantitative estimate of future heat wave mortality under global climate change. *Environmental Health Perspectives 119*(5), 701–706.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics 16*(2-3), 123–146.

Rabl, A. (2005). Air pollution mortality: Harvesting and loss of life expectancy. *Journal of Toxicology and Environmental Health, Part A 68*(13-14), 1175–1180.

Robinson, A., D. Gyrd-Hansen, P. Bacon, R. Baker, M. Pennington, and C. Donaldson (2013). Estimating a WTP-based value of a QALY: The 'chained' approach. *Social Science & Medicine 92*(C), 92–104.

Schwartz, J., M. Lee, P. Kinney, S. Yang, D. Mills, M. Sarofim, R. Jones, R. Streeter, A. Juliana, J. Peers, and R. Horton (2015). Projections of temperature-attributable premature deaths in 209 U.S. cities using a cluster-based poisson approach. *Environmental Health 14*(1), 85.

Schwartz, J., J. M. Samet, and J. A. Patz (2004). Hospital admissions for heart disease: The effects of temperature and humidity. *Epidemiology 15*(6), 755–761.

SOEP (2012). Data for the years: 1984-2011. doi:10.5684/soep.v29, version 29.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human Resources 50*(2), 301–316.

Son, J.-Y., M. L. Bell, and J.-T. Lee (2014). The impact of heat, cold, and heat waves on hospital admissions in eight cities in Korea. *International Journal of Biometeorology 58*(9), 1893–1903.

Son, J.-Y., N. Gouveia, M. A. Bravo, C. U. de Freitas, and M. L. Bell (2016). The impact of temperature on mortality in a subtropical city: Effects of cold, heat, and heat waves in São Paulo, Brazil. *International Journal of Biometeorology 60*(1), 113–121.

Stafoggia, M., F. Forastiere, P. Michelozzi, and C. A. Perucci (2009). Summer temperature-related mortality: Effect modification by previous winter mortality. *Epidemiology 20*(4), 575–583.

Stern, N. (2006). Stern review on the economics of climate change. http://webarchive.nationalarchives.gov.uk/+/http://www.hm-treasury.gov.uk/sternreview_index.htm, last accessed on March 1, 2013.

United States Global Change Research Program (2009). Global Climate Change Impacts in the United States. In T. R. Karl, J. M. Melillo, and T. C. Peterson (Eds.), *United States Global Change Research Program*. Cambridge University Press.

van den Berg, G. J., M. Lindeboom, and F. Portrait (2006). Economic conditions early in life and individual mortality. *American Economic Review 96*(1), 290–302.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP)--Scope, Evolution and Enhancements. *Journal of Applied Social Science Studies (Schmollers Jahrbuch) 127*(1), 139–169.

White, C. (2017). The dynamic relationship between temperature and morbidity. *Journal of the Association of Environmental and Resource Economists 4*(4), 1155–1198.

Wilde, J., B. H. Apouey, and T. Jung (2017). The effect of ambient temperature shocks during conception and early pregnancy on later life outcomes. *European Economic Review 97*(Supplement C), 87 – 107.

Zhang, X., X. Zhang, and X. Chen (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management 85*(Supplement C), 81 – 94.

Zivin, J. G. and M. Neidell (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics 32*(1), 1 – 26.

Zivin, J. S. G. and M. J. Neidell (2012). The impact of pollution on worker productivity. *American Economic Review 102*(7), 3652–3673.

Zivin, J. S. G. and M. J. Neidell (2013). Environment, health, and human capital. *Journal of Economic Literature 51*(3), 689–730.

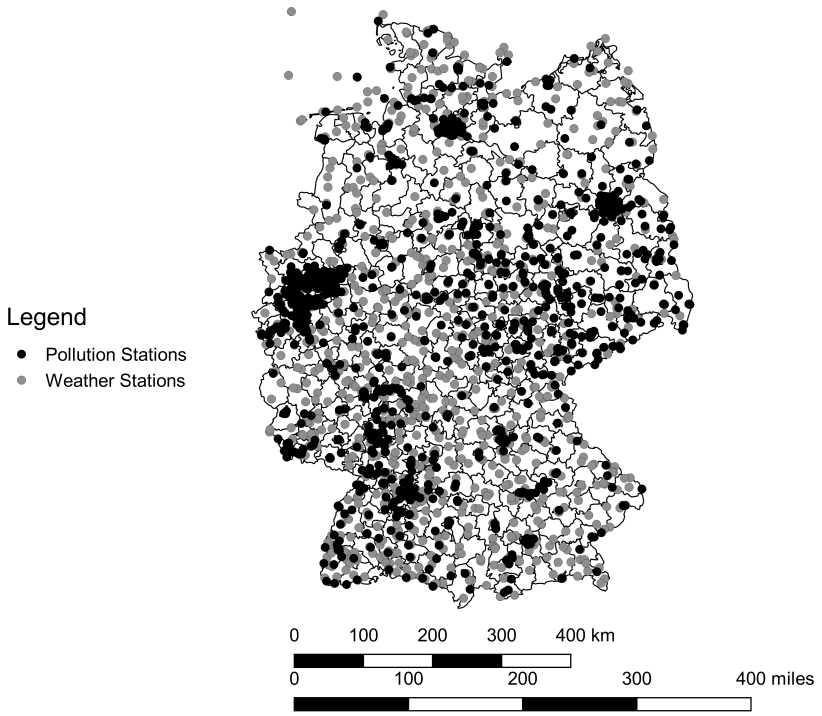
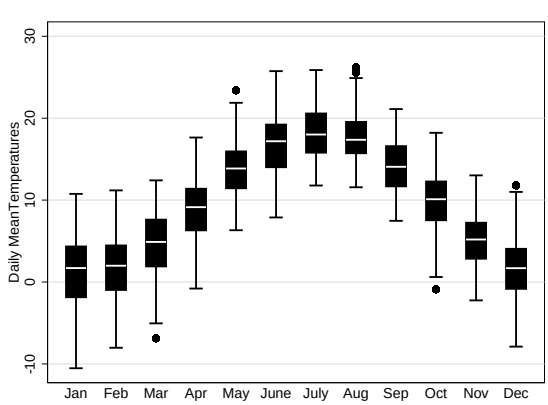
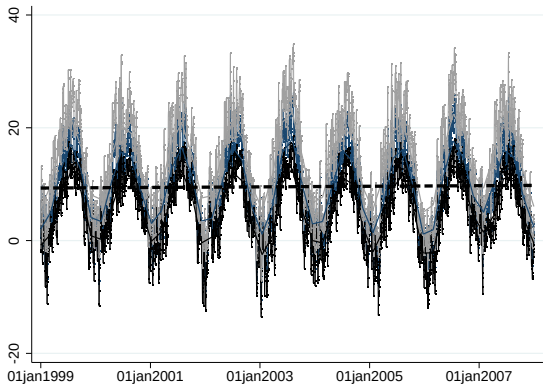


Figure 1: Distribution of Official German Ambient Weather and Pollution Monitors

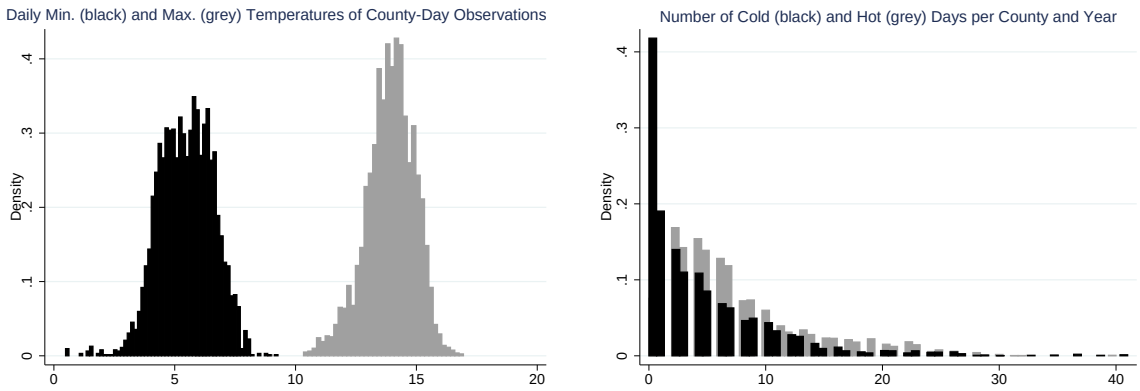


(a) Mean Temperatures (in $^\circ$C) by Months



(b) Mean, Min. and Max. Temperatures (in $^\circ$C) over Time

Figure 2: Distribution of Temperatures 1999-2008

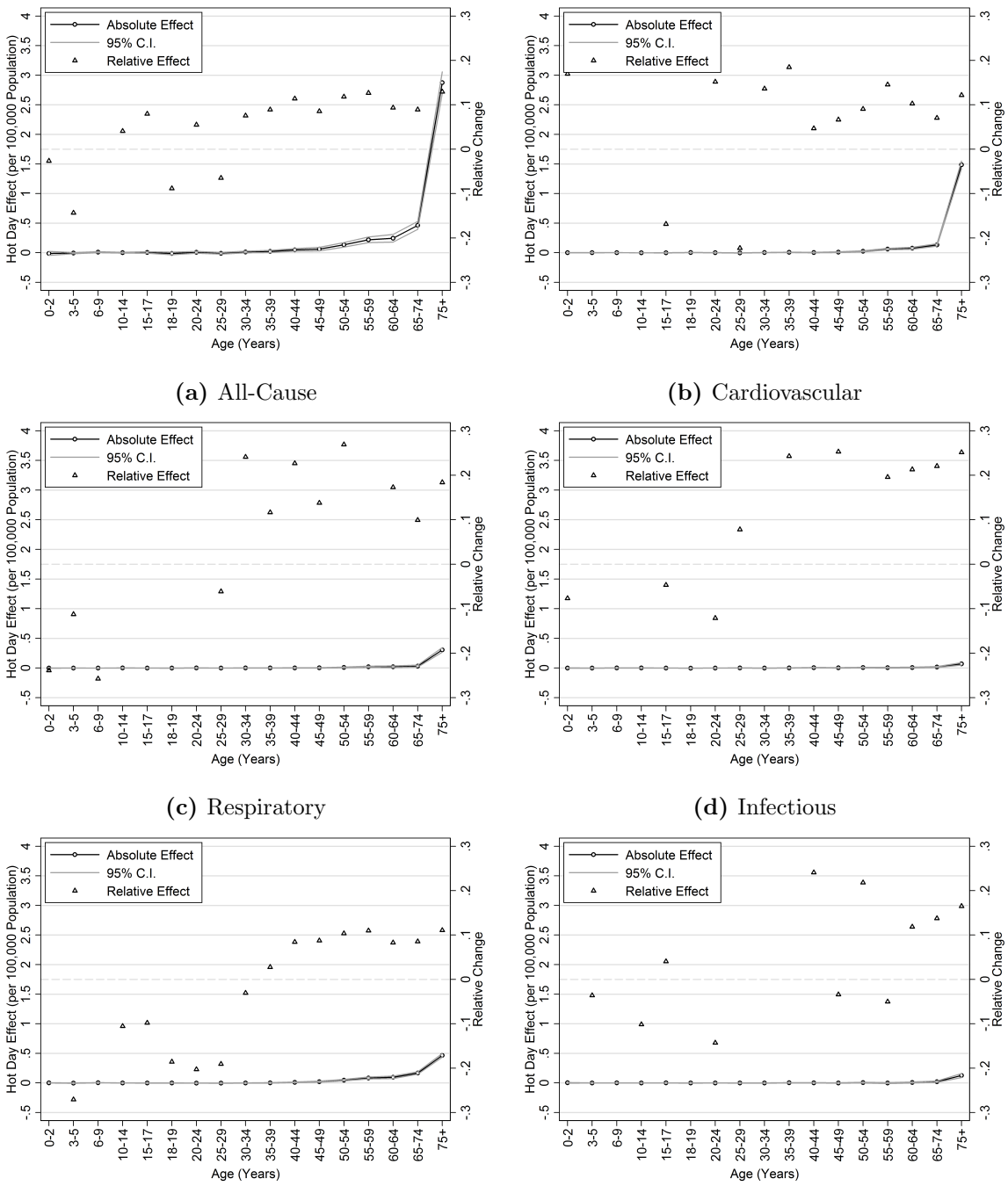


(a) Min. and Max. Temperatures (in $^\circ$C)

(b) Number of Hot and Cold Days

Figure 3: Distribution of Temperatures and Hot + Cold Days

Note: Figure 3a shows the county-day distributions of minimum temperatures (avg. 5.5 $^\circ$C; 42 $^\circ$F) and maximum temperatures (avg. 13.9 $^\circ$C; 57 $^\circ$F). Figure 3b shows the distribution of *Hot* (max. temp>30 $^\circ$C) and *Cold Days* (min. temp<-10 $^\circ$C).



(a) All-Cause

(b) Cardiovascular

(c) Respiratory

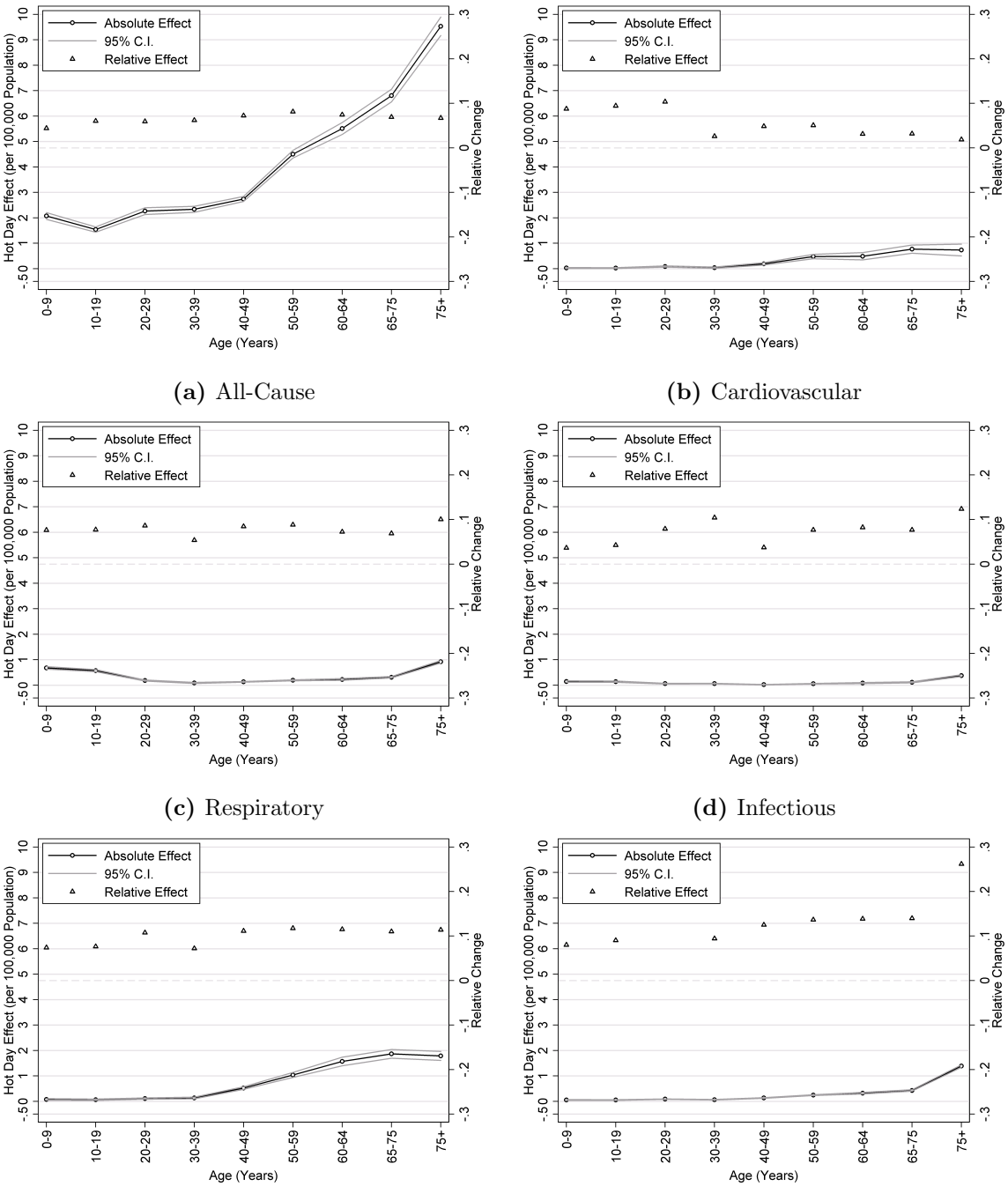
(d) Infectious

(e) Neoplasm

(f) Metabolism

Figure 4: Age Structure of Heat-Related Mortality by Disease Type

Note: The figures plot the coefficient of *Hot Day* of a model similar to equation (1), estimated for 17 age groups separately, with the mortality rate as the dependent variable. The left y-axis provides the absolute effect per 100,000 population and the right y-axis provides the relative effect in percent.



(a) All-Cause

(b) Cardiovascular

(c) Respiratory

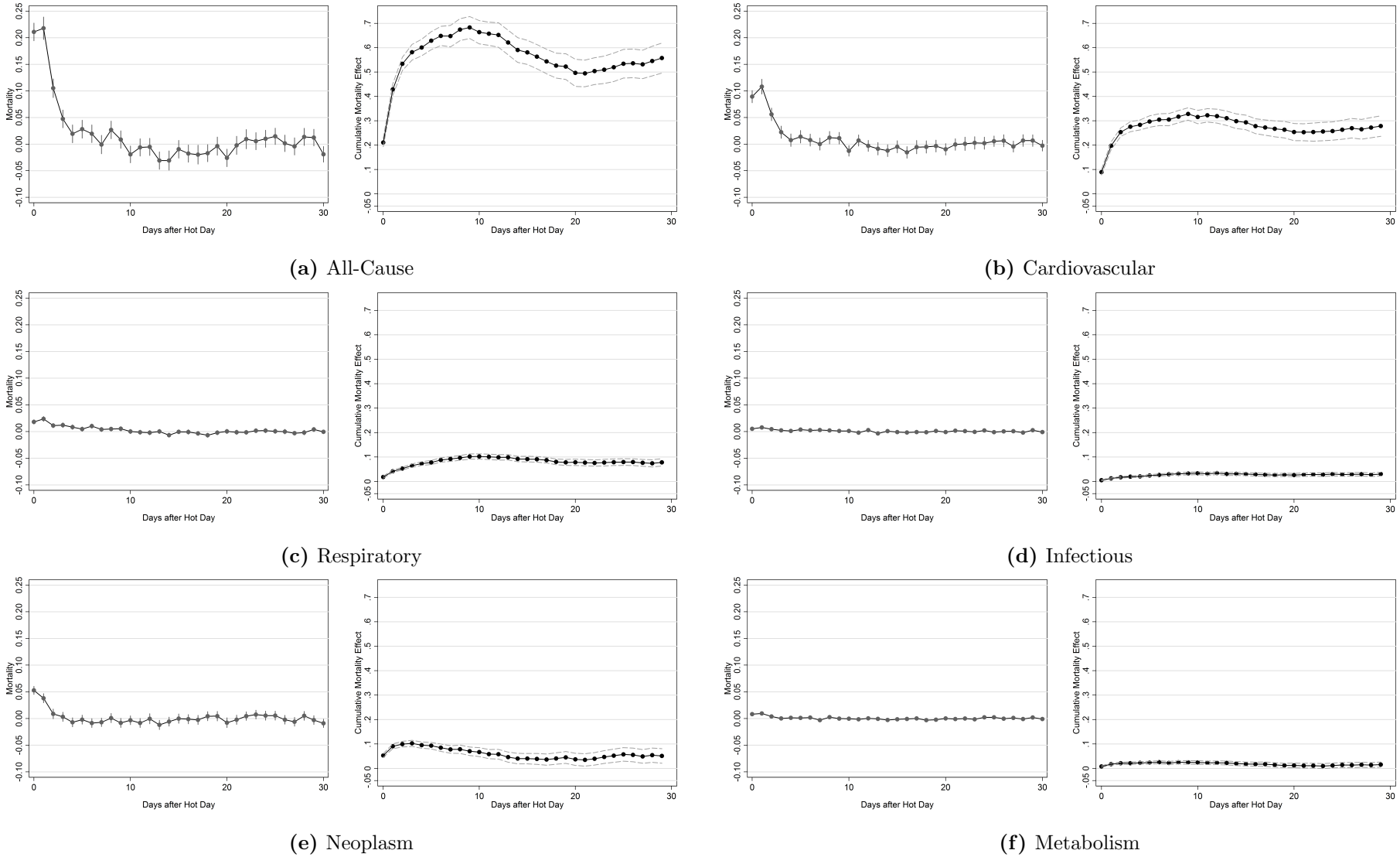
(d) Infectious

(e) Neoplasm

(f) Metabolism

Figure 5: Age Structure of Heat-Related Hospitalizations by Disease Type

Note: The figures plot the coefficient of *Hot Day* of a model similar to equation (1), estimated for 9 age groups separately, with the hospitalization rate as the dependent variable. The left y-axis provides the absolute effect per 100,000 population and the right y-axis provides the relative effect in percent.



(a) All-Cause

(b) Cardiovascular

(c) Respiratory

(d) Infectious

(e) Neoplasm

(f) Metabolism

Figure 6: Mortality by Cause Up to 30 Days Following a *Hot Day*

Note: The figure plots *Hot Day* along with 30 lags in a model similar to equation (1). The plotted coefficients represent changes in percent.

Table 1: Select Published Empirical Papers on the Health Effects of Extreme Temperatures, Developed Countries

A: Economic Journals	Area	Period	Unit of Obs.	Method	Outcomes	Var. of Interest	Climate Controls
Deschênes and Moretti (2009), REStat	continental US, white deaths	'72-'88	county-day	county-month-year FE	mortality rate (cause, age, gender)	hot days, cold days	precip.
Deschênes et al. (2009), AER: PP	continental US	'72-'88	ind.-pregn.,	county-year FE;	birth wgt, LBW during gestation	# days in 5 temp. bins	N/A N/A
Deschênes and Greenstone (2011), AEJ: Applied	continental US	'68-'02	county by year	county FE, state-year FE	mortality rate	# days in 10 temp. bins	precip.
Barreca (2012), JEEM	373/3,100+ US counties	'73-'02	county-month	county-month FE, county-month trends	mortality rate	mean temp.& precip.	temp.-precip. interactions
Barreca et al. (2016), JPE	US	1900-2004	state-month	state-month FE, year-month FE	mortality rate	# days in 10 temp. bins	precip.
White (2017), JAERE	CAL/US, all hospitalizations	1900-2004 2005-2014	state-month zip-day	state-month FE, zip-week, county-year FE	mortality rate hospitalization rate	# days in 10 temp. bins	precip. precip.
B: Epidemiological & Bio-Stat.	**Journals**						
Curriero et al. (2002), Am J Epidemiology	11 US cities	'73-'94	daily	log-linear poisson, distr. lag models	#deaths 3 ICD groups	avg. temp., 10 lags	smooth. spline
Braga et al. (2002), Env Health Persp	12 US cities	'86-'93	daily	log-linear poisson, distr. lag models	#deaths 4 ICD groups	temp., 20 lags	humidity, air press., smooth. param.
Hajat et al. (2005), Epidemiology	Delhi, London, São Paulo	'91-'94	daily	log-linear poisson, distr.lag models	#deaths 3 ICD groups	heat/cold day, 28 lags	humidity, precip. PM_{10}, cubic splines
Hajat et al. (2006), Epidemiology	London, Budapest, Milan	'70-'03	daily, June-Sep	log-linear poisson, AR-structure	#deaths, 2 lags 3 ICD groups	heat wave var. def.	black smoke, O_3 cubic splines
Anderson and Bell (2002), Epidemiology	107 US commun.	'87-'00	daily	log-linear poisson,	#deaths 2 ICD groups	heat/cold wave, 25 lags	O_3, PM_{10} cubic splines
Peng et al. (2011), Env Health Persp	Chicago, US	'87-'05	daily, May-Oct	log-linear poisson	#deaths	heat wave, climate scenarios	temp., O_3 splines
Goldberg et al. (2011), Env Research	Montreal, CA	'84-'07	daily	log-linear poisson distr. lag models	#deaths 3 ICD groups	max temp., hot/cold day	O_3, NO_2 cubic splines
Heaton and Peng (2012), J Agric Biol Env S	4 metrop. areas, US	'01-'05	daily, April-Sep	distributed lag models, Gaussian procedures	#deaths	avg temp., 60 lags	cubic splines
Barnett et al. (2012), Env Research	99 cities, US	'87-'00	daily	log-linear poisson, Bayesian, lags	#deaths 3 ICD groups	heat/cold wave, var. def.	splines
Bobb et al. (2014), JAMA	1943 US counties Medicare enrollees	'99-'10	daily	log-linear regressions matching by county-week	#admissions, 283 disease groups	heat wave, 7 lags	year FE
Bobb et al. (2014), Env Health Persp	105 US cities	'87-'05	daily, May-Oct	log-linear poisson, hier. Bayes	#deaths	avg. temp.	age cat. cubic splines
Son et al. (2014), Int J Biometeorol	8 Korean cities	'03-'08	daily, Mar-Aug + Sep-Feb	log-linear poisson, hier. Bayes	#admissions	avg. temp.	cubic splines
Schwartz et al. (2015), Env Health	209 US cities	'73-'06	daily	log-linear poisson, cluster analysis	#deaths	avg. temp.	temp
Nordio et al. (2015), Env Health	211 US cities	'62-'06 (no '67-'73)	monthly	log-linear poisson, cluster anal. + meta reg.	#deaths 5 lags	avg. temp.	temp
Gasparrini et al. (2015), Lancet	384 locations 13 countries	depends, max. '85-'12	daily	log-linear poisson, distr. lag models	#deaths 21 lags	extreme heat + cold	temp cubic splines
Chung et al. (2015), Epidemiology	3+6+6 cities in Taiwan, Korea, Japan	depends, max. '85-'12	daily	log-linear poisson, hier. Bayes	#deaths 30 lags	heat/cold var. def.	humidity, air press. cubic splines

Table 2: Temperature, Mortality and Hospitalizations: Economic Fixed Effects Models

	Temperature Bin Model					Threshold Model	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Mortality per 100,000 population							
<10 °F	-0.0187	-0.0188	-0.0466	0.1209*	0.1067		
	(0.048)	(0.048)	(0.048)	(0.049)	(0.055)		
10-20 °F	0.0050	0.0050	-0.0194	0.1138***	0.0782***		
	(0.014)	(0.014)	(0.014)	(0.016)	(0.017)		
20-30 °F	-0.0332***	-0.0332***	-0.0304***	0.0683***	0.0332***		
	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)		
30-40 °F	-0.0229***	-0.0229***	-0.0172***	0.0361***	0.0144***		
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)		
40-50 °F	ref.	ref.	ref.	ref.	ref.		
50-60 °F	0.0408***	0.0408***	0.0281***	-0.0206***	0.0008		
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)		
60-70 °F	0.1328***	0.1328***	0.0936***	-0.0062	0.0151*		
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)		
70-80 °F	0.3544***	0.3545***	0.2478***	0.0965***	0.0780***		
	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)		
80-90 °F	0.9559***	0.9557***	0.6893***	0.5065***	0.2409***		
	(0.040)	(0.040)	(0.039)	(0.040)	(0.033)		
Hot Day Effect	0.3475***	0.3475***	0.2470***	0.1123***	0.0763***	0.3198***	0.0948***
s.e.	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Pct. change	12.177	12.178	8.659	3.938	2.674	11.209	3.324
Cold Day Effect	-0.0579***	-0.0579***	-0.0571***	0.0862***	0.0499***	0.0250*	0.0195
s.e.	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(0.011)	(0.013)
Pct. change	-2.029	-2.029	-2.004	3.021	1.750	0.877	0.684
R-squared	0.03	0.03	0.03	0.03	0.03	0.03	0.03
N. of cases	1,590,501	1,590,501	1,429,899	1,429,899	1,429,059	1,590,501	1,429,059
Panel B: Hospital Admissions per 100,000 population							
<10 °F	2.1188**	2.1244**	-0.3325	11.0228***	13.7168***		
	(0.745)	(0.750)	(0.822)	(0.822)	(1.036)		
10-20 °F	-2.2236***	-2.2201***	-4.6711***	2.4397***	4.8323***		
	(0.213)	(0.213)	(0.286)	(0.297)	(0.367)		
20-30 °F	-2.5313***	-2.5329***	-3.4483***	0.7142***	2.0300***		
	(0.096)	(0.096)	(0.120)	(0.132)	(0.149)		
30-40 °F	-0.7699***	-0.7694***	-1.5315***	0.1811	0.9333***		
	(0.055)	(0.055)	(0.085)	(0.093)	(0.088)		
40-50 °F	ref.	ref.	ref.	ref.	ref.		
50-60 °F	1.1232***	1.1216***	0.8890***	-0.0963	0.2784***		
	(0.061)	(0.061)	(0.092)	(0.084)	(0.078)		
60-70 °F	1.7562***	1.7558***	0.2636	-0.5795***	-0.6134***		
	(0.063)	(0.063)	(0.148)	(0.148)	(0.149)		
70-80 °F	3.5465***	3.5433***	1.1668***	0.4125	0.6047		
	(0.117)	(0.117)	(0.273)	(0.267)	(0.262)		
80-90 °F	8.8118***	8.8174***	5.3014***	2.9667***	3.480***		
	(0.329)	(0.330)	(0.499)	(0.445)	(0.433)		
Hot Day Effect	3.4429***	3.4409***	1.6834***	0.5743**	0.3663	2.9080***	1.0790***
	(0.103)	(0.103)	(0.241)	(0.235)	(0.229)	(0.225)	(0.147)
Change in %	5.937	5.934	2.897	0.988	0.631	5.015	1.860
Cold Day Effect	-2.6452***	-2.6430***	-3.8223***	3.6777***	2.5616***	-1.2194***	2.1980***
	(0.122)	(0.122)	(0.166)	(0.199)	(0.2446)	(0.121)	(0.212)
Change in %	-4.562	-4.558	-6.578	3.737	6.339	-2.103	3.789
R-squared	0.48	0.48	0.52	0.53	0.53	0.48	0.53
N	1,590,454	1,590,454	1,429,928	1,429,928	1,408,356	1,590,454	1,408,356
County, week + month-year fixed effects	yes	yes	yes	yes	yes	yes	yes
Age, gender + hospital controls	yes	yes	yes	yes	yes	yes	yes
Annual county-level controls	no	yes	yes	yes	yes	no	yes
Pollution measures	no	no	yes	yes	yes	no	yes
Additional weather controls	no	no	no	yes	yes	no	yes
Lags of temperature variables	0	0	0	0	30	0	30

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are clustered at the county level. Regressions are weighted by the yearly county population. Data sources are discussed in Section 2. All specifications estimate the model in equation (1) by OLS, where the first five columns estimate the temperature bin model and the last two columns estimate the threshold model (see main text). In columns (1) to (5) the reported "Hot Day Effect" and "Cold Day Effect" are based on the approach described in Section 3.3. In Panel A, the dependent variable is the daily mortality rate per 100,000 population at the county level (mean: 2.99). In Panel B, the dependent variable is the daily hospital admission rate per 100,000 population at the county level (mean: 57.99). Columns (3) to (5) and (7) have fewer observations because PM_{10} data for 2000 are not available.

Table 3: Extreme Temperatures, Mortality, and Hospitalizations: Epidemiological Poisson Spline Models

Extreme Temperature	**Absolute** ($< 14^{\circ}$F and $> 86^{\circ}$F)				**Relative** (Percentiles 2.5 and 97.5)			
Panel A: Mortality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hot Day	0.0751***	0.0754***	0.0680***	0.0425***				
	(0.000)	(0.000)	(0.000)	(0.000)				
Cold Day	0.0884***	0.0876***	0.0190***	0.0006***				
	(0.000)	(0.000)	(0.000)	(0.000)				
Hot Day II					0.0699***	0.0703***	0.0607***	0.0391***
					(0.000)	(0.000)	(0.000)	(0.000)
Cold Day II					0.0846***	0.0838***	0.0327***	0.0141***
					(0.000)	(0.000)	(0.000)	(0.000)
N. of cases	1,590,501	1,590,501	1,429,899	1,429,059	1,590,501	1,590,501	1,429,899	1,429,059
Panel B: Hospital Admissions								
Hot Day	0.0138***	0.0145***	-0.0125***	-0.0077***				
	(0.000)	(0.000)	(0.000)	(0.000)				
Cold Day	-0.0105***	-0.0114***	-0.0254***	-0.0284***				
	(0.000)	(0.000)	(0.000)	(0.000)				
Hot Day II					0.0116***	0.0125***	-0.0144***	-0.0093***
					(0.000)	(0.000)	(0.000)	(0.000)
Cold Day II					-0.0354***	-0.0366***	-0.0538***	-0.0506***
					(0.000)	(0.000)	(0.000)	(0.000)
N. of cases	1,590,454	1,590,454	1,429,928	1,408,356	1,590,454	1,590,454	1,429,928	1,408,356
Cubic date splines and day of week	yes	yes	yes	yes	yes	yes	yes	yes
Age, gender & hospital controls	yes	yes	yes	yes	yes	yes	yes	yes
Annual county-level controls	no	yes	yes	yes	no	yes	yes	yes
Cubic CO, NO_2, SO_2, PM_{10}, and O_3 splines	no	no	yes	yes	no	no	yes	yes
Lags of *Hot Day* (II) and *Cold Day* (II)	0	0	0	30	0	0	0	30

* p<0.1, ** p<0.05, *** p<0.01; standard errors are in parentheses. Regressions are weighted by the yearly county population. Data sources are discussed in Section 2. All specifications estimate the Poisson model in equation (2) with random effects. Each column in each panel represents one model. Models only differ by the sets of covariates included as indicated. In Panel A, the dependent variable is the daily mortality rate per 100,000 population at the county level (mean: 2.99). In Panel B, the dependent variable is the daily hospital admission rate per 100,000 population at the county level (mean: 57.99). All variables of interest, i.e. *Hot Day II*, are defined as discussed in Section 2.3 and Appendix C. *Hot Day II* equals one when the average daily temperature exceeds the 97.5^{th} percentile of the county-level temperature distribution over all years. *Cold Day II* equals one when the average daily temperature does not exceed the 2.5^{nd} percentile of the county-level temperature distribution over all years. Columns (3), (4), (7) and (8) have fewer observations because PM_{10} data for 2000 are not available.

Table 4: The Impact of Extreme Heat on Hospital Admissions: Robustness Checks

Panel A	**cluster at state level (1)**	**2-way cluster (2)**	**correct for spatial dependence (3)**	**linear & quadr. time trends (4)**	**linear & quadr. state time trends (5)**	**linear & quadr. county time trends ['06-'08] (6)**	**county by year FE ['06-'08] (7)**
Hot Day	2.9083***	2.9083***	2.9083***	2.9083***	2.6515***	2.1616***	2.2053***
	(0.2376)	(0.2474)	(1.0899)	(0.1585)	(0.1303)	(0.2188)	(0.2287)
change in %	+5.0%	+5.0%	+5.0%	+5.0%	+4.6%	+3.7%	+3.8%
N	1,590,454	1,590,454	1,590,454	1,590,454	1,590,454	467,770	467,770

Panel B	**no 2000 (1)**	**log (2)**	**×(temp.>32 ° C) (3)**	**× warm region (4)**	**× cold region (5)**	**No counties without monitors (6)**
Hot Day×[column header]			0.5751***	-1.0389***	0.4892**	
			(0.0603)	(0.1911)	(0.2459)	
Hot Day	3.1113***	0.0573***	1.6456***	3.3788***	2.8509***	2.7047***
	(0.1741)	(0.0042)	(0.1598)	(0.1684)	(0.1685)	(0.9136)
max. daily temp.			0.1654***			
			(0.0039)			
N	1,429,928	1,590,454	1,590,454	1,590,454	1,590,454	1,274,615

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are clustered at the county level except for columns (1) of Panel A which clusters at the state, column (2) of Panel A which clusters at the county and day level (2-way cluster), and column (3) of Panel A which corrects for spatial correlation (Driscoll and Kraay, 1998). Regressions are weighted by the yearly county population. Data sources are discussed in Section 2. Each column in each panel represents a model that does not control for other weather and pollution conditions. However controlling for them yields similar results (available upon request). The dependent variable is always the hospitalization rate (mean: 57.99, see Table B1); the reference estimate is the one in Column (1) of Table 2. All specifications estimate a model similar to equation (1) by OLS. Column (4) of Panel A adds a nation-wide linear and quadratic time trends. Column (5) adds state-level time trends and column (6) adds county-level time trends (for 2006-2008 only because of computer memory constraints). Column (7) of Panel A includes county-year FE. The first column in Panel B excludes the year 2000 for which no PM_{10} data are available and column (2) uses the logarithm of the normalized hospitalization rate as dependent variable. Column (3) adds a continuous measure for the maximum daily temperature as well as an interaction term between the maximum daily temperature and the average maximum *Hot Day* temperature (32 ° C, 89 ° F). Thus, the interaction term estimates the marginal effect of one temperature degree above 32 ° C. Columns (4) and (5) add a dummy for *warm region* (mean annual county-level temperature falls into the highest temperature quartile for Germany (>10 ° C, 50 ° F)) and *cold region* (mean annual temperature below the lowest temperature quartile (<9 ° C, 48 ° F)) as well as their interactions with *Hot Day*. Column (6) of Panel B excludes all county observations in years without an active weather monitor within a radius of 60 km (37.5 miles) of the county centroid.

Table 5: Testing the Harvesting Hypothesis

Panel A: Mortality	**Daily Data (1)**	**Aggr. at Monthly Level (2)**	**Aggr. at Annual Level (3)**
Hot Day		-0.0004	0.0025***
		(0.0006)	(0.0006)
First Week of Hot Day	0.3331***		
	(0.0187)		
Second Week after Hot Day	-0.1188***		
	(0.0112)		
Third Week after Hot Day	-0.0860***		
	(0.0125)		
Fourth Week after Hot Day	0.1313***		
	(0.38512)		
N	1,516,538	52,248	4,354
Panel B: Hospitalizations			
Hot Day		0.0288***	0.0297**
		(0.0024)	(0.0133)
First Week of Hot Day	0.9306*		
	(0.5511)		
Second Week after Hot Day	-0.2353		
	(0.4544)		
Third Week after Hot Day	-0.2157		
	(0.4138)		
Fourth Week after Hot Day	-0.1374		
	(0.38512)		
N	1,590,454	52,272	4,356

* p<0.1, ** p<0.05, *** p<0.01; standard errors are in parentheses clustered at the county level. Each column in each panel represents one model. The models in the first two columns are as in equation (1) with county fixed effects, week fixed effects, and month-year fixed effects (but without pollution and other weather controls). The models in the last two columns are OLS models with year fixed effects. The variables of interest in column (1) indicate the first, second, third, and fourth week after a *Hot Day*, respectively. The variable of interest in columns (2) and (3) is the number of *Hot Days* per year. Column (2) aggregates the data at the monthly level and column (3) aggregates the data at the annual level.

Table 6: The Monetized Health Effects of One Additional Hot Day

	Hospitalizations				*Mortality*	*Total*	
	Health Care Expenditures (1)	**Lost Labor (2)**	**Lost QALYs (upper bound) (3)**	**Lost QALYs (lower bound) (4)**	**remaining life years (5)**	**(1)-(3) +(5)**	**(1)+(2) +(4)+(5)**
Fixed Effects Model, daily	19,000×€500 =€9.5m	0.5×19,000×€150 =€1.4m	(19,000/365)×€100,000×1.0 =€5.2m	(19,000/365)×€100,000×0.5 =€2.6m	270×1×€100,000 =€27m	∼€43.1m	∼€40.5m
Fixed Effects Model, pollution+ weather controls, daily	8,000×€500 =€4.0m	0.5×8,000×€150 =€0.6m	(8,000/365)×€100,000×1.0 =€2.2m	(8,000/365)×€100,000×0.5 =€1.1m	78×1×€100,000 =€7.8m	∼€14.6m	∼€13.5m
Fixed Effects Model, annual	180×€500 =€90,000	0.5×180×€150 =€13,000	(180/365)×€100,000×1.0 =€50,000	(180/365)×€100,000×0.5 = €25,000	2×30×€100,000 =€6m	∼€6.2m	∼€6.1m

The table shows the health-related costs associated with one *Hot Day*. The first row is based on the model in equation (1) that does not consider additional weather or pollution controls. The models that estimate how many hospital days are triggered by a *Hot Day* are similar to equation (1) but use *Hospital Days* as dependent variable (see Appendix B1). The second row uses the model that considers other weather and pollution controls. These first two approaches are based on daily county-level observations and do not consider potential harvesting effects, i.e., focus on short-term effects. The third row considers harvesting and is based on aggregated annual county-level data (see column (4) of Table 5). Column (1) considers that an average hospital day in Germany is reimbursed with €500. Column (2) considers that the average daily wage in Germany is €150. Columns (3) and (4) assume that 365 hospital days equal a loss of 1 and 0.5 QALYs, respectively. One QALY is evaluated with €100,000. Column (5) assumes that the remaining life expectancy for those who die during heat events is one year for rows one and two (excluding harvesting) and 30 years for row three (including harvesting). We do not discount the monetized health-related loss in welfare. For the approach in the first row, a discount rate of 2.5% would reduce the costs over 80 years from €3.2bn to €1.4bn or €17 per resident. The table does not consider health issues that lead to outpatient treatments. The table also does not consider health-related avoidance behavior costs or adverse health effects due to tornadoes, hurricanes, or floods.

Online Appendix

Appendix A: *Mortality Census*

The first administrative dataset is the *Mortality Census*. It contains the universe of deaths on German territory from 1999 to 2008. This is a restricted access dataset provided by the German Federal Statistical Office (*Statistische Ämter des Bundes und der Länder*). We observe all of the 0.8 million annual deaths. The data contain the following information at the individual admission level:

- age in years
- gender *(binary indicator)*
- county of residence [*between 442 (1999) and 413 (2008) counties*]
- day of death
- primary cause of death *(ICD-10, 3 digit)*

As described in Section F1, we normalize, aggregate, and merge this dataset with the other datasets at the day-county level. As such, we obtain the following descriptive statistics.

Table A1: Mortality Census: Dependent Variables per 100,000 pop. (Daily County-Level, 1999-2008)

Variable	Mean	Std. Dev.	N
Mortality rate	2.9897	1.5229	1,518,000
Cardiovascular mortality rate	1.3839	1.0788	1,518,000
Respiratory mortality rate	0.1918	0.4039	1,518,000
Infectious mortality rate	0.0374	0.1749	1,518,000
Metabolic mortality rate	0.0973	0.2889	1,518,000
Neoplasmic mortality rate	0.7676	0.2889	1,518,000

Source: German Federal Statistical Office (*Statistische Ämter des Bundes und der Länder*). The mortality statistic includes the county of residence and the day of death. The *mortality rate* counts the daily mortality rate per 100,000 population at the county level. German data protection laws prohibit us from reporting min. and max. values.

Appendix B: *Hospital Admission Census*

The second administrative dataset is the *Hospital Admission Census*. It contains the universe of hospital admissions from 1999 to 2008. This is a restricted access dataset provided by the German Federal Statistical Office (*Statistische Ämter des Bundes und der Länder*). We observe more than 17 million annual hospital admissions. The data contain the following information on the individual admission level:

- age in 18 age groups
  (0-2 yrs., 3-5 yrs., 6-9 yrs., 10-14 yrs.,..., 60-64 yrs., 65-75 yrs., >75 yrs.)
- gender *(binary indicator)*
- county of residence [*between 442 (1999) and 413 (2008) counties*]
- day of admission
- length of stay *(censored at 85 days)*
- died in hospital *(binary indicator)*
- primary diagnosis *(ICD-10, 3 digit)*
- surgery needed *(binary indicator)*
- primary hospital department *(43 categories)*
- #hospital beds *(12 categories)*
- hospital location *(federal state level; 16 states)*
- private hospital *(binary indicator)*
- hospital identifier

As described in Section F1, we normalize, aggregate, and merge this dataset with the other datasets at the day-county level. As such, we obtain the following descriptive statistics for the hospital admission data:

Table B1: Hospital Admission Census: Dependent Variables per 100,000 pop. (Daily County-Level, 1999-2008)

Variable	Mean	Std. Dev.	N
All-cause hospitalization rate	57.99	25.71	1,590,454
Hospital days	488.87	267.21	1,590,454
Cardiovascular hospitalization rate	9.1116	4.9216	1,590,454
Cardiovascular hospital days	83.69	55.96	1,590,454
Cardiovascular deaths	0.4532	0.6423	1,590,454
Respiratory hospitalization rate	3.6013	2.5195	1,590,454
Respiratory hospital days	27.93	23.39	1,590,454
Respiratory deaths	0.1557	0.3685	1,590,454
Infectious hospitalization rate	1.3442	1.1759	1,590,454
Infectious hospital days	10.45	13.36	1,590,454
Infectious deaths	0.0509	0.2072	1,590,454
Neoplastic hospitalization rate	6.54	5.1076	1,590,454
Neoplastic hospital days	56.92	49.24	1,590,454
Neoplastic deaths	0.2812	0.5022	1,590,454
Metabolic hospitalization rate	1.6476	1.5454	1,590,454
Metabolic hospital days	15.48	18.39	1,590,454
Metabolic deaths	0.02534	0.1489	1,590,454

Source: German Federal Statistical Office (*Statistische Ämter des Bundes und der Länder*). The *German Hospital Admission Census* includes the county of residence and the day when the patient was hospitalized. The *hospitalization rate* counts the daily incidence of hospitalizations per 100,000 population on the county level. *Hospital days* is the sum of all hospital days that were triggered on a given day, i.e., it is the product of the hospitalization rate and the length of stay. *Deaths* count the number of hospital deaths per 100,000 population on the county level. The reference point is always the day when the patient was hospitalized. The patient died sometime after being admitted, but not necessarily on the day of admission. German data protection laws prohibit us from reporting min. and max. values.

Appendix C: Official Weather Data

The third register dataset contains daily weather measures from up to 1,044 ambient weather stations. The data are provided by the German Meteorological Service (*Deutscher Wetterdienst (DWD)*). It covers the years from 1999 to 2008.

Table C1: Weather Data (Daily County-Level, 1999-2008)

Variable	Mean	Std. Dev.	Min.	Max.	N
A. Raw Daily Measures					
Average temperature in °C (*2 m (6'7") above ground*)	9.5573	7.3047	-19	30.6	1,590,454
Minimum temperature in °C (*2 m (6'7") above ground*)	5.4671	6.4965	-25.01	23.8	1,590,454
Maximum temperature in °C (*2 m (6'7") above ground*)	13.8912	8.5608	-14.1	39.07	1,590,454
Total hours of sunshine	4.6252	4.2373	0	16.7	1,590,454
Precipitation level in mm	2.2246	4.2154	0	144.98	1,590,454
Average humidity in %	78.3161	11.4307	10	100	1,590,454
Average cloud coverage in %	5.3128	2.1534	0	8.23	1,590,454
Average storm force	3.6065	2.0856	0	26.3	1,590,454
Max. wind speed in km/hr.	10.4964	4.4462	0	54	1,590,454
Vapor pressure in hPA	9.8876	3.9981	0.5	25.9	1,590,454
Min. air pressure in hPA (*5 cm (2 inches) above ground*)	3.8456	6.5299	-29.01	22	1,590,454
B. Extreme Temperature Indicators					
Economic Models					
Hot Day (max temp. >30 °C (86 °F))	0.0198	0.1392	0	1	1,590,454
Cold Day (min temp. <-10 °C (14 °F))	0.0124	0.1108	0	1	1,590,454
Epidemiological Models					
Hot Day II (mean temp. > percentile 97.5)	0.0249	0.1558	0	1	1,590,454
Cold Day II (mean temp. < percentile 2.5)	0.0249	0.1558	0	1	1,590,454

Source: German Meteorological Service (*Deutscher Wetterdienst (DWD)*). The information was recorded on a daily basis by up to 1,044 ambient weather monitors (see Figure 1). The number of weather stations varies from year to year. The weather indicators displayed cover the years 1999 to 2008. As described in Section F1, all point measures from the stations are interpolated into the county space by means of deterministic inverse distance weighting (IDW). Level of analysis is the day×county level. Hence, with exactly 400 counties in each year, we would obtain $400 \times 365 \times 10 = 1,460,000$ observations. However, the number of counties varies across years from 442 (1999) to 413 (2008). Vapor and air pressure are measured in hectopascal (hPa).

As described in Section F1, in a first step, we interpolate the point measure into the county space. Then we merge the weather data with the other data at the day-county level.

Panel A of Table C1 shows the descriptive statistics for the interpolated weather measures

as collected by the DWD. The mean daily air temperature is 10 ° C (49 ° F), averaged over the whole time period and over all counties. Note the rich variation in the average daily temperatures which ranges from -19 ° C (-2 ° F) to 31 ° C (87 ° F). Equally rich is the variation of the minimum and maximum temperatures, hours of sunshine and other weather measures.

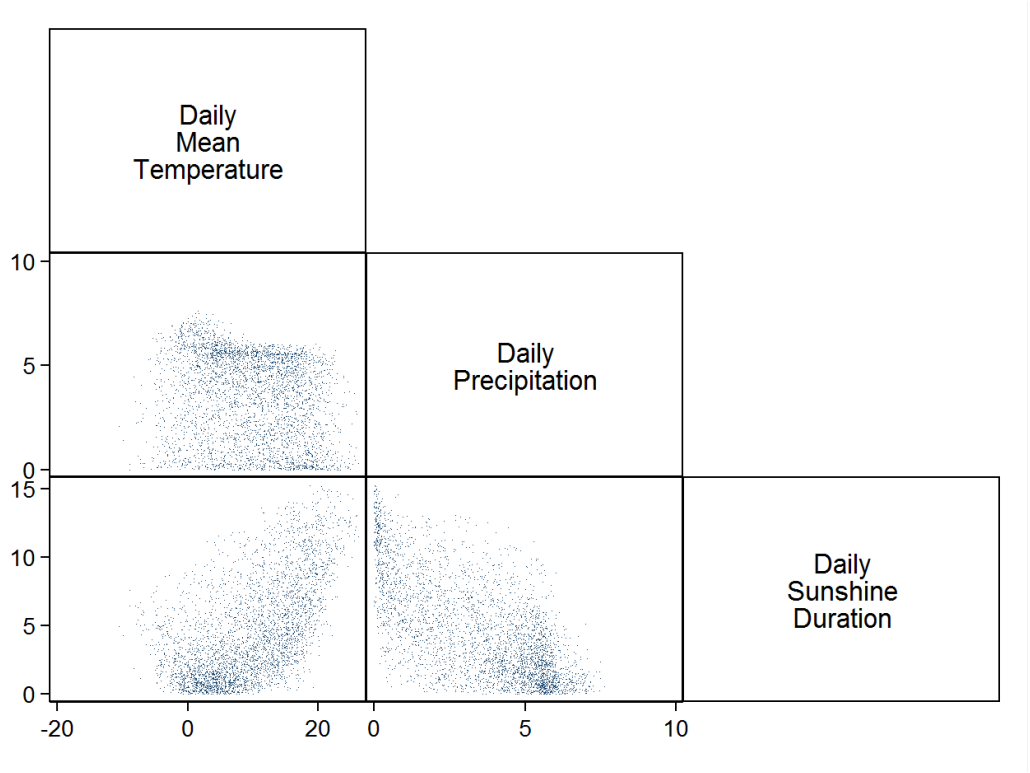
The raw weather data that we use for interpolation have missing information on some weather indicators. Table C2 displays the frequency of missing observations by year (at the monitor-day level). In general, missing observations are very rare: the overall frequency is 0.78%; the year 2001 shows the highest frequency of missing information with 1.0% of observations missing.

Table C2: Missing Weather Information

	Mean	SD	*N*
1999	0.009	0.093	201,233
2000	0.018	0.132	202,486
2001	0.010	0.099	203,615
2002	0.008	0.090	196,632
2003	0.007	0.086	188,879
2004	0.007	0.081	181,321
2005	0.007	0.085	187,114
2006	0.004	0.065	188,128
2007	0.004	0.059	184,415
2008	0.003	0.051	185,785
Total	0.008	0.088	1,919,608

Figure C7 displays a scatter matrix which shows, illustratively, the associations between some raw weather measures. Not surprisingly, one finds a strong positive association between the hours of sunshine and the temperature, as well as a strong negative association between the hours of sunshine and the precipitation level.

Figure C7: Scatter Matrix Illustrating Associations Between Temperature, Sunshine, and Precipitation



Note: The scales of the x and y-axes correspond to the scales of the plotted variables of interest, i.e., temperature, precipitation, and hours of sunshine.

Appendix D: Official Pollution Data

The fourth register dataset contains daily pollution measures from up to 1,314 ambient monitors. The data are provided by the German Federal Environmental Office (*Umweltbundesamt (UBA)*). It covers the years from 1999 to 2008. As described in Section F1, in a first step, we interpolate the point measure into the county space via IDW. Then we merge the pollution dataset with the other datasets at the day-county level. We make use of the following pollutants as additional control variables in extended model specifications:

Table D1: Pollution Data (Daily County-Level, 1999-2008)

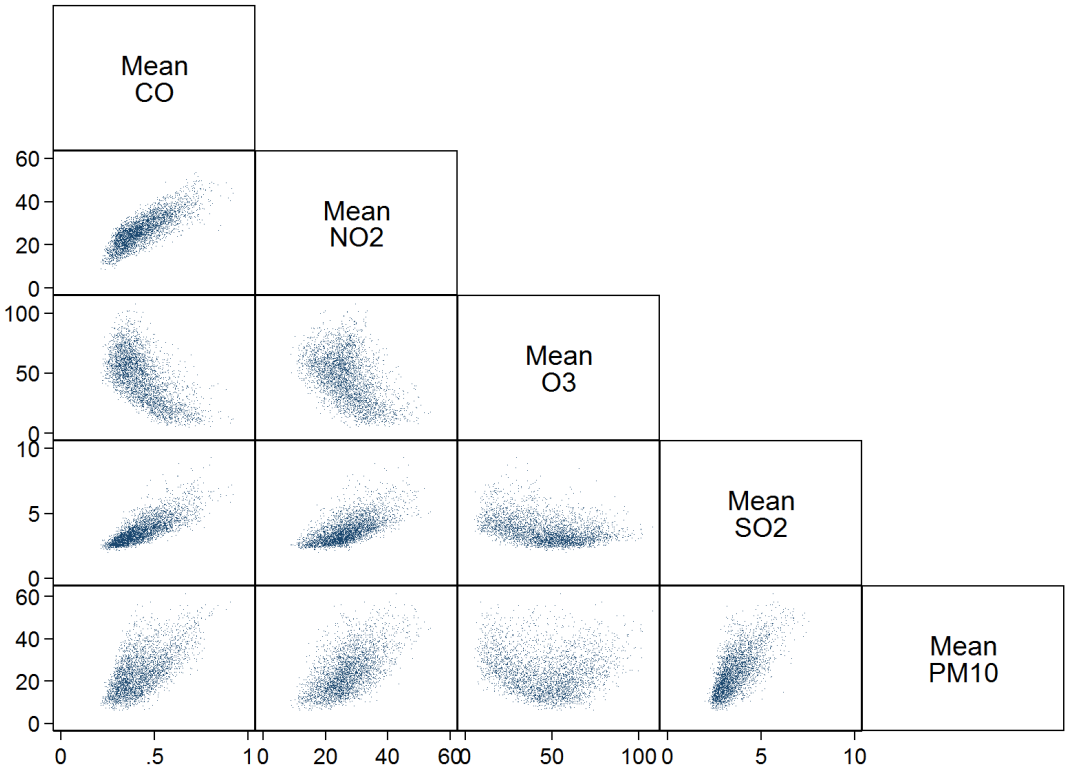
Variable	Mean	Std. Dev.	Min.	Max.	N
Average CO in ppm	0.4342	0.1794	0.0023	1.3083	1,594,154
Average O_3 in $\mu g/m^3$	45.9786	22.0423	0.8612	135.79	1,594,154
Average NO_2 in $\mu g/m^3$	26.8907	10.6284	0.0278	80.3095	1,594,154
Average SO_2 in $\mu g/m^3$	3.7256	1.6115	0.0654	12.5435	1,594,154
Average PM_{10} in $\mu g/m^3$	24.3097	11.4625	2.0625	64.625	1,432,822

Source: German Federal Environmental Office (*Umweltbundesamt (UBA)*). The information was recorded on a daily basis by up to 1,317 ambient pollution monitors (see Figure 1). The number of counties and weather stations vary from year to year. The pollution measures displayed cover the years 1999 to 2008. As described in Section F1, all point measures from the stations are interpolated into the county space by means of deterministic inverse distance weighting (IDW). Level of analysis is the day×county level. Hence, with exactly 400 counties in each year, we would obtain $400 \times 365 \times 10 = 1,460,000$ observations. However, the number of counties varies across years from 442 (1999) to 413 (2008). CO stands for "carbon monoxide" and ppm for "parts per million." NO_2 stands for "nitrogen dioxide," O_3 stands for "ozone," SO_2 stands for "sulphur dioxide," and PM_{10} stands for "particular matter." $\mu g/m^3$ stands for micrograms per cubic meter of air.

D1 Associations Between All 5 Pollutants

Figure D8 shows the associations between all five air pollutants discussed above. NO_2 is positively correlated with SO_2 and PM_{10}, but negatively correlated with O_3. The same is true for CO. O_3 exhibits only very noisy and weak associations with SO_2 and PM_{10}. However, SO_2 and PM_{10} show a strong and positive association. Several recent papers in environmental epidemiology and biostatistics have developed innovative approaches to model the joint and partial health effects of several pollutants (Katsouyanni et al., 2001; Dominici et al., 2010; Bobb et al., 2013; Zhang et al., 2017).

Figure D8: Scatter Matrix Illustrating Associations Between Pollutants



Appendix E: Annual Socio-Economic County-Level Data

Finally, this paper makes use of yearly county-level data provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (2012) (*Bundesinstitut für Bau-, Stadt- und Raumforschung*) in their INKAR (*Indicators and Maps on Spatial Development*) database. The data vary by year.[16] To normalize the dependent variables and calculate hospitalization and death rates, we use annual county-level population counts. In the models, we control for the *unemployment rate* and *GDP per capita*. Supply-side constraints are captured by the *# hospitals per county*, *hospital beds per 10,000 pop.* and *physicians per 10,000 pop.*

Table E1: Descriptive Statistics Other (County-Level, 1999-2008, Annual)

Variable	Mean	Std. Dev.	Min.	Max.	N
Unemployment rate	10.47	5.28	1.6	29.3	4,354
GDP per capita	24971	10146	11,282	86,728	4,354
# hospitals per county	4.84	5.49	0	76	4,354
Hospital beds per 10,000 pop.	1211.19	1593.88	0	24,170	4,354
Physicians per 10,000 pop.	152.72	52.59	69	394	4,354
County population	189,450	219,753	34,525	3,431,675	4,354

Source: Federal Institute for Research on Building, Urban Affairs and Spatial Development (2012). The data vary on the county-year level. In addition, in contrast to the register databases in Appendices A and B, the INKAR data refers to the county codes and boundaries as of January 1, 2012. Since various county reforms were implemented between 1999 and 2008, we imputed information for pre-reform counties with post-reform data (if possible). For example, if counties A and B simply merged to county C and only the GDP per capita for county C was available, we imputed the GDP per capita for A and B using the population information on A and B which is available for all years and counties. If, as another example, data was surveyed in every other year, we took the mean value of t_0 and t_2 to impute information for t_1.

[16] The hospitalization and mortality data contain the county of residence according to the county codes and boundaries of the specific year. In contrast, the INKAR database contains all information according to the county codes and boundaries as of January 1, 2012. From 1999 to 2008, various county reforms, mostly mergers between two counties, led to changes in the county codes and boundaries. Consequently, the number of counties varies across years from 442 (1999) to 413 (2008). For counties with county reforms, we imputed pre-reform values using the post-reform boundary data as of January 1, 2012. In addition to reforms, not all information listed above have been collected in every single calendar year. We imputed missing values for these cases. See notes to Table E1 for more details.

Appendix F: Interpolation of Weather and Pollution Measures

F1 Interpolation of Weather and Pollution Measures

To obtain the working datasets, we *(i)* interpolate the point measures of the weather and pollution monitors into the county space, *(ii)* aggregate and normalize all information at the daily county level, and *(iii)* merge the register datasets with the pollution, weather, and the socioeconomic dataset (see Appendix E) at the day-county level. Assuming that the number of counties is time-invariant and 400, we would obtain $400 \times 365 \times 10 = 1,460,000$ rows, each representing one county on a given day.

Hanigan et al. (2006) discuss and compare different approaches to calculating estimates of population exposure to daily weather and pollution conditions from monitors. Here, we rely on inverse distance weighting (IDW) within a certain radius. We first determine the centroid of each county. Then the distance between each county and each monitor is calculated. In the final step, we calculate the weighted average for each county where weights are based on the inverse distance to all monitors within a radius of 60 km (37.5 miles) of the county centroid. Thus, by denoting δ_{ij}—the distance between a location i (a county centroid) and a monitor j—one can define the weighting scheme as:

$$w_{ijd} = \begin{cases} \frac{1}{\delta_{ij}} & \text{if } i \neq j \text{ and } \delta_{ij} < 60 \\ 1 & \text{if } \delta_{iM_{id}} > 60 \text{ and } j = M_{id} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where M_{id} denotes the nearest station outside location i. Thus, whenever there are no stations within a radius of 60 kilometers, the measure from the nearest station outside this radius is used. Henceforth, we call this interpolation approach simply Inverse Distance Weighting. This section discusses econometric and methodological issues related to the IDW interpolation method.

Twenty percent of all counties over all years have no weather monitor within a radius of 60 km. In 26% of all observations, there is one monitor, in 27% of all observations two monitors, and in another 27% more than two monitors. In a robustness check, we omitted all counties in years when there wasn't an active weather monitor within a radius of 60 km (37.5 miles); our results were very robust (Panel B, column (7) of Table 4).

Below we present the results of specification tests for measurement errors and concludes that

the IDW method works well for heat events. However, our tests suggest some caution is required when it comes to pollution, which seems to be measured with more noise. Because we only use the measures for the five pollutants to control for contemporaneous pollution conditions in our conditional models, it is no major methodological threat to our main findings.

F2 Cross-Validation of Interpolated Measures

When mapping ambient monitor point measures into space, one has to deal with measurement errors. It is known that classical measurement error attenuates parameter estimates. In case of non-classical measurement error, the direction of the bias is unclear. Moreover, measurement error in the dependent variables inflates standard errors (Chen et al., 2011).

To assess the measurement error that is introduced via the IDW method, following Currie and Neidell (2005), we perform the following (indirect) test: For each weather and pollution *monitor* (*not* county centroid), we calculate the IDW value using the weighting scheme in equation (4). The crucial point is that the weighting scheme attaches weight 0 to the own station.[17] Thus, for each ambient monitor and all weather and pollution measures from that monitor Z_d, we calculate a cross-validated $\tilde{Z}_d = Z_d \Omega_d$; where Ω_d is the symmetric matrix of weights for day d with elements $\omega_{ijd} = w_{ijd} / \sum_k w_{ikd}$. In other words, we predict the values of each monitor using all surrounding monitors and the IDW interpolation method. Then, we assess the accuracy of the IDW interpolation by calculating Pearson's correlation coefficient for the variables Z and $\tilde{Z}$. The results of this exercise are in column (1) of Table F1.

Table F1 illustrates that *(a)* the IDW method dominates the simpler NN weighting scheme: The NN method delivers only better accuracy for air pressure. Besides, it becomes clear that *(b)* our IDW interpolation algorithm delivers a very acceptable accuracy with correlation coefficients ranging up to 0.98 for the mean temperatures. Note that this paper particularly relies on minimum, mean, and maximum temperature measurements, all of which deliver excellent accuracy results with correlation values ranging above 0.95 (column (1) of Table F1). This means that we are able to predict with 95% accuracy the temperature measured by monitor X using our IDW method and all surrounding monitors. However, the correlation values for pollution range between 0.4 and 0.8 suggesting a substantial degree of measurement error. Recall, though, that we only use the pollution measures as additional controls in some model specifications.

[17] When using the county centroid in the IDW interpolation of point measures into county space, the closest monitor obviously gets the largest weight.

One concern with this interpolation test is that a seemingly high degree of accuracy might be driven by time trends and seasonal variation in the variables. Thus, we calculate alternative accuracy correlation measures that are based on transformed versions of Z. These are first nonparametrically adjusted for individual day effects. As seen, the correlation coefficients in columns (3) and (4) of Table F1 drop somewhat, but still show that there is a considerable correlation between imputed and actual values. For the temperature measures, the time trend and season-adjusted correlation values all lie around 0.7. Controlling for 3,650 day fixed effects is very conservative and likely removes "too much" variation from the data because one cannot disentangle the "true" correlation between monitors and climatic measures from day fixed effects. By removing the daily mean, one obviously also removes part the non time-trend correlation. Note that the results are robust to considering individual years instead of the entire pooled sample (results available upon request).

Table F1: Cross-Validation of IDW Interpolation

Variable	Raw Correlation		Time and Season-Adjusted Correlation	
	IDW Method	NN Method	IDW Method	NN Method
Temperature	0.981	0.972	0.733	0.661
Min Temperature	0.968	0.953	0.713	0.637
Max Temperature	0.977	0.966	0.659	0.587
Precipitation	0.788	0.740	0.688	0.634
Sunshine	0.934	0.922	0.556	0.535
Cloud	0.874	0.821	0.585	0.508
Humidity	0.876	0.826	0.643	0.566
Vapor Pressure	0.979	0.970	0.735	0.678
Air Pressure	0.549	0.579	0.239	0.257
Wind Speed	0.497	0.478	0.219	0.156
CO Mean	0.477	0.363	0.149	0.082
NO2 Mean	0.562	0.450	0.407	0.321
O3 Mean	0.862	0.797	0.435	0.362
SO2 Mean	0.616	0.532	0.306	0.265
PM10 Mean	0.837	0.814	0.239	0.212

Source: German Meteorological Service (*Deutscher Wetterdienst (DWD)*) and German Federal Environmental Office (*Umweltbundesamt (UBA)*). The table shows the cross-validation of the weather and pollution interpolation as discussed in Section F1. The underlying data stems from up to 1,044 ambient weather monitors and up to 1,317 ambient pollution monitors between 1999 and 2008. Columns (1) and (3) display the Pearson's Correlation Coefficient between the original values of monitor X and its predicted values solely using all surrounding monitors and Inverse Distance Weighting (IDW). Columns (2) and (4), in contrast, simply use the Nearest Neighbor (NN) method and thus predict values of monitor X with the measurement of its nearest neighbor monitor. Columns (3) and (4) are based on values that have been non-parametrically adjusted for all 3,650 day effects, i.e., the nationwide daily mean of a specific measure was first removed from all monitor measurements. This exercise removes time trends, but also the "true" correlation in measurements between monitors and has to be regarded as a very conservative test.

Table F2 show results of a similar test for the generated extreme weather indicators and

confirms the results of Table F1. Basically, one finds that the overall share of correctly predicted heat and cold indicator values is above 99%, as is the share of correctly predicted zeros. Since there is only a small percentage of extreme temperature events, "false positives" have a larger impact on estimates than "false negatives." Thus, it is reassuring to see that *(i)* IDW clearly outperforms NN, and *(ii)* the share of false positives is low and less than 20% in the case of heat.

Finally, we calculate the Reliability Ratio (RR) α that indicates the magnitude of measurement errors and thus the attenuation bias (Hyslop and Imbens, 2001):

$$\alpha = \frac{\text{Cov}\left(Z, \tilde{Z}\right)}{\text{Var}\left(\tilde{Z}\right)} \tag{5}$$

Table F2: Share of Correctly Predicted Extreme Weather Indicators

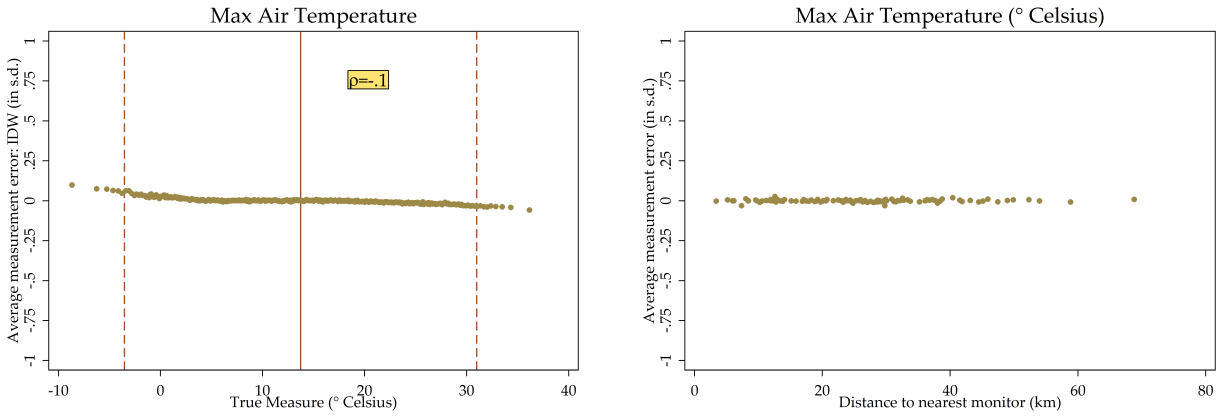
Panel A: IDW	Overall Correct Predicted	Positives Correct Predicted	Zeros Correct Predicted	Reliability Ratio
Hot Day	0.9904	0.8133	0.9939	0.8071
Cold Day	0.9927	0.7680	0.9954	0.7634
Panel B: NN	Overall Correct Predicted	Positives Correct Predicted	Zeros Correct Predicted	Reliability Ratio
Hot Day	0.9881	0.7286	0.9937	0.72233
Cold Day	0.9908	0.6699	0.9951	0.6651

Source: German Meteorological Service (*Deutscher Wetterdienst (DWD)*). The underlying data stems from up to 1,044 ambient weather monitors between 1999 and 2008. Panel A tests the predictive quality of the Inverse Distance Weighting (IDW) interpolation method into the county space and Panel B the Nearest Neighbor (NN) method. All numbers are shares of predicted relative to actual values. The predicted value for monitor X are calculated using solely all surrounding monitors and assuming that monitor X is non-existent. Column (1) reports the overall share of correctly predicted positive or negative extreme weather indicator values. Column (2) reports the share χ of correctly predicted positives and column (3) the share δ of correctly predicted zero values. Consequently, 1-χ represent false positives and 1-δ false negatives. Column (4) shows the Reliability Ratio (RR) α which indicates the ratio between OLS and IV estimates and thus assesses the size of the potential attenuation bias (Hyslop and Imbens, 2001).

In a bivariate regression, the RR measures the attenuation bias and can thus be used to adjust estimates. In a multivariate setting, the issue is less straightforward. Under the assumption that covariates are uncorrelated with the measurement error, a specific RR—which is typically lower than the RR for the bivariate case—can be derived. However, as we include covariates which are prone to measurement error themselves, it is not possible to draw general conclusions about the size of the bias (Maddala, 1977). Nevertheless, it is reassuring that the RR is relatively high and lies around 0.8 for the most important indicators. Because the RR assumes classical measurement error, we follow Knittel et al. (2016) and carry out the following test: First,

we generate an 'error' variable for each monitor and weather measure by *(i)* predicting the weather measure using all surrounding ambient monitors, similar to above. Then we *(ii)* take the difference between the true ambient monitor measure and the predicted measure and label it 'error.' Figure F9 shows the mean measurement error in standard deviations by *(a)* the maximum daily temperature recorded, and *(b)* the distance to the next ambient monitor.

Figure F9: Temperature Measurement Error by (a) Baseline Level and (b) Distance to Next Monitor



Note: Figure F9a shows the average measurement error of the IDW method by baseline temperature level. Figure F9b shows the average measurement error of the IDW method by the distance to the next ambient monitor.

As seen, both scatterplots provide evidence for a very small degree of average measurement errors. The dots are tightly lined up around the zero line on the y-axis. Not unexpectedly, in Figure F9a, the plotted dots slightly deviate from the zero line for very high or low maximum temperatures but the deviations are very small and below 0.1 of a standard deviation. In Figure F9b, showing the measurement error as a function of the distance to the next monitor, the curve is straight flat. Overall, this additional test makes us very confident that the degree of measurement error is likely to be small and, if existent, likely to be of classical form.

As a final conceptual point, we would like to emphasize that the issue of introducing measurement error when extrapolating point measures into space is methodologically not fundamentally different from the issue of unknown individual exposure to weather and pollution conditions. We approximate the individual level exposure to weather and pollution on a given day by taking inverse distance weighted averages of the daily measures of the next monitors. Even if we knew the exact ambient weather and pollution conditions at the exact locations of residence of all German residents, we would still *(i)* have to take daily averages in ambient conditions, *(ii)* lack knowledge about the exact length, place, and time of the day spent outdoors by the individuals, and thus *(iii)* deal with exposure-related measurement error of unknown form.