Racial disparities in temperature related deaths in the U.S. between 1993 and 2005

Risto Conte Keivabu[1], Ugofilippo Basellini[2], and Emilio Zagheni[2]

[1]*European University Institute, Florence, Italy*
[2]*Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany*

November 26, 2021

Abstract

Extreme temperatures increase mortality in the population, but some individuals are more vulnerable than others. Here, we investigate how extreme temperatures affect mortality and how race stratifies this relationship in the U.S.. We use the Berkeley Unified Numident Mortality Database (BUNMD) dataset on approximately 15 million deaths in more than 3 thousand U.S. counties from 1993 to 2005 for individuals aged above 65. Also, we aggregated this data with meteorological information provided by Gridmet. We employed Poisson regression models with fixed effects to estimate how extreme temperatures affect mortality. Moreover, we added an interaction between temperature and race to test for a stratified effect of temperature on mortality. Results in the pooled sample show an increase in deaths with warm and hot days and especially cold days. Looking at differences by race, we find a comparable effect of cold on mortality but the effect size of hot days on mortality is largest for Blacks. Finally, we simulate the number of additional deaths that would have occurred between 1993 and 2005 if temperatures were to increase to those expected in 2051 based on the RCP4.5 emissions scenario. Our findings highlight great racial disparities in the effect of temperature change on mortality.

1 Introduction

Climate change is predicted to enhance the occurrence of extreme temperatures posing health risks for the U.S. population, especially for the most vulnerable. One of the ultimate metrics for assessing the impact of climate change is the number of deaths and years of life lost. While the relative importance of economic costs may vary across societies, the value of life is essentially universal and counting excess deaths offers a key metric for assessing the burden of climate change. There is an important and growing body of literature on the impact of extreme temperatures on mortality (Barreca et al., 2016; Hovdahl, 2020; Zanobetti and Schwartz, 2008; Nordio et al., 2015; Weinberger et al., 2020) including on how different demographic groups are unequally affected by changes in temperature (Son et al., 2019). Notably, some racial groups might be more at risk than others.

Racial disparities in heat related mortality are determined by differences in exposure and vulnerability (Stafoggia et al., 2006; Liu et al., 2021). Disparities in exposure relate to poorer housing conditions and the higher likelihood of living in warmer areas of a city for minority race groups (Gronlund, 2014). Disparities in vulnerability are expected as some race groups could have preexisting health conditions that can be aggravated by extreme temperatures. For example, the black population is more likely to suffer from obesity that makes heat less bearable (Petersen et al., 2019).

Despite the expected racial differences in heat related deaths, there is currently only limited or suggestive evidence on this matter in the U.S., as highlighted in a recent literature review (Son et al., 2019). Studies on racial disparities in the U.S. have shown a higher risk for the black population in North Carolina, South Carolina, Georgia from 2007 to 2011 (Lee et al., 2016) and in 4 U.S. cities (O'Neill, Zanobetti, Schwartz, 2005). However, no racial differences in temperature related morbidity have been found in 9 counties in California from 1999 to 2005 (Green et al., 2010). Moreover, there is some evidence of a lower risk of heat related mortality for hispanics and asians (Hansen et al., 2013; Noe et al., 2012).

In this paper we go beyond the state of the art by leveraging new and untapped linked datasets, the Berkeley Unified Numident Mortality Database (BUNMD) combined with meteorological information provided by Gridmet (Abatzoglou, 2013). These data enable us to shed new light on longstanding questions for which there is only suggestive evidence. Our research is guided by two main questions: how does extreme temperature affect mortality in the U.S.? How does race stratify the relationship between temperature and mortality?

In the following sections, we describe the dataset, the variables and the empirical strategy that we employ in our study, and we present some preliminary results.

2 Methodology

2.1 Data

In this study we use three main datasets: the Berkeley Unified Numident Mortality Database (BUNMD), Population Estimates from the National Cancer Institute (2021), and the Gridmet: Data on a 4-km grid for North America (Abatzoglou, 2013).

The BUNMD is a unique dataset comprising 50 million individuals that were present in the 1940 Census and that died from 1988 to 2005 (Goldstein and Breen, 2020). The dataset provides the ZIP code of residence at the time of death permitting to accurately identify the geographical location of an individual. However, to link death counts to population estimates provided at the county level, we connect the zip codes to county identifiers using information from the Housing & Urban Development Office (HUD)[1]. In our datasets, we include only individuals aged above 65 for two main reasons. First, the elderly are the most vulnerable to extreme temperatures (Achebak et al., 2019); (Cheshire, 2016); (Oudin Åström et al., 2011). Secondly, the BUNMD covers over a substantial proportion of all elderly deaths (aged 65+) occurred in the US during this time period (Goldstein and Breen, 2020), while coverage at younger ages is more limited. As such, including younger individuals in the analysis could bias our estimates. To further reduce bias determined by missing values we limit the analysis to the years 1993 to 2005 and we aggregate this information to generate our main dependent variable composed of month-by-year-by-county death counts (Goldstein and Breen, 2020). Furthermore, in order to reduce the discrepancy of the BUNMD with respect to the observed deaths in the U.S., we adjust death counts and exposures so that they match the observed ones in the (HMD, 2021) for each year, sex and age-group.

We link the death counts to population estimates provided by the U.S. (N.C.I, 2021). Population estimates are available at the racial and county levels for each year analysed. Moreover, we compute monthly mortality rates by disaggregating annual population estimates into monthly ones. This is achieved by dividing the race-, county- and year- specific populations by 12 (the number of months), and by using linear interpolation between consecutive years.

[1]We used the identifiers provided by the HUD as some zip codes are overlapping between counties. We allocated zip codes to a county if more than 50% of the residential addresses were part of that county. The zip-code and county identifiers are available here: https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook

Moreover, we combine the month-by-year-by-county information on death counts and population estimates with precise meteorological data provided by Gridmet (Abatzoglou, 2013). The Gridmet dataset provides daily information on several climatic variables such as maximum temperature, minimum temperature, precipitation, wind velocity and shortwave radiation at a 4x4km resolution. We computed the average minimum and maximum temperature of the grid cells falling within each county boundaries and used it to compute: i) the daily average temperature (as the mean between minimum and maximum temperatures), and ii) the monthly temperature categories counting the number of days in 7 bins, computed based on percentiles of the county temperature distribution, respectively: days below and equal to the 5th percentile; from the 5th to the 10th percentile; from the 10th to the 25th percentile; above the 25th and below the 75th percentile; from the 75th to the 90th percentile; from the 90th to the 95th percentile; above the 95th percentile.

Finally, we retrieve information on the monthly average level of the air pollutant particulate matter 2.5 (PM2.5) at the county level from the Atmospheric Composition Analysis Group (ACAG) dataset (Hammer et al., 2020).

2.2 Methods

Let $c = 1, \ldots, 3096$ denote the U.S. counties, $r =$ W,B,O denote race, $s =$ F,M denote sex, and $a = 1, 2, 3$ denote age groups (65–74, 75–84, 85+). To ease notation, let j denote a given combination of county, race, sex and age group. Furthermore, let $t = 1, \ldots, 156$ denote the time observations in the dataset, corresponding to twelve monthly observations from year 1993 to 2005. Let us also denote $\tilde{t} = 1, \ldots, 12$ the month corresponding to time t. We assume that deaths Y_{jt} in group j at time t are realizations of a Poisson distribution with expected value equal to the product of exposures E_{jt} and force of mortality μ_{jt} (Brillinger, 1986).

In our analysis, we model Poisson death counts in the standard Generalized Linear Model (GLM) framework using a log-link function and exposures as an offset. In particular, the expected value of the Poisson distribution $\mathbb{E}(Y_{jt})$ can be described as:

$$\ln\left[\mathbb{E}\left(Y_{jt}\right)\right] = \ln\left(E_{jt}\right) + \sum_{k=1}^{9} \theta_k \text{TEMP}_{jt}^{k} \times R_j + \boldsymbol{X}_j \boldsymbol{\beta}_j + \alpha_t + \gamma_{j\tilde{t}} \tag{1}$$

where $\ln(E_{jt})$ is the offset term, TEMP_{jt}^{k} denotes the number of days in the k-th temperature bin ($k = 1, \ldots, 9$, see Section 2.1 for their description) to which individuals in group j were exposed at time t. The temperature variable is interacted with the race variable R_j, which denotes the race of group j. Days in the comfort zone are defined as those days between the 25th and 75th percentile. This bin is not introduced in the model, so that it becomes the baseline to which other bins are compared to. Moreover, we add a 1×3 matrix of time-unvarying covariates $\boldsymbol{X}_j$ with associated coefficients $\boldsymbol{\beta}_j$ that include: i) the monthly average level of the air pollutant particulate matter 2.5 (PM2.5) at the county level, ii) the sex and iii) the age group of group j. Furthermore, we include month-by-year and county-by-month fixed effects to capture specific yearly and seasonal variations in mortality in each county that could affect the outcome variable (as in, for example, Barreca et al., 2016). Specifically, α_t captures the month-by-year fixed effects for each time t in the dataset, and $\gamma_{j\tilde{t}}$ is the county-by-month fixed effect, corresponding to the county in group j and the month $\tilde{t}$. Finally, we cluster standard errors at the county and race group level allowing them to correlate over time within the units of analysis (Cameron & Miller, 2015).

3 Results

First, we present results based on Equation (1) but without the interaction term in Figure 1 to visualize how temperatures affect mortality in the total U.S. population. We observe positive coefficients for the temperature variables, corresponding to an increase in mortality rates for all temperature bins that depart from the comfort zone set at temperatures between the 25th and 75th percentile. Interestingly, we find stronger effects for exposure to cold days compared to hot days. Respectively, we observe an increase in the mortality rate with exposure to days with temperatures above or equal to the 5th percentile of 3 per 1,000 and with days with temperatures above the 95th percentile of about 1.2 per 1,000. Nevertheless, these results might hide heterogeneous effects based on race.

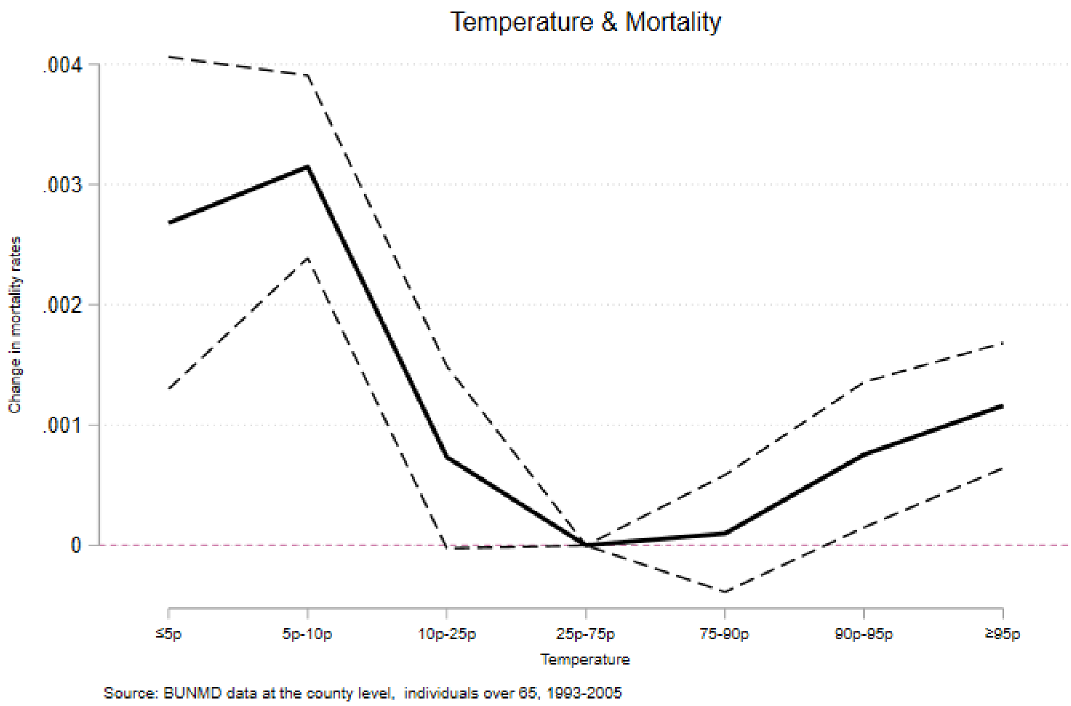

Figure 1. Temperature related deaths and Mortality. *Notes*: Estimates are obtained by estimating equation (1) without the interaction with race. Standard errors clustered at the county level. 95% confidence intervals.

In Figure 2, we present results of the interaction model presented in Equation (1). Here, we observe cold temperatures to increase mortality for Whites, Blacks and Other races. More precisely, we observe an increase in the mortality rate by 7 per 1,000 with exposure to temperatures below the 5th percentile in Whites and Blacks, but not for the Other race group. Nevertheless, we observe comparable effect sizes for all race groups with exposure to days with temperatures between the 5th and 10th percentile. Respectively, we observe an increase in the mortality rate of about 7 per 1,000 for all race groups. Focusing on the results for heat we denote larger race disparities. Temperatures above the 95th percentile increase the mortality rate for Blacks by 8 per 1,000 but for Whites by 2 per 1,000 and Other race groups by 1 per 1,000.

Table 1 shows the number of excess deaths and excess mortality rate that would result from 1993 to 2005 if temperatures were to change to levels predicted for 2051 (data retrieved based on the RCP4.5 emissions scenario and provided by the Multivariate Adaptive Constructed Analogs dataset - MACA) (Taylor K.E. and Meehl, 2012; Abatzoglou and T.J, 2012). This counterfactual analysis shows important racial disparities in the effect of temperature changes on mortality.

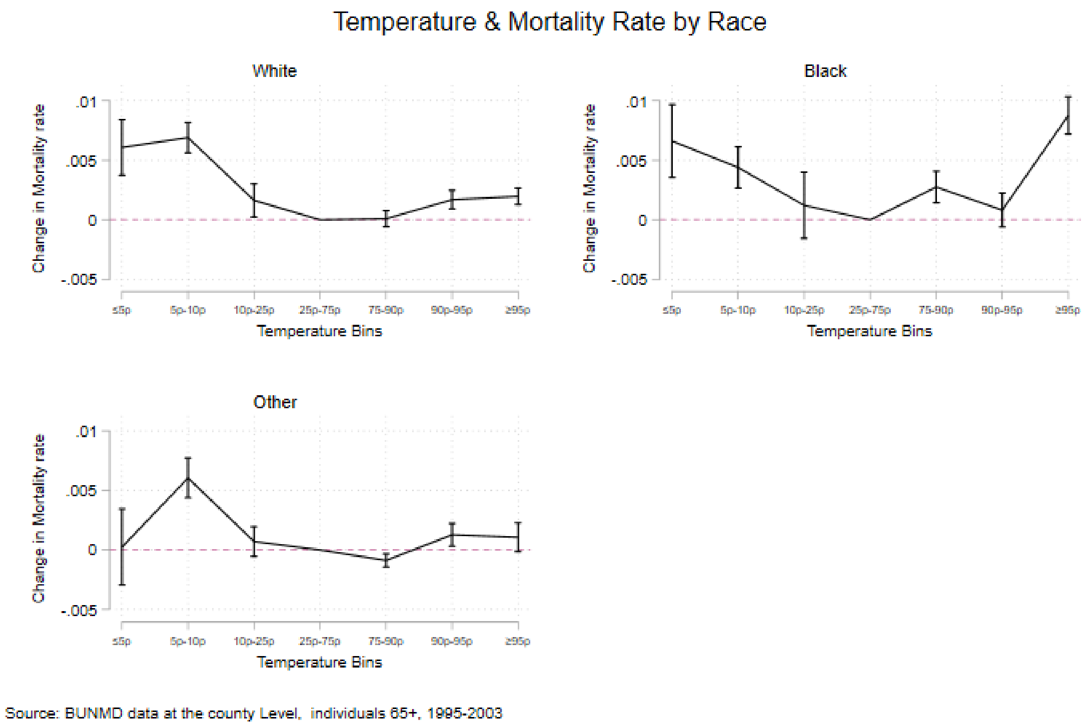

Figure 2. Temperature related deaths and interaction with race. *Notes*: Estimates are obtained by estimating equation (1). Standard errors clustered at the county and race group level. 95% confidence intervals.

In particular, while for Whites and Other race group the increase in temperature would not largely increase the mortality rate, for Blacks we observe a substantively larger increase in risk.

	Observed deaths	**Counterfactual deaths**	**Excess deaths**	**Total Exposure**	**Excess rate (per 100,000)**
White	19,797,674	19,802,521	4,847	378,012,978	1.28
Black	2,148,544	2,162,227	13,682	36,557,640	37.4
Other	819,408	821,205	1,797	33,631,825	5.34
Total	22,765,627	22,785,953	20,326	448,202,443	4.5

Table 1. Table 1. Excess deaths and mortality rate if temperature were to raise to 2051 level based on RCP4.5 *Notes*: Estimates are obtained by predicting the number of deaths based on the estimates retrieved in the analysis of Equation (1) using daily data on temperatures in 2051 based on the RCP4.5 emission scenario.

References

Abatzoglou, J. and T.J, B. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*.

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1):121–131.

Achebak, H., Devolder, D., and Ballester, J. (2019). Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: a national time-series analysis. *The Lancet Planetary Health*, 3(7):e297–e306.

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

Brillinger, D. R. (1986). A biometrics invited paper with discussion: The natural variability of vital rates and associated statistics. *Biometrics*, 42(4):693–734.

Cheshire, W. P. (2016). Thermoregulatory disorders and illness related to heat and cold stress. *Autonomic Neuroscience: Basic and Clinical*, 196:91–104. Publisher: Elsevier.

Goldstein, J. and Breen, C. (2020). Berkeley Unified Numident Mortality Database: Public Administrative Records for Individual Level Mortality Research. page 33.

Green, R. S., Basu, R., Malig, B., Broadwin, R., Kim, J. J., and Ostro, B. (2010). The effect of temperature on hospital admissions in nine California counties. *International Journal of Public Health*, 55(2):113–121.

Gronlund, C. J. (2014). Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review. *Current Epidemiology Reports*, 1:165–173.

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K., Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., and Martin, R. V. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environmental Science & Technology*, 54(13):7879–7890. Publisher: American Chemical Society.

Hansen, A., Bi, L., Saniotis, A., and Nitschke, M. (2013). Vulnerability to extreme heat and climate change: is ethnicity a factor? *Global Health Action*, 6(1):21364. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/gha.v6i0.21364.

HMD (2021). Human mortality database. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).*

Hovdahl, I. (2020). Deadly Variation: The Effect of Temperature Variability on Mortality. Working Papers No 01/2020, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.

Lee, M., Shi, L., Zanobetti, A., and Schwartz, J. D. (2016). Study on the Association Between Ambient Temperature and Mortality Using Spatially Resolved Exposure Data. *Environmental research*, 151:610–617.

Liu, X., Behrman, J., Hannum, E., Wang, F., and Zhao, Q. (2021). Same Environment, Stratified Impacts? Air Pollution, Extreme Temperatures, and Birth Weight in South China. SSRN Scholarly Paper ID 3759319, Social Science Research Network, Rochester, NY.

N.C.I (2021). National Cancer Institute, U.S. population estimates, 1969-2019.

Noe, R., Jin, J., and Wolkin, A. (2012). Exposure to natural cold and heat: hypothermia and hyperthermia Medicare claims, United States, 2004-2005. *American journal of public health.*

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

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

Petersen, R., Pan, L., and Blanck, H. M. (2019). Racial and ethnic disparities in adult obesity in the united states: Cdc's tracking to inform state and local action. *Preventing chronic disease*, 16.

Son, J.-Y., Liu, J. C., and Bell, M. L. (2019). Temperature-related mortality: a systematic review and investigation of effect modifiers. *Environmental Research Letters*, 14(7):073004.

Stafoggia, M., Forastiere, F., Agostini, D., Biggeri, A., Bisanti, L., Cadum, E., Caranci, N., de'Donato, F., De Lisio, S., De Maria, M., Michelozzi, P., Miglio, R., Pandolfi, P., Picciotto, S., Rognoni, M., Russo, A., Scarnato, C., and Perucci, C. A. (2006). Vulnerability to Heat-Related Mortality: A Multicity, Population-Based, Case-Crossover Analysis. *Epidemiology*, 17(3):315–323.

Taylor K.E., Stouffer, R. and Meehl, G. (2012). An overview of cmip5 and the experiment design. *MS-D-11-00094.1*.

Weinberger, K. R., Harris, D., Spangler, K. R., Zanobetti, A., and Wellenius, G. A. (2020). Estimating the number of excess deaths attributable to heat in 297 united states counties. *Environmental Epidemiology (Philadelphia, Pa.)*, 4(3).

Zanobetti, A. and Schwartz, J. (2008). Temperature and mortality in nine us cities. *Epidemiology*, 19(4).