Fatal Errors:
The Mortality Value of Accurate Weather Forecasts

Jeffrey G. Shrader, Laura Bakkensen, and Derek Lemoine*

March 11, 2022

Abstract

We provide the first revealed-preference estimates of the benefits of routine weather forecasts. The benefits come from how people use the information to reduce mortality from heat or cold. We show that more accurate forecasts reduce mortality if and only if mortality risk is convex in forecast errors, which in turn depends on how people use forecasts. Using data on the universe of mortality events and weather forecasts for a twelve-year period in the U.S., we show that making forecasts 50% more accurate would save 1,700 lives per year, for gross annual benefits of $18 billion. The effects of forecast errors indicate that adaptation becomes less effective if temperatures are either higher or lower than expected. Forecast-driven adaptation is especially important in extreme heat, which suggests that short-run weather forecasts could be an important tool for managing the effects of climate change. (JEL:D83,I12,Q51)

*Shrader: School of International and Public Affairs, Columbia University (email: jgs2103@columbia.edu); Bakkensen: School of Government and Public Policy, University of Arizona (email: laurabakkensen@arizona.edu); Lemoine: Department of Economics, University of Arizona, NBER, and CEPR (email: dlemoine@arizona.edu). We thank Michael Best, Matthew Gibson, Matthew Neidell, and Wolfram Schlenker for helpful comments and suggestions. Special thanks to Anouch Missirian for providing code and pointers on dealing with wind data. Gabriel Gonzalez Sutil provided excellent research assistance. All errors are our own.

1 Introduction

Routine weather forecasts are the product of sophisticated scientific and policy efforts that require global cooperation. Data from around the world are continuously gathered and processed to produce forecasts that are disseminated multiple times per day, largely for free, to the public. Despite the ubiquity of weather forecasts, the number of people who rely on them, and the effort involved in their production, surprisingly little is known about their economic value. Valuing improvements to forecasting systems has long been of policy interest (e.g. (Chapman, 1992, WMO et al., 2015)). To date, valuations have relied on simulation models, stated preference surveys, or other heuristics rather than on real-world behavior.[1] Our study provides the first revealed preference estimates of the benefits of weather forecasts.

We focus on the value of forecasts for reducing temperature-related mortality. While forecasts have many uses, enabling preparation for extreme weather is among the most prominent, and extreme temperatures cause more loss of life in the U.S. than any other form of extreme weather (Pielke and Carbone, 2002). A large literature across economics, epidemiology, and other fields has investigated the effects of realized temperature extremes on mortality, corroborating the finding that extreme temperatures are a major source of mortality.[2] To date, this literature has not investigated the role of weather forecasts in helping individuals avoid mortality.

We formally demonstrate that whether more accurate forecasts reduce mortality depends on whether mortality risk is convex in forecast errors. We show that two plausible models of forecast-driven, "ex-ante" adaptation to temperature risk can have opposite implications for the value of more accurate forecasts.[3] If adaptation is "protective", then forecasts that are too extreme call forth additional adaptation that further reduces mortality risks whereas forecasts that are too mild increase mortality risk by suppressing adaptation. With this type of adaptation, more accurate forecasts can either increase or decrease mortality, as they reduce excess deaths from insufficiently extreme forecasts but also reduce lives saved from overly extreme forecasts. In contrast, if adaptation is "appropriate", then any forecast errors reduce the effectiveness of adaptation by worsening the match between the chosen adaptation

[1] In justifying its budget, the U.S. National Weather Service relies either on decade-old, stated-preference surveys of willingness to pay for forecasts or on a measure of the total amount of economic activity that is "sensitive to weather," without any estimate of the degree to which forecasts could be useful for reducing that sensitivity (NOAA, 2021). Many have emphasized the need for good estimates of the economic value of forecasts (e.g., Freebairn and Zillman, 2002, Pielke and Carbone, 2002, National Research Council, 2010, Katz and Lazo, 2011).

[2] See, for instance, Anderson and Bell (2009), Deschênes and Moretti (2009), Deschênes and Greenstone (2011), Gasparrini et al. (2015), Barreca et al. (2016), Carleton et al. (2020).

[3] We remain agnostic about whether adaptation is an individual choice or is a public investment, such as cooling shelters. Our estimates encompass the net effect of all ex-ante adaptation. "Ex-post" adaptation responds to temperature realizations and is absorbed by our controls for realized weather

and realized temperature. With this type of adaptation, more accurate forecasts unambiguously reduce mortality by improving the match between adaptation and temperature. It is ultimately an empirical question whether more accurate forecasts reduce mortality or not.

To estimate the effect of weather forecasts on mortality, we combine the universe of deaths reported by the Centers for Disease Control and Prevention (CDC) with daily temperature realizations and forecasts issued by the National Weather Service (NWS). We study the continental U.S. from 2005 through 2017. We focus on day-ahead forecasts of temperature, which households say are among the most important forecast products (Stratus Consulting Inc., 2002). Our regression framework accounts for potential location-specific and time-varying confounders as well as for the potential direct effect of temperature and other weather on mortality. Across the full sample, we find that mortality risk is indeed convex in forecast errors and its shape is consistent with appropriate adaptation but not with protective adaptation. As a result, reducing the standard deviation of forecast errors by 50% would save 1,700 lives per year, creating $18 billion in value annually. This value is comparable to the entire value for current forecasts previously estimated via survey methods.[4]

We assess whether it is more important to improve forecasts' accuracy on cold, hot, or moderate days. One might expect extreme weather to drive forecasts' value, but we show that much of the value from improved forecasts comes from days with more moderate weather because they are so frequent: extreme heat and cold represent only around 10% of the sampled days.[5] Nonetheless, halving the standard deviation of forecast errors either only on extremely hot days or only on extremely cold days would still generate around $2 billion in value. The source of this value differs between extreme cold and extreme heat. For extreme cold, this value arises because forecasts are currently especially noisy on these days and because these days are nearly five times more common than extremely hot days. For extreme heat, this value arises because mortality risk is especially convex in forecast errors, suggesting that ex-ante adaptation is especially responsive to forecasts in extreme heat and/or especially consequential for mortality in extreme heat.

We estimate gross benefits that are much larger than the direct costs associated with producing and disseminating weather forecasts. The National Oceanic and Atmospheric Administration (NOAA) had a total budget of $5.5 billion in fiscal year 2021, with $1.2 billion allocated to the National Weather Service (NOAA, 2021). In 1999, the U.S. government reported spending $2.2 billion on producing and disseminating forecasts and $0.5 billion on research to improve them, with the private sector spending another $1 billion broadcasting

[4]Stratus Consulting Inc. (2002) estimate that the total value for forecasts is $11.4 billion in 2001 dollars, which is around $16 billion in 2020 dollars.

[5]Our primary results define extreme cold as a day with an average temperature below 0°C and define extreme heat as a day with an average temperature above 30°C. The exact share of value attributed to moderate days is sensitive to how one defines "moderate". Our estimates suggest that these days contribute $12 billion to the value of more accurate forecasts.

the forecasts (Hooke and Pielke Jr., 2000).

Many decisions about future investment in forecasts will require information about how valuable forecasts are in their current state and how that value might change if the forecasts were to become more accurate. Short-run weather forecasts have steadily improved since 1980 (Bauer et al., 2015). Decade-on-decade, forecast skill has risen so that, currently, 7-day-ahead forecasts are typically as skillful as 3-day-ahead forecasts were in the 1980s. Forecast skill will likely continue to increase in the near future as weather forecasting groups continue to improve data assimilation, modeling, and computing power (Toth and Buizza, 2019).

Our results are the first revealed preference estimates of the benefits of routine weather forecasts.[6] Recent theoretical work emphasizes that short-run forecasts such as those studied here can be especially valuable for planning purposes (Millner and Heyen, 2021). Previous valuations of routine weather forecasts either tally up the value of sectors judged to be sensitive to weather (National Research Council, 1998) or use stated preference methods based on surveys of 381 individuals (Stratus Consulting Inc., 2002). Many authors in the forecasting literature have recognized that it would be ideal to find a market in which people reveal their value for forecasts with real bets but lament that such markets do not exist for publicly provided forecasts (e.g., Freebairn and Zillman, 2002, Letson et al., 2007, Morss et al., 2008, Katz and Lazo, 2011). For instance, Katz and Lazo (2011, 5740) observe, "Perhaps because of the scarcity of links to market transactions that would permit revealed preference applications, there has been virtually no work to date using revealed preference methods for assessing values for weather forecasts." We here infer that agents act on forecasts by exploring how forecasts affect observed mortality, and we value forecasts from the reduction in mortality they enable.[7]

In the existing literature on the effects of realized temperature on mortality, both hot and cold temperatures are typically associated with excess mortality. Cold events are associated with little acute (i.e., same-day) mortality but exhibit elevated mortality in days following the event. Heat exhibits the reverse pattern, with elevated mortality only on the day of the event

[6]Seasonal weather forecasts have been subject to more substantial study. Meza et al. (2008) summarize 33 papers studying the effect of seasonal forecasts on agricultural outcomes. Some more recent work has also estimated the value of seasonal forecasts for climate-exposed production (e.g., Rosenzweig and Udry, 2019, Shrader, 2021). We focus on routine, shorter-run forecasts that predict a particular day's weather only a few days in advance. Short-run forecasts are much more accurate, widely produced, and widely used than seasonal forecasts. Other work studies the effects of extreme weather warnings (e.g., Craft, 1998, Bakkensen, 2016, Miller, 2018, Weinberger et al., 2018, Kruttli et al., 2019), but even there some lament the dearth of rigorous work (Sutter and Ewing, 2016). We study the effects of the forecasts that underpin excessive heat warnings and wind chill warnings. The consequences of warnings are included in our estimated effect of forecasts.

[7]We do not observe the costs of acting on forecasts (except insofar as those actions themselves affect mortality), so our estimates should be taken as a gross measure of benefits. However, we also do not observe all the other ways that forecasts provide value, which will tend to lead us to underestimate forecasts' value.

and in the immediate aftermath. The dominant methodology uses fixed effects to account for the average weather in a time and place. To gain exogenous variation, this methodology needs to isolate the consequences of weather shocks, but shocks relative to average weather (that the fixed effects models identify) may not be shocks relative to expectations.[8] When people act on their information about coming weather, accurately forecasted shocks can have very different implications from inaccurately forecasted shocks. It is important to disentangle these effects when assessing policy responses to extreme temperatures and also when extrapolating to the effects of climate change.[9]

Section 2 theoretically analyzes how more accurate forecasts can reduce expected mortality. Section 3 describes the data used for the empirical analysis. Section 4 describes the empirical strategy. Section 5 reports results. Section 6 concludes.

2 Formal Analysis

We first analyze the mortality value of more accurate weather forecasts. We then show how responses to forecast errors can identify the interaction between adaptation and temperature. We conclude by showing that certain types of interaction can be responsible for forecasts' value.

2.1 The Mortality Value of Forecasts

To value forecasts, we extend the workhorse single-period model of the value of a statistical life (VSL), the marginal rate of substitution between money and the risk of sudden mortality.[10] This model originates with Drèze (1962) and Jones-Lee (1974) and has been extensively applied in the literature (see Viscusi, 1993, Andersson and Treich, 2011).

An individual's indirect utility over wealth w is $u(w)$ when alive and $v(w)$ when dead. As is conventional, assume that $u(\cdot) > v(\cdot)$, $u'(\cdot) > v'(\cdot) \geq 0$, $u''(\cdot) \leq 0$, and $v''(\cdot) \leq 0$. The individual's hazard of death following temperature T and forecast f is $h(T, f)$. Forecasts can affect the risk of death only through the actions people take in response to them, but to start, we subsume any changes in actions within the forecast argument. The frequency

[8] In fact, we show in Section 3 that forecasts substantially outperform average weather at predicting coming weather.

[9] In cross-sectional regressions, Fishback et al. (2011) show that controlling for access to information (measured as rates of literacy and radio ownership) changes the effect of a year's average daily high temperature on mortality from positive to negative in the U.S. during the Great Depression. The relevant information may include information about coming weather and public health information helpful for managing hot weather.

[10] Simon et al. (2019) review the use of VSL in economics and policy and provide standard definitions. They propose that "value of reduced mortality risk" would more accurately denote the concept and avoid the common misinterpretation of VSL as indicating the value of a life.

of each type of temperature over a year is $p(\cdot)$. On average, each temperature is correctly forecasted: $E[f|T] = T$. The error from forecast f is $e = f - T$, with $E[f|T] = T$ implying that $E[e|T] = 0$.

Consider adding a small daily mortality risk ϵ. The individual maximizes expected indirect utility

$$V = \int E_{f|T}\Big[[1 - h(T, f) - \epsilon]\, u(w) + [h(T, f) + \epsilon]\, v(w)\Big]\, p(T)\ \mathrm{d}T,$$

where $E_{f|T}$ indicates expectations over forecasts given temperature. Second-order approximating V around $f = T$, we find

$$V \approx \int \left\{[1-h(T,T)-\epsilon]\, u(w)+[h(T,T)+\epsilon]\, v(w)-\frac{1}{2}h_{ff}(T,T)\,[u(w)-v(w)]\, Var[f|T]\right\} p(T)\ \mathrm{d}T,$$

where subscripts on h indicate partial derivatives. (Observe that $Var[f|T] = Var[e|T]$.) Totally differentiating while holding V constant, willingness to accept the small risk is

$$\begin{aligned} VSL &\triangleq \left.\frac{\mathrm{d}w}{\mathrm{d}\epsilon}\right|_{\epsilon=0} \\ &= \frac{u(w) - v(w)}{\int \{(1 - h(T,T))u'(w) + h(T,T)v'(w) - \frac{1}{2}h_{ff}(T,T)\,[u'(w) - v'(w)]\, Var[f|T]\}\, p(T)\ \mathrm{d}T}. \end{aligned}$$

This willingness to accept is the VSL.[11]

Willingness to accept an increase in the variance of errors around some particular temperature $\hat{T}$ is[12]

$$\left.\frac{\mathrm{d}w}{\mathrm{d}Var[f|T=\hat{T}]}\right|_{\epsilon=0} = \frac{1}{2} VSL\ h_{ff}(\hat{T}, \hat{T})\ p(\hat{T}).$$

The right-hand side is positive if and only if the risk of death is convex in the forecast within the neighborhood of the correct forecast. To better match the empirical specifications below, we can also write the risk of death as $\tilde{h}(T, e) \triangleq h(T, T + e)$, in which case

$$\left.\frac{\mathrm{d}w}{\mathrm{d}Var[f|T=\hat{T}]}\right|_{\epsilon=0} = \frac{1}{2} VSL\ \tilde{h}_{ee}(\hat{T}, 0)\ p(\hat{T}). \tag{1}$$

[11] We assume throughout that the variance of forecasts errors is small enough to ensure that $VSL > 0$. In the empirical application, the variance of forecast errors is in fact reasonably small.

[12] One might consider instead fixing the distribution of forecasts over the year, assuming that forecasts are unbiased on average, and analyzing a reduction in the variance of each forecast's error. The problem with that approach is that the comparative static implicitly alters the distribution of realized temperature, which can create value even if agents never act on forecasts.

When the risk of death is convex in forecast errors, expected mortality increases in the variance of forecast errors. The value of more accurate forecasts follows from this reduction in expected mortality. It is most valuable to improve the accuracy of forecasts for temperatures around which the risk of death is especially convex in forecast error (with large $\tilde{h}_{ee}(\hat{T}, 0)$) and for temperatures that are especially frequent (with large $p(\hat{T})$).

Note that $h_f \neq 0$ (equivalently, $\tilde{h}_e \neq 0$) only if agents act on forecasts, as forecasts are irrelevant to mortality if agents do not use them. Such actions are adaptations to weather risk. Consequently, adaptation actions also must be responsible for making $h_{ff} \neq 0$ (or $\tilde{h}_{ee} \neq 0$). The mortality value of forecasts derives not merely from how agents act on forecasts (h_f) but from how those actions interact with mortality risk when forecasts are mistaken (h_{ff}).

2.2 How Forecast Errors Identify the Form of Ex-Ante Adaptation

Before considering the sources of convexity in mortality risk, consider what we learn from the effects of forecast errors. In the foregoing analysis, $h(T, f)$ subsumed any adaptation as an effect of f, but we now make adaptation explicit. Let A indicate ex-ante adaptation, chosen based on forecasts but without knowing realized temperature, and express the probability of death as $H(T, A(f))$.[13] The function $A(f)$ may result from agents trading off the costs and benefits of adaptation in response to forecasts, but we here avoid writing down an explicit optimization problem and take $A(f)$ to be a reduced-form representation of agents' adaptation decisions, for whatever situations they face.

Consider an extreme high temperature, for which $H_T > 0$ locally. And measure ex-ante adaptation such that the chosen actions $A(f)$ are monotonically increasing in forecasts around this temperature ($A'(f) > 0$). The actions could represent the chosen magnitude of some adaptation option to implement or could represent the chosen set of adaptation options to undertake. There are two plausible models for how ex-ante adaptation may affect the risk of death.

The first model is one of *protective adaptation*. Here additional adaptation always reduces mortality risk over the relevant domain ($H_A < 0$), as when $H(T, A) \propto T - A$. If an individual happens to receive a too-high forecast, then A is larger than would have been optimal based on perfect knowledge of temperature. When adaptation is protective, mistakenly undertaking too much adaptation reduces mortality risk compared to a case in which the realized temperature had been forecasted accurately. Conversely, mistakenly undertaking too little adaptation because of a too-mild forecast increases mortality risk.

The second model is one of *appropriate adaptation*. Here adaptation is targeted to par-

[13] This representation still subsumes any ex-post adaptation (i.e., adaptation chosen based on knowledge of realized temperature) into the effect of temperature.

ticular temperatures and is less effective if temperatures are either higher or lower than the target: $H_A > 0$ for high temperatures and $H_A < 0$ for low temperatures, as when $H(T, A) \propto [T - A]^2$.[14] When adaptation is protective, the chosen ex-ante adaptation may minimize mortality risk at temperatures near the forecast but will be less effective at higher or lower temperatures.[15]

In the empirical setting, we will estimate mortality risk as a function of forecast errors. Continue considering a case of extremely high temperatures. A positive error means that the forecast was too warm. If adaptation is protective, then positive errors would actually reduce mortality risk, but if adaptation is appropriate, then positive errors would increase mortality risk relative to a case with the same temperature but no forecast error. In contrast, negative errors (i.e., a too-cold forecast) would increase mortality risk in either model. Therefore we can divide the possible empirical outcomes into three cases:

1. Mortality risk decreases in forecast errors. This case is consistent with the protective adaptation model.

2. Mortality risk has a U-shape in forecast errors. This case is consistent with the appropriate adaptation model.

3. Mortality risk has some other shape. This case is inconsistent with either adaptation model on its own, although it potentially reflects a combination of the two.

Analogous cases hold for extreme cold, with the first case modified to have mortality risk increasing in forecast errors.

2.3 How Ex-Ante Adaptation Can Make Forecasts Valuable

Finally, consider the conditions under which $h_{ff} > 0$, as was required for forecasts to reduce expected mortality. Recalling that $h(T, f) = H(T, A(f))$, $h_{ff} > 0$ if and only if

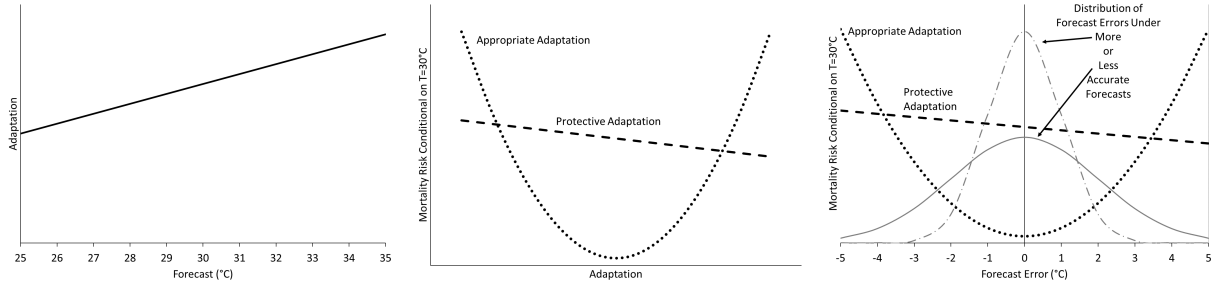
$$H_{AA}[A']^2 + H_A A'' > 0.$$

Again measure ex-ante adaptation so that $A'(f) > 0$ and consider dangerously high temperatures. If ex-ante adaptation is protective ($H_A < 0$), then forecasts are valuable only if either ex-ante adaptation tails off as forecasts increase ($A'' < 0$) or the protective effect increases

[14] $H_A > 0$ could reflect interactions between ex-ante adaptation and the physical effects of temperature or between ex-ante adaptation and ex-post adaptation based on realized temperature.

[15] In the presence of adaptation costs, optimal ex-ante adaptation will not minimize mortality risk exactly at the forecasted temperature, but we may in general expect the mortality-minimizing temperature to be in the general neighborhood of the forecast. Intertemporal considerations such as adjustment costs or resource depletion would further complicate the dynamics of ex-ante adaptation (Lemoine, 2021).

Figure 1: Illustrating how the form of adaptation is critical to the mortality value of forecasts.



(a) Adaptation vs Temperature Forecast (b) Mortality Risk vs Adaptation (c) Mortality Risk vs Forecast Errors

as ex-ante adaptation increases ($H_{AA} > 0$). If only one of these conditions holds, then improving forecast accuracy may not reduce expected mortality when ex-ante adaptation is protective.[16] In contrast, if ex-ante adaptation is appropriate with a mortality-minimizing forecast near T ($H_A < 0$ for $T < f$ and $H_A > 0$ for $T > f$), then $H_{AA} > 0$ around $T = f$ and $H_A \approx 0$ around $T = f$, so forecasts generically reduce expected mortality. Improved accuracy increases social value by enabling agents to better target ex-ante adaptation to realized temperatures.

Figure 1 illustrates this analysis. The left panel depicts ex-ante adaptation as a linear function of forecasted temperature. It assumes, as above, that adaptation increases in the temperature forecast, based on whatever costs, benefits, and constraints agents face when choosing actions. The middle panel plots two plausible relationships between adaptation and mortality risk, conditional on a realized temperature. In the protective adaptation case (dashed), mortality risk declines linearly in adaptation, but in the appropriate adaptation case (dotted), mortality risk is minimized when adaptation is most suited to the given temperature and increases as adaptation moves away from that level in either direction. The right panel plugs the adaptation from the left panel into the mortality relationship in the right panel, thereby depicting mortality risk as a function of forecast errors. It also plots example densities for more and less accurate forecasts, again conditional on realized temperature. In this example, reducing the variance of forecast errors does not reduce expected mortality in the protective adaptation case because mortality risk is linear in forecast errors (because $A'' = 0$ in the left panel and $H_{AA} = 0$ in the middle panel): the extra mortality from too-mild forecasts is offset by the reduction in mortality from too-extreme forecasts. However, as is apparent either by visual inspection or by Jensen's inequality, reducing the variance of forecast errors does reduce expected mortality in the appropriate adaptation case. These

[16] Of course, improved accuracy may nonetheless provide social value by avoiding the costs of accidentally excessive adaptation.

benefits arise from the smaller probability of errors that push risk up either end of the U-shape.

The mortality value of forecasts therefore depends on how forecasts affect adaptation choices and on how adaptation choices affect mortality risk when forecasts turn out to miss the mark. It is not *a priori* obvious which type of adaptation predominates in the real world or, following the example from Figure 1, whether forecasts would have much value at all. It is ultimately an empirical question whether more accurate forecasts in fact reduce mortality.

3 Data

To estimate the effect of forecasts on mortality risk, we combine data on mortality events, realized temperature, and temperature forecasts.

3.1 Weather and Weather Forecasts

Our two primary explanatory variables are daily average temperature and forecasted temperature. The NWS issues weather forecasts at horizons between 1 and 7 days. We focus on daily minimum and maximum temperature point forecasts, from which we calculate daily average temperature by taking the simple average of the two measures. The NWS runs the forecasting model multiple model times during the day, with the most important runs at noon and midnight UTC. We use the noon UTC run because it is typically the one reported in morning news broadcasts. These are also the forecasts that are available at any time of the day on the public NWS website, `weather.gov`.

The forecasts are stored in the National Digital Forecast Database (NDFD) which was created in the early 2000s to standardize the processing, storage, and dissemination of weather forecasts in the U.S. (Glahn and Ruth, 2003). Meteorologists in different locations across the country, known as Weather Forecasting Offices (WFOs), work in shifts to produce forecasts for their local area, known as County Warning Areas or CWAs (see a map of these areas in Figure A3). The NDFD stores the forecasts on a consistent spatial grid with resolution of 2.5km or 5km, depending on the time period.[17]

We use forecast data from April 13, 2005 onward, which is the universe of data available in the NDFD containing both minimum and maximum temperature forecasts. Roughly 5% of the county-day values are missing due to missing values in the underlying, raw NDFD

[17]The NDFD was created after the completion of a major modernization and restructuring of the NWS that occurred between 1990 and 2000 (National Research Council, 2012). The modernization was based around installation of the Next Generation Weather Radar (NEXRAD) stations. NEXRAD has a range of roughly 200 km, and a grid of NEXRAD stations were constructed to provide continuous coverage of weather systems across the U.S. WFOs were relocated or created near the NEXRAD stations. Our study period occurs entirely after the restructuring was complete.

data (owing, for example, to data corruption in the NWS archives or to a missed forecast deadline).[18] We aggregate the forecasts to the county level by taking the population-weighted average, based on the 2010 population grids from CIESIN (2017).

For weather realizations, we use PRISM (Parameter-elevation Regressions on Independent Slopes) Climate Group data (PRISM Climate Group, 2004). PRISM combines weather station observations with an interpolation procedure that accounts for causes of weather gradients such as elevation, weather inversions, rain shadows, and coastal proximity. The output is daily measures of weather on a consistent 4km resolution grid across the country. The PRISM data provide more consistent geographic coverage than raw weather station data.[19]

We aggregate the gridded measures to the county level using the same procedure as the forecasts. Maps of the spatial variation in weather and comparisons of weather and forecast values can be found in the Appendix Figure A4. The final weather dataset contains measures of daily minimum and maximum temperature as well as control variables for total daily rainfall and average dew point temperature for each continental U.S. county from April 13, 2005 to December 31, 2017. We calculate a day's average realized temperature and average forecasted temperature by averaging of the day's minimum and maximum temperature.

3.2 Mortality

The primary outcome we study is mortality. The effect of forecasts on mortality is of particular interest to the NWS. Both the NWS and the CDC point to statistics showing that extreme heat is the number one source of weather-related fatalities, on average, in the U.S., and one of the goals of the NWS when issuing forecasts and extreme weather alerts is to minimize the loss of life.

Mortality data come from the CDC's National Center for Health Statistics Multiple Cause of Death (MCOD) file. It contains records of all vital events that occurred in the U.S. from 2004 to 2017. We use the restricted access version of the dataset, which records the day and county of each mortality event. From the set of all mortality events, we calculate county mortality rates per 100,000 people by dividing the total mortality each day by the county population in that year. Population figures are from the NIH Surveillance, Epidemiology, and End Results (SEER) Program (see Section 3.3). The race, sex, and cause of death for the decedent are also recorded and are discussed further in the heterogeneity analysis results.

[18] The results we report below are robust to estimating with a sample that interpolates the missing values based on the most recent available forecast.

[19] We use the original PRISM data rather than the version produced by Schlenker and Roberts (2006, 2009) so that we can also measure dewpoint temperature for humidity robustness checks. One of the advantages of the Schlenker-Roberts version of the data is greater temporal consistency. Over the relatively short time series analyzed in this project, the difference between the original PRISM data and the Schlenker-Roberts data is minimal.

3.3 Additional Data

In the primary estimation sample, we include control variables for population shares in four different age groups (less than 1, 1 to 44 years old, 45 to 64 years old, and 65 or older). These population shares come from the same NIH SEER data from which we extract annual, county population. Additional datasets are used for heterogeneity, robustness, and mechanism analysis. These datasets are described in the sections where results using the data are shown.

3.4 Data Structure and Summary Statistics

The primary estimation sample consists of all non-missing observations of all-cause mortality, average temperature, total daily precipitation, average temperature forecasts, and population shares for each county in the continental U.S. and each day from April 13, 2004 to December 31, 2017. The initial results focus on the 1-day-ahead temperature forecast, and results investigating dynamics analyze longer-horizon forecasts. Summary statistics for the sample are shown in Table 1.

Table 1: Summary Statistics

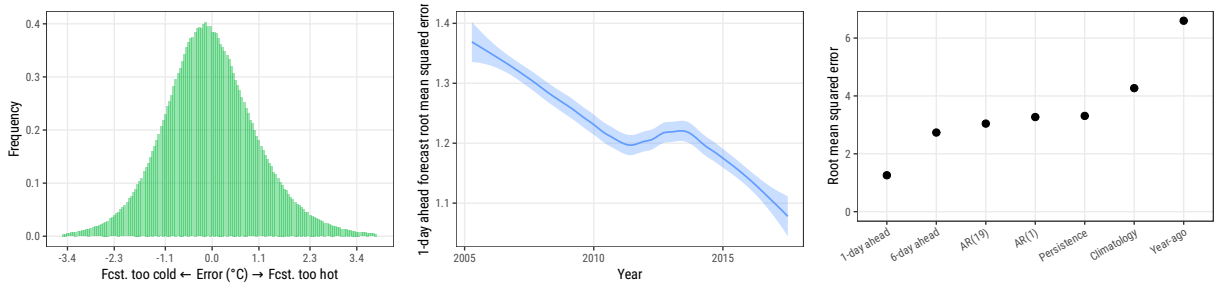
Variable	Mean	S.D.	Observations
Daily all-cause mortality rate (per 100,000)	2.25	1.73	13,704,398
Average temperature (°C)	14.6	10.1	13,704,398
1-day-ahead avg. temperature forecast (°C)	14.5	9.96	13,704,398
1-day-ahead forecast error (°C)	-0.041	1.15	13,704,398

Notes: The table shows summary statistics for the primary variables in the estimation sample, weighted by county population. The difference between average realized temperature and average forecasted temperature does not necessarily equal the average forecast error due to rounding.

In the estimation sample, the average number of deaths reported for all causes per day across the U.S. is 2.3 per 100,000 people. The average county population in our sample is 97,356, so the average death rate is also almost identical to the number of deaths, on average, per county per day. Multiplying the value in Table 1 by total population (per 100,000) indicates that there are about 7,000 mortality events per day in the Continental U.S., which agrees with aggregate statistics from the CDC during our sample period.

Forecasts tend to be roughly correct on average. The standard deviation of forecast errors is just over 1°C. In the estimation sample, there is a slight cool bias to the forecasts of about -0.04°C. The median bias (weighted by population) is -0.08°C. Both of these biases

Figure 2: Forecast Errors and Comparison With Alternatives



(a) 1-day-ahead fcst. errors (b) Accuracy over sample (c) NWS fcst. vs. alternatives

Notes: Panel 2a shows the errors from the 1-day-ahead NWS forecast. The distribution is truncated at the 0.025 and 99.75 percentiles. The x-axis tick marks are at standard deviations relative to 0. The distribution is weighted by annual, county-level population. Panel 2b shows the trend 1-day-ahead forecast RMSE over the sample period. The line is a local linear regression fit to the daily, national average RMSE. Panel 2c shows the root mean squared error (RMSE) for the 1-day-ahead and 6-day-ahead NWS forecasts compared to four alternatives: an AR(1) forecast using yesterday's observation, a persistence forecast based on yesterday's observation, the average value for that location and day, and last year's value for that location.

are small relative to the standard deviation of forecast errors.[20]

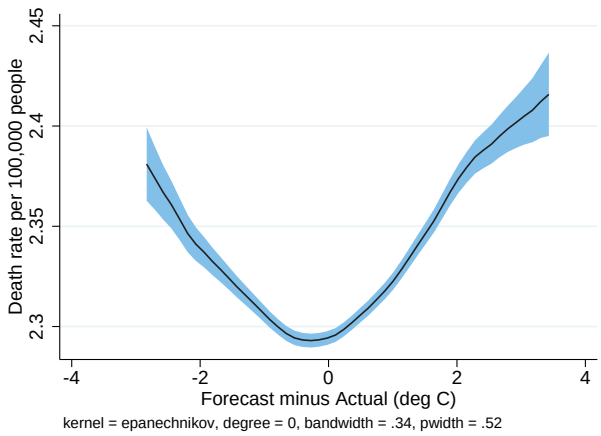
The county population-weighted distribution of 1-day-ahead forecast errors is shown in Figure 2a. The figure shows all errors between the 0.025 and 99.75 percentile, and the x-axis tick marks indicate standard deviations relative to 0, based on the value reported in Table 1. The errors, in addition to being nearly unbiased, are also largely symmetric around zero and exhibit a bell-curve shape. Direct comparison with a normal density shows that the forecast error distribution exhibits slightly thicker tails and excess mass close to 0.

Figure 2b shows how 1-day-ahead forecast accuracy has evolved over our sample period. The line in the figure is a local linear regression fit to the daily average forecast RMSE across the U.S. Over the 12 years of the sample, forecast RMSE has fallen from about 1.35°C to about 1.05°C, an improvement of almost 1/3.

The forecast error for our sample is compared to error from other forecasting methods in Figure 2c. The figure shows the root mean squared error (RMSE) for the forecasts we use in our baseline analysis (the 1-day-ahead forecasts issued by the NWS) as well as five additional forecasts: the 6-day-ahead NWS forecast, an AR(1) forecast using yesterday's value (calculated within sample, so likely overstating more general accuracy), a persistence forecast based on yesterday's observation, a climatological forecast based on the average

[20]See Myrick and Horel (2006) for an early analysis of forecast error and discussion of the creation of a routine verification system that can now be found at `https://sats.nws.noaa.gov/~verification/ndfd/`. Myrick and Horel (2006) emphasize the role that terrain features like rapid changes in elevation can have on forecast error. In our estimation sample, we adjust for location fixed effects to mitigate bias from these types of differences.

Figure 3: Raw Data Relationship Between 1-Day Ahead Forecast Error and Mortality



Notes: The figure shows the relationship between forecast error from the 1-day ahead forecast and the daily mortality rate, estimated using the raw data for counties with populations greater than 10,000 people. The forecast error is Winsorized at the 1% level. The fitted line is a local polynomial regression with an Epanechnikov kernel and the default, plugin bandwidth. The shaded area is the 95% confidence interval but should be interpreted with caution because they treat all observations as i.i.d. Clustered inference is presented in Section 5.

weather for that location and day, and the observation from one year ago.

The 1-day-ahead forecast substantially outperforms even the best-performing non-NWS competitor, the within-sample AR(1) forecast. The RMSE of the official forecast is less than half of the value for the AR(1) forecast. The RMSE for the climatological forecast is almost 4 times higher. The climatological forecast comparison is notable because conventional weather-mortality regression specifications implicitly account for this forecast via location fixed effects. The results in Figure 2c show that these controls do not capture all of the information available to an agent one (or even six) days ahead of a weather realization.

3.5 Motivating Evidence

For initial motivation, Figure 3 plots local polynomial regressions fit to the raw data. The y-axis is the daily mortality rate. The x-axis is the forecast error in degrees C (forecast temperature minus realized temperature). The figure shows the relationship between mortality and forecasts when the forecasts are too hot relative to the realized temperature (values greater than 0 on the x-axis) and too cold relative to the truth (values less than 0 on the x-axis). The forecast error is Winsorized at the 1% level for legibility, and to avoid outliers the plot shows only counties with populations larger than 10,000 people. Otherwise, the relationship is unconditional.

The figure shows that the relationship between mortality and forecast error is convex

and U-shaped: accurate forecasts are associated with the lowest mortality whereas forecast errors in either direction are associated with substantially higher mortality. In the raw data, a one standard deviation forecast error (about 1.15°C) is associated with about 0.05 more deaths per 100,000 people per day, which increases average deaths (of 2.2–2.3) by around 2%. The marginal effect of a greater forecast error is roughly the same for all negative and positive forecast errors outside of a narrow range around 0. In the context of Section 2, these results strongly favor a model of appropriate adaptation.

4 Empirical Strategy

We study the relationship between temperature, temperature forecast accuracy, and mortality via the following estimating equation:

$$y_{ct} = \sum_{\ell=0}^{L} \left[h_\ell(e_{c,t-\ell}) + f_{1,\ell}(T_{c,t-\ell}) + f_{2,\ell}(\text{prec}_{c,t-\ell}) \right] + X_{ct}\gamma + \alpha_{cm} + \rho_t + \varepsilon_{ct}. \tag{2}$$

Some analyses (described in Section 5) extend this equation to allow for heterogeneous effects.

The dependent variable y_{ct} is the daily mortality rate in county c on day t. The primary right-hand side variable is forecast error e_{ct}, defined as $e_{ct} \triangleq \hat{T}_{ct} - T_{ct}$ where $\hat{T}_{ct}$ denotes the forecast of temperature and T_{ct} denotes the realized temperature. This definition matches that in Section 2.[21] In the main results presented below, we study the effect of forecasts issued one day ahead. We also directly include functions of realized temperature in the estimating equation to assess the effect of temperature on mortality, conditional on forecasts.

To allow for different effects at extreme versus moderate temperatures and at positive versus negative forecast errors, we estimate flexible relationships between both of these variables and mortality. For temperature realizations, we estimate $f_1(T)$ semi-parametrically by discretizing temperature into bins.[22] In the baseline results, we use 4 bins: one for temperatures below 0°C, another for temperatures between 0 and 15°C, another for temperatures between 15 and 30°, and a final one for temperatures above 30°. Prior work has shown that this number of bins is sufficient to roughly capture the shape of the realized temperature-mortality relationship in a parsimonious manner (Barreca et al., 2016).[23]

The shape of the relationship between mortality and forecast errors is important for determining whether forecast improvements are beneficial. From Section 2.1, if we find that $h(e)$ is convex, then a reduction in the variance of forecast error will lead to a reduction

[21] To reduce the influence of potential outliers, we Winsorize the forecast error at the 1% level.

[22] This is a widely used method for estimating temperature-mortality relationships (e.g., Deschênes and Greenstone, 2011).

[23] In robustness checks, we also use a denser set of 5°C bins and show that the baseline results are, if anything, conservative.

in average mortality. And from Section 2.2, $h(e)$ needs to be flexible enough to capture a potentially non-monotonic relationship between mortality and forecast errors. In the baseline results, we report estimates using a one-knot linear spline and a global quadratic specification. In sensitivity analysis, we also use a series of specifications using quantile bins. We show that the quadratic and spline provide a useful balance between parsimony and flexibility, and we focus on those results for heterogeneity and mechanism analyses.

We estimate distributed lag models to account for lagged effects or temporal displacement of mortality. Effects that occur on day t (with $\ell = 0$) are the "acute" effects of temperature and forecast errors. We also include up to L lags of temperature and forecast errors to account for dynamic effects. For models with small L ($\leq$ 14 days), we use a standard distributed lag estimator.[24]

The additional covariates include L lags of indicators for daily precipitation below the 25^{th} or above the 75^{th} percentile for that county, denoted $prec_{c,t-\ell}$, to account for potentially correlated effects of rainfall. Date fixed effects (denoted ρ_t) remove any national, time-based confounders including day-of-week effects, holidays, overall patterns in economic activity, changes in national policy, and large-scale weather patterns. County-by-month fixed effects (α_{cm}) are included to remove local seasonal patterns. These fixed effects also adjust for any time-invariant differences across locations, including long-run climatic differences, differences in average medical care availability or economic conditions, and information provision and acquisition. X_{ct} denotes county-month fixed effects interacted with quadratic time trends to account for any overall trends in county- or season-specific weather or mortality, and we include month fixed effects interacted with continuous measures of the share of the population in four age categories (less than 1 year old, 1 to 44, 45 to 65, and 65 or older) to address the age-profile of mortality and associated potential confounders. Finally, the regressions are weighted by annual population in the county to better estimate nationally representative values and to increase precision.

Standard errors are clustered at the CWA level. As discussed in Section 3, CWAs are collections of counties that receive weather forecasts from the same NWS forecast office (mapped in Figure A2). The counties are grouped based on technological considerations related to weather observation equipment and so that meteorologists can specialize in forecasting weather for particular areas of the country. The CWA is a natural level for clustering because its counties have similar weather and because its counties all receive forecasts from the same group of meteorologists. Econometrically, CWAs also have some further appealing

[24]For models with large L, we reduce the computational burden of estimating high-dimensional models by approximating the lag terms via a distributed lag spline. We use b-splines to approximate the lag terms, following Barwick et al. (2018). Similar methods have been widely used in previous studies of temperature-related mortality (Gasparrini et al., 2010). Details on the spline construction and estimation can be found in the appendix, Section **??**.

features, particularly when compared to state-level clustering. First, there are 116 CWAs in the Continental U.S., above the typical rule-of-thumb for the minimum number of clusters (Cameron and Miller, 2015).[25] Second, state borders are not set based on meteorological considerations while CWA borders are, making it more plausible that meteorological data is approximately i.i.d. across CWAs than across states.

For identification, we need forecast errors to be as-good-as-randomly assigned within counties over time, conditional on our other covariates. Potential selection effects are minimized by our use of high-frequency variation for identification. Because of our fixed effects, identification is threatened by omitted variables only if idiosyncratic shocks to some unobserved variable correlate with idiosyncratic shocks to forecast errors. Assuming that forecasting systems always strive for accuracy, historical atmospheric conditions will not confound forecast errors because the meteorologist takes them into account when formulating the forecast. A remaining potential threat would be unobserved aspects of realized contemporaneous weather that affect mortality directly and also affect forecast quality. To address the concern, we conduct robustness checks (described further below) that include additional controls for contemporaneous weather and atmospheric conditions, including humidity, wind, and pollution. We find that effects are largely unchanged when these additional covariates are added.

5 Results

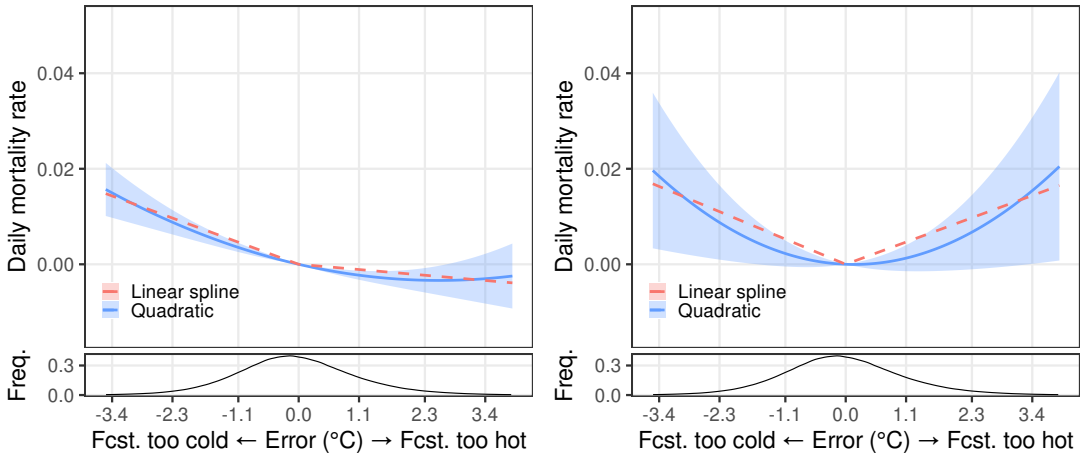
5.1 The Effect of Forecast Errors on Mortality

Estimating Equation 2, we find that forecasts consistently have a convex relationship with mortality. Based on the model presented above, that means forecast improvements will be valuable in terms of reduced mortality, and section 5.3 calculates that value under alternative counterfactuals. In this section, we assess the strength of the convex relationship and whether the effect is U-shaped (consistent with appropriate adaptation) or not (consistent with protective adaptation).

Figure 4 presents estimates graphically, allowing for a visual test of the type of behavior driving the results. The outcome variable is the daily mortality rate (deaths per 100,000 people) in a county. For comparison, there are just over 2 deaths per 100,000 people per day in the average county. The forecast error, given on the x-axis, is the forecasted average temperature for a day minus the realized average temperature. Negative values indicate that the forecast was colder than the weather turned out to be and positive values indicate that

[25] Some counties fall into more than one CWA, and for those locations, we assign the county an identifier that is the combination of all CWAs to it belongs. This results in 130 CWA ids in our sample. See Section 3.3 for details.

Figure 4: Effect of Forecast Accuracy on Mortality



(a) Acute effect

(b) 8-day cumulative effect

Notes: The top panel of each figure shows estimates of the effect of forecast error on the mortality rate. The bottom panels show the distribution of forecast errors (the x-axis). The top panels show cumulative effects every 4 days. Panel (a) shows the acute effect (day t) and Panel (b) shows the effect through 1 week (t through $t+7$). The blue, solid lines and shaded 95% confidence intervals come from a single regression estimated using Equation 2 fit to the baseline data, with a quadratic forecast error function. The red, dashed lines show estimates from a single regression using a linear spline with a knot at 0 for the error function, with confidence intervals suppressed for legibility. Standard errors are clustered at the CWA level.

the forecast was too hot relative to the realized weather. The x-axis tick marks are spaced according to standard deviations of forecast error, and the empirical distribution of forecast errors is shown below each figure.

The panels give an indication of the dynamics of the estimated effects. The left panel shows the same-day ("acute") effect, and the right panel shows the cumulative effect through the end of 1 week. One can see that the relationship between mortality and forecast errors is convex at both horizons, although the convexity gets much stronger over time, with the acute effect appearing nearly linear. Even in that case, however, the marginal effects of a negative and positive forecast error are different. For example, the slopes of the left and right portions of the linear spline (shown by the red, dashed lines) are significantly different from each other (difference of -0.0032, standard error of 0.0015). A 1°C too-cold forecast increases the mortality rate by 0.004, or about 0.2%. In contrast, a 1°C too-warm forecast reduces the mortality rate by 0.001, or about 0.05%.

The effect is substantially more symmetric around 0 error over time. By 8 days, mortality risk is nearly perfectly symmetric in forecast errors. This is consistent with initial action being protective then becoming increasingly appropriate over longer, cumulative shocks. The overall effect of forecasts is also stronger, with the marginal effect of either a positive

or negative forecast error of more than 0.004 deaths per 100,000 people.

Table 2, shows all cumulative effect estimates from the quadratic forecast error specification. The first two columns report the linear and quadratic coefficients, showing that there is substantial and significant convexity of the mortality-forecast relationship across the range of effects. This is further reflected in the marginal effects at error of -1 and error of 1 reported in Column (3) and (4). Column (5) tests the difference in magnitudes of these effects, showing that forecast errors are not only convex but significantly U-shaped after just 3 or 4 days.

5.2 Effects Across the Realized Temperature Distribution

The foregoing results show that forecasts affect mortality on average across all temperatures. We now explore whether forecast errors matter differently on days with different realized temperatures. To do so we estimate a version of equation (2) modified to interact the temperature and forecast error functions

$$y_{ct} = \sum_{\ell=0}^{L} \left[h_{2,\ell}(T_{c,t-\ell}, e_{c,t-\ell}) + f_{1,\ell}(T_{c,t-\ell)}) + f_{2,\ell}(\text{prec}_{c,t-\ell}) \right] + X_{ct}\gamma + \alpha_{cm} + \rho_t + \varepsilon_{2,ct}. \quad (3)$$

Temperature-specific effects are estimating using a function h_2 that interacts bins of realized temperature (the same 4 bins used in the baseline specification) with quadratic functions of forecast error.

Figure 5 shows the predicted effect of forecast errors on mortality stratified by realized temperature bins. Panel (a) shows acute effects and Panel (b) shows cumulative effects over 1 week. For comparison, note that the overall estimates presented in Figure 4 have a magnitude of around 0.02 for forecast errors of 3.4 degrees.

The immediate effect is convex and U-shaped for all realized temperatures aside from the moderate bin between 15 and 30°. On average, the effect is roughly the same as the overall estimate for the two colder bins. Mortality on hot days (above 30° or the hottest 2% of days in the U.S.), in contrast, have not only a symmetric effect around 0 error but also show an extremely strong response to forecasts. The effect of a 1 degree forecast error on a hot day is as strong as the effect of a 3 degree error on a more mild day.

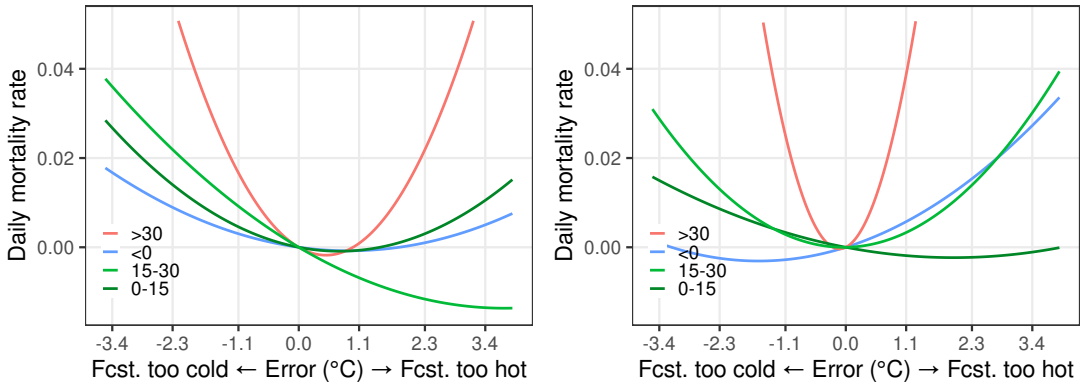
The point estimate is six times larger on days that are hotter than 30°C. Figure 5 suggests that forecast accuracy on a single hot day has substantially more of an impact on the mortality rate than forecast accuracy on other days. Across the year, however, colder days are much more common. In our sample, days above 30°C currently occur only about seven days per year in the average county (which also explains why the estimated effects are so noisy on hot days). As the climate warms, however, hotter days will become more common,

Table 2: Estimates of Forecast Effect on Mortality

Cumulative lag length	(1) Linear coef.	(2) Quadratic coef.	(3) Marginal eff. at -1	(4) Marginal eff. at 1	(5) Test (3) vs (4) =2×convexity
1	-0.0026	0.0005	-0.0037	-0.0016	0.0021
	(0.000500)	(0.000300)	(0.000900)	(0.000500)	(0.001100)
2	-0.0044	0.0015	-0.0073	-0.0014	0.0059
	(0.000700)	(0.000300)	(0.001200)	(0.000800)	(0.001400)
3	-0.0031	0.0016	-0.0063	0.0001	0.0064
	(0.000800)	(0.000400)	(0.001400)	(0.001000)	(0.001700)
4	-0.0027	0.0015	-0.0058	0.0003	0.0061
	(0.000900)	(0.000500)	(0.001400)	(0.001200)	(0.002000)
5	-0.0022	0.0012	-0.0046	0.0003	0.005
	(0.001000)	(0.000600)	(0.001600)	(0.001300)	(0.002300)
6	-0.0014	0.0013	-0.0039	0.0012	0.0051
	(0.001000)	(0.000600)	(0.001700)	(0.001400)	(0.002400)
7	-0.001	0.0009	-0.0028	0.0009	0.0038
	(0.001200)	(0.000600)	(0.001800)	(0.001500)	(0.002500)
8	-0.0004	0.0015	-0.0034	0.0025	0.0059
	(0.001200)	(0.000600)	(0.001800)	(0.001600)	(0.002600)
9	-0.0005	0.0017	-0.0039	0.003	0.0069
	(0.001200)	(0.000700)	(0.001800)	(0.001700)	(0.002600)
10	0.0006	0.0014	-0.0022	0.0035	0.0057
	(0.001200)	(0.000700)	(0.002000)	(0.001800)	(0.002700)
11	0.0009	0.0015	-0.002	0.0038	0.0058
	(0.001400)	(0.000700)	(0.002200)	(0.002000)	(0.003000)
12	0.0021	0.0012	-0.0003	0.0046	0.0049
	(0.001400)	(0.000800)	(0.002100)	(0.001900)	(0.003100)
13	0.0026	0.0014	-0.0003	0.0054	0.0056
	(0.001400)	(0.000800)	(0.002300)	(0.002100)	(0.003200)
14	0.0036	0.0015	0.0005	0.0066	0.0061
	(0.001500)	(0.000800)	(0.002400)	(0.002300)	(0.003400)
15	0.0043	0.0011	0.0021	0.0066	0.0044
	(0.001600)	(0.000900)	(0.002500)	(0.002400)	(0.003500)

Notes The table shows the cumulative effect estimates from a single estimation of Equation 2 with a quadratic specification. The dependent variable is the daily mortality rate per 100,000 people. Column (1) shows the linear terms and Column (2) the quadratic terms. Columns (3) and (4) show marginal effects at error of -1 and 1. Opposite signs on those term indicates U-shaped response, and Column (5) tests for the significance of the U-shape by comparing -1 times Column (3) to 1 times Column (4). This value is also 2 times the convexity. Standard errors, clustered at the CWA-level and calculated using the delta method, are below each estimate.

Figure 5: Effect of Forecast Accuracy on Mortality by Realized Temperature



(a) Acute effect (b) 8-day cumulative effect

Notes: The figure shows predicted effects of forecast error on the daily mortality rate based on estimating Equation 3. Panel (a) shows the acute effect (day t), and Panel (b) shows the effect through 1 week (t through $t+7$). The blue line is the effect for cold realized temperatures ($< 0°C$), dark green is for 0–15°, light green is 15–30°, and red is hot temperatures ($> 30°$).

potentially raising the importance and value of weather forecasts.

5.3 Expected Lives Saved and Economic Value

We have seen that forecast errors are convexly related to mortality. Therefore, from the formal analysis in Section 2.1, increasing the accuracy of forecasts will save lives. Here, we quantify this effect for two counterfactual changes in forecast accuracy. Table 3 shows the important elements of the calculation. All estimates are annual values based on two assumptions: (i) that the estimated changes in mortality are persistent, and (ii) that the estimated response to forecasts is invariant to the postulated changes in forecast accuracy. We assess these assumptions in Section 5.4.

The row labelled "All temps" shows the average value of improved forecasts across the full temperature distribution (i.e., in the pooled sample). Column 1 gives the convexity of mortality risk in forecast errors, which Section 2.1 shows is the marginal effect of the variance of forecast errors on the mortality rate. The estimate comes from the same baseline regression results reported in Section 5.1. As seen there, mortality risk is convex in forecast errors. Column 2 shows the frequency of the observed temperature, which is trivially 1 in this case of the unconditional effect.

Column 3 is the marginal effect of the standard deviation of forecast errors on annual, nationwide mortality. From equation (1), it is the convexity multiplied by the frequency,

Table 3: Annual Lives Saved and Economic Value From Forecast Accuracy Improvements

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Marginal effect of error s.d.			50% reduction of error s.d.	
	Convexity h_{ee}	Frequency $p(\hat{T})$	Lives saved	Value ($ billions)	Current error s.d.	Lives saved	Value ($ billions)
All temps	0.0031	1	3915	37.2	1.15	1882	17.9
			(1258)	(11.95)		(605)	(5.7)
By realized temperature							
≤ 0°C	0.0017	0.093	245	2.3	1.39	139	1.3
			(232)	(2.2)		(131)	(1.2)
0–15°C	0.002	0.375	1048	9.95	1.25	546	5.2
			(708)	(6.7)		(369)	(3.5)
15–30°C	0.0038	0.513	2212	21.01	1.02	944	8.96
			(911)	(8.7)		(388)	(3.7)
≥ 30°C	0.034	0.019	592	5.6	0.89	208	1.97
			(293)	(2.8)		(103)	(0.98)

Notes: The table shows the degree of forecast error function convexity and the associated lives saved and economic value from reductions in the standard deviation (s.d.) of forecast errors. The first column indicates the sample over which the values are estimated and calculated. "Convexity" is the second derivative of the quadratic forecast error function based on the estimates reported in Section 5.1. It measures the mortality rate increase from a marginal change in mean absolute forecast error. "Frequency" gives the fraction of observations in each temperature bin. "Lives saved" is the expected number of lives saved across the Continental U.S. from the indicated forecast error improvement. It is the convexity translated into the marginal effect of forecast error standard deviation times frequency times days per year times population, divided by 100,000. "Value" is the expected lives saved times the EPA VSL in billions of 2020 dollars. Terms in parentheses are standard errors clustered at the CWA level.

the population of the U.S. in hundred thousands,[26] and the number of days per year.[27] Equation (1) gives the marginal effect of variance, which we convert to the marginal effect of standard deviation by multiplying by two times the standard deviation. More accurate forecasts save lives: marginally reducing the standard deviation of forecasts errors would save 4,000 lives per year. Notably, this effect is just from day-ahead temperature forecasts, omitting potential benefits from other types of temperature forecasts and from forecasts of other dimensions of weather.

Column 4 translates the number of lives saved into economic value by multiplying by the U.S. Environmental Protection Agency's value of a statistical life (VSL), which is $9.5 million in 2020 dollars (EPA, 2021). The marginal value of forecast accuracy is $38 billion. As discussed in the introduction, the U.S. federal government and the private sector spend roughly $5 billion per year on production and dissemination of forecasts. If NWS forecasts are 1°C more accurate than alternatives, then their value-added via avoided mortality is almost 8 times their total cost.

The final three columns present an alternative counterfactual, which estimates the lives saved and economic value from halving the standard deviation of forecast errors.[28] Currently, temperature forecast errors have a standard deviation of 1.15° (population-weighted), so this counterfactual is close to half the size of the estimates reported in columns 3 and 4. This second counterfactual is more interesting in the lower panel, where we explore the value of improvements at different levels of realized temperature.

That lower panel of Table 3 report the value of improving the accuracy of particular types of forecasts. To motivate these experiments, Figure 6 depicts partial densities of forecast errors. The densities are conditional on realized temperature in four temperature bins. The different error distributions have notably different features. Forecasts of hot temperatures have the lowest standard deviation but the largest average bias.[29] In contrast, the errors during periods with the coldest temperatures have less average bias but a larger spread. The differences in the spread of errors is likely due to underlying meteorology: warm air masses are more stable and easier to forecast than colder, more volatile air masses. The average bias could be due to hedging by the meteorologists themselves. In interviews we conducted with professional meteorologists, they stated that they were historically reluctant to forecast record temperatures for a given location. This behavior would tend to produce negative bias at high temperatures and positive bias at low temperature, exactly the pattern we find in Figure 6.

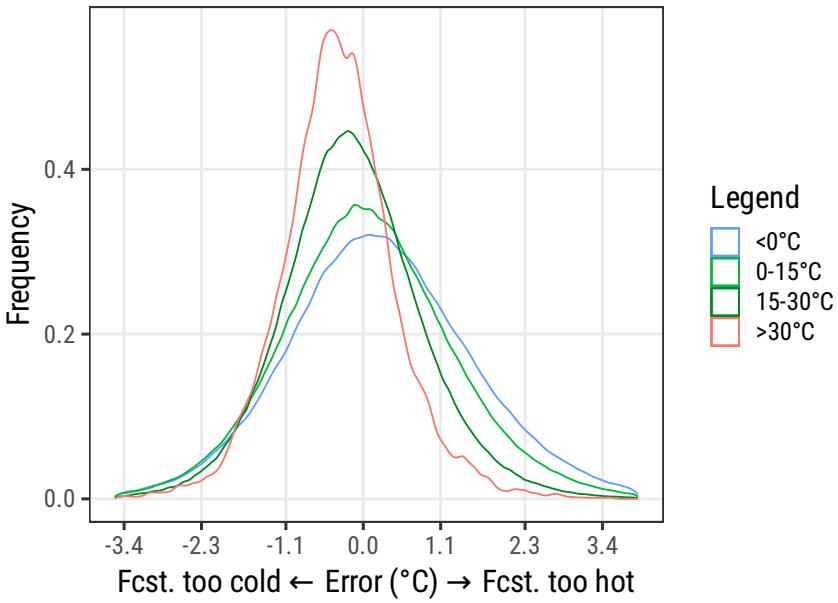
[26] We use 310 million for our sample period population.

[27] We use a mean solar year as our year.

[28] These calculations linearly extrapolate the marginal effect from equation (1). The 50% reduction in the standard deviation corresponds to a 75% reduction in the variance.

[29] As we will see, there are many fewer hot days than other types of days. Some of this larger bias could therefore be due to sampling.

Figure 6: Conditional Densities of Forecast Errors



Notes: Figure shows the empirical density errors from the 1-day-ahead NWS forecast conditional on the realized temperature, broken down into 4 bins. The distribution is truncated at the 0.025 and 99.75 percentiles. The x-axis tick marks are at standard deviations relative to 0 for the unconditional distribution. Each density is weighted by annual, county-level population.

The lower panel of Table 3 reports that a large portion of the total value of forecast accuracy improvements comes from days with moderate temperatures: improving accuracy on days with an average temperature of 15–30°C provides almost half the value of improving accuracy on all days. This high value is despite mortality risk not being especially convex on these days and despite forecasts already being relatively accurate for these temperatures. The high value arises from the frequency of moderate days, which constitute more than half the sample. The value per day may be small, but there are many such days.[30]

Although most value may be concentrated in moderate days, we do see substantial value from the relatively infrequent days with extreme weather. The marginal value of accuracy is slightly larger for extremely hot days than for extremely cold days, but the value of halving the standard deviation is larger for extremely cold days. From Column 1, mortality risk is eight times more convex in forecast errors on hot days, which is why marginal improvements in accuracy are more valuable on those days. However, extremely cold days are five times

[30]One may heuristically think that forecast accuracy is most valuable on days where mortality tends to be high. However, the theoretical analysis in Section 2 shows that the value of forecast accuracy derives from how the effectiveness of ex-ante adaptation at reducing mortality risk varies with forecast errors, not from typical mortality on that day or even from the typical effect of ex-ante adaptation at reducing mortality.

more common and currently have a 50% larger standard deviation in their forecast errors. These effects are enough to offset the smaller convexity of mortality risk on extremely cold days when we consider the policy experiment of halving the standard deviation of forecast errors.

To further contextualize these values, one can also ask by how much forecasts would need to improve in order to justify the cost of NOAA expenditures. NOAA is currently installing a supercomputer that will become operational in 2022, at a cost of $505 million over 8 years. How much better would temperature forecasts need to be to justify $63 million of expenditure per year? First imagine that the new facility improves the standard deviation of all forecasts. Based on the values from the top panel of Table 3, the current standard deviation of forecast errors would need to fall by less than 0.1% to justify such an expense. Next, imagine that the new facility improves the standard deviation of forecasts only on either extremely hot or extremely cold days (but not on both types of days). Based on the values from the bottom panel of Table 3, the current standard deviation of forecast errors would need to fall by about 2% to justify such an expense. As Figure 2b showed, the standard deviation of forecast errors fell by almost 1/3 over the the sample period, making improvements of these magnitudes plausible given recent experience.

5.4 Testing Assumptions Underlying Forecast Value Estimate

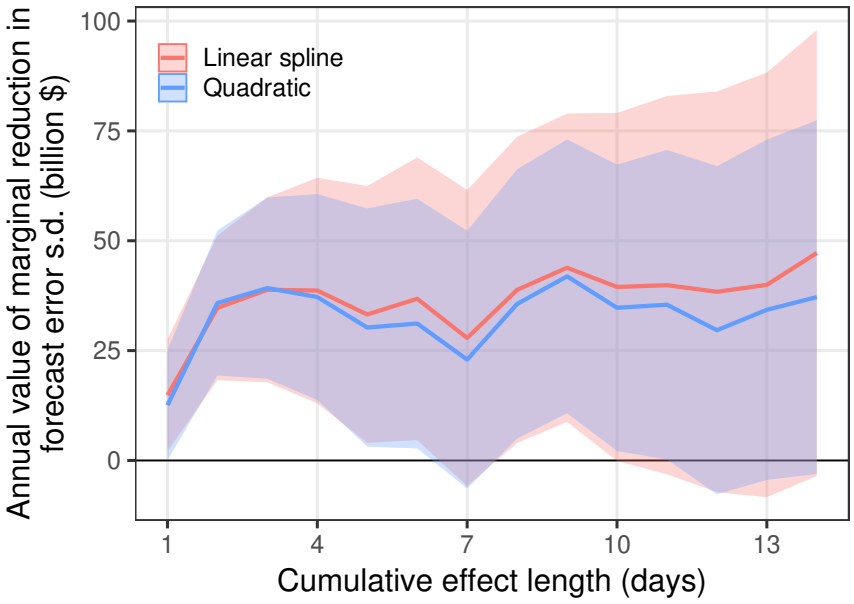
The previous results show the value of forecasts based on the 4-day cumulative effect of forecasts on mortality. One assumption underlying those results is that the effect of forecasts on mortality is persistent (so that VSL is appropriate). Previous research on the effect of realized temperature on mortality has shown that there can be dynamic effects over a few days or weeks after the actual temperature realization. Here, we explore the cumulative effect of forecasts over similar time periods.

Figure 7 shows the value, in billions of dollars, of a marginal reduction in the standard deviation of forecast errors (the same counterfactual reported in Column 4 of Table 3) using a two-week distributed lag model.[31] The results show that on the day the forecast arrives, the annual value of a marginal change in forecast standard deviation is roughly $15 billion. This rises to above $30 billion after just one additional day, and the value stays about the same after that point. The value is significantly positive over all but one horizon less than ten days. After ten days, the point estimate remains constant, but the 95% confidence interval begins to include zero.

The estimates help us assess the plausibility of the assumption that the death reductions from more accurate forecasts are persistent. Prior analyses of mortality from extreme tem-

[31] Examining longer ranges of cumulative effects is computationally infeasible using a standard distributed lag model given the high dimensionality of the estimating equation.

Figure 7: Cumulative Effect of Forecasts on Mortality



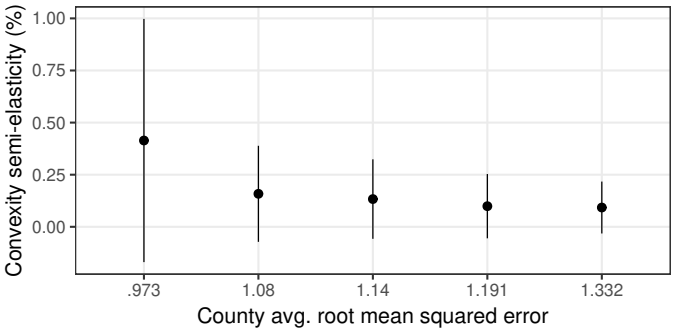
Notes: The figure shows estimates of the cumulative value of a marginal improvement in the standard deviation of forecast error. Each line is based on a single regression estimated using Equation 2 fit to the baseline data. The estimates correspond to Column 4 of Table 3. The red line estimates a quadratic relationship between mortality and forecast error while the blue line uses a linear spline with a knot at 0. The areas show 95% confidence intervals based on standard errors clustered at the CWA level.

peratures has shown important dynamics and intertemporal displacement of mortality over a few days after the initial temperature event. Deschênes and Greenstone (2011) examine daily mortality responses to hot and cold temperatures, finding that cold temperature effects have significant daily effects up to 10 days after the initial cold day. Hot temperatures have significant effects over 3 to 4 subsequent days. Recently, Heutel et al. (2021) showed stable cumulative effects of temperature on mortality after less than one week. The range of our estimates should capture the most substantial dynamic effects if the temporal response of mortality to forecast errors follows that of realized temperature.

The second main assumption underlying Table A1 is that the estimated forecast-mortality relationship still holds under the counterfactual forecast. We can generate descriptive evidence on this assumption by looking at how the effect varies by average forecast quality (measured by RMSE) in the sample. These estimates are non-causal—they simply rely on cross-sectional variation in forecast accuracy. Estimating a version of Equation (2) that interacts the forecast error function, h, with the average RMSE in a county, we find a small, negative interaction term between RMSE and convexity of -0.0013 (standard error

of 0.005). Interpreting the point estimate, this implies that mortality responds about 0.5% more strongly in areas with 1% lower RMSE. If anything, this result suggests that our counterfactual estimates are lower bounds because stronger responses are associated with more accurate forecasts.

Figure 8: Heterogeneity by Average Forecast Accuracy: Assessing Credibility



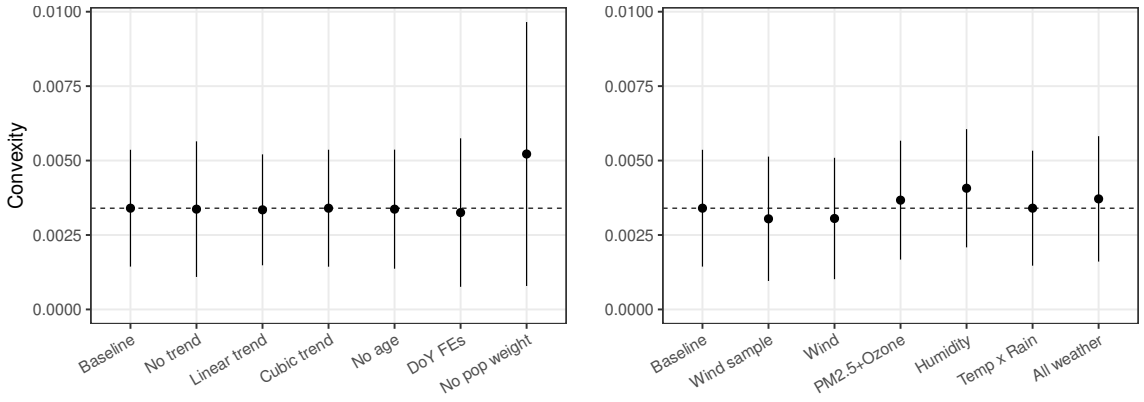
Notes: The figure shows the 4-day cumulative percent reduction in the mortality rate from a 1°C reduction in forecast absolute error (marginal effect) based on a models fit using Equation (4) on the baseline data. The forecast error function is a quadratic. The dimension of heterogeneity is the average RMSE in each county over our full sample. The circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level.

Figure 8 shows this difference in response graphically by breaking down forecasts' effects by the average RMSE of forecasts in the county across the full sample. The estimates are generated by fitting Equation (4) which interacts quintiles of average RMSE in the county with a quadratic forecast error function. The horizontal axis tick marks indicate the quintiles of RMSE. The value of a 1° reduction in absolute forecast error is about 6 times greater in counties in the lowest quintile of RMSE compared to counties in the highest quintile, although the difference is not significant. This evidence suggests that long-run forecast quality and the effectiveness of forecasts are complementary. There are two possibilities for this effect: people in more accurate locations may learn to trust and act on forecasts or those locations may differ from ones with poorer forecasts in unobserved ways (although linear cross sectional confounding is ruled out by our county fixed effects).

5.4.1 Robustness and Sensitivity

In addition to testing the assumptions underlying our counterfactuals, we can also assess sensitivity and robustness of the results to variations in the control set and specification.

Figure 9: Robustness and Sensitivity Checks



(a) Non-weather (b) Weather

Notes: The figure shows robustness and sensitivity checks on the main results reported in Section 5.1. The points are the four-day cumulative estimates of the marginal effect of a increase in forecast mean absolute error ("convexity"), estimated using Equation 2 and a quadratic specification for the forecast error function. The lines are 95% confidence intervals based on standard errors clustered at the CWA level. "Baseline" reproduces the baseline estimate, with controls described in Section 4, for comparison. "No trend" removes all year trends, "Linear trend" removes the quadratic trend, "Cubic trend" adds a cubic trend, "DoY FEs" replaces month fixed effects with day-of-year, "No age int." removes the age interaction, and "No pop weight" removes population weighting. The right panel adds realized weather controls. In all cases, the weather is controlled for non-parametrically using quantile bins, and 4 lags of the bins are included (to match the lag length of the forecast error). The labels indicate the added variable. Note that wind is only available for a subset of the observations, so the sample changes. "All weather" simultaneously includes functions of all additional weather variables listed in the figure.

The results have been conditioned on fixed effects to remove time series and location-specific confounders. They were also conditioned on realized temperature and rainfall to account for potentially correlated effects of weather on both mortality and forecast error.

In Figure 9, the leftmost point and whiskers reproduce the baseline estimate with a quadratic specification for the forecast error function. For legibility, we report the estimate of the 4-day cumulative effect on the mortality rate of a 1° increase in the standard deviation of forecast error. As Figure 7 shows, the effect is roughly constant after a few days, so these estimates are indicative of robustness across the range of cumulative effects we consider.

The left panel of Figure 9 varies the controls and other aspects of the baseline specification. The second through fourth points remove or add additional terms to the time trend control. In all cases, the point estimate is essentially unchanged. Including at least linear trends is helpful for improving precision.

The fifth point removes age interactions. In the baseline specification, age was interacted with the month fixed effects, following previous literature on the temperature-mortality relationship. Here the age controls have little effect. Similarly, making the seasonal controls more granular by replacing month fixed effects with day-of-year fixed effects (in the sixth

point) does not change the point estimate but does reduce precision.

The rightmost point in the left panel removes county population weighting. The baseline results weight by population in order to better represent the forecast effects for a representative individual in the U.S. and also to improve precision. The robustness check suggests that this last motivation is especially well-founded.

The right panel considers the effect of adding additional weather controls. For all weather controls, we include the level and three lags of nonparametric functions of the weather variables. The specifics of the functions for each variable are described below. The number of lags matches the lag length of the forecast error variables.

The first two robustness checks address the possibility that unobserved variation in wind may correlate with forecast errors. We here include wind data from NOAA's North American Regional Reanalysis (NARR) dataset. Like PRISM, the NARR dataset combines individual weather observations with a model (in this case, the NCEP Eta weather model) to produce weather measures on a consistent grid across the U.S (Mesinger et al., 2006). The grid has a spatial dimension of roughly 32km, and we take a spatial average of the values in each county to match our estimation sample.[32] Because the wind data grid is coarser than the PRISM grid, we lose some county observations when we include wind. Therefore, the first point in the second panel re-estimates our baseline effect on the sample for which we observe wind information. The effect is slightly, but not significantly, smaller than the overall baseline estimate.

The second point adds nonparametric functions of wind speed and direction. For wind speed, we include the level and lags of 4 indicators for the quartiles of speed. For wind direction, we include 8 indicators, again divided by quantiles. Comparing the second estimate to the first, one can see that the point estimate and inference are unchanged by the inclusion of these wind controls.

The third estimate returns to the baseline sample and adds eight quantile-based indicators for both daily average PM2.5 and ozone. Pollution has a direct effect on mortality, and weather conditions affect both the creation and the dispersal of some pollutants. Including these controls increases the point estimate, but the difference from the baseline estimate is minor. A similar result holds when including 4 bins of heat index to account for humidity.

The baseline controls include functions of realized temperature and rain. The second-to-last estimate interacts these two functions. This addition has no impact on the estimated forecast error effect. The final point adds all of the previously mentioned weather controls simultaneously. Because this set of controls also includes wind, the estimate is based on the smaller wind sample. We again find similar results to the baseline case.

[32]Further details on the steps we follow to process the wind data can be found in Missirian (2020).

5.5 Heterogeneity

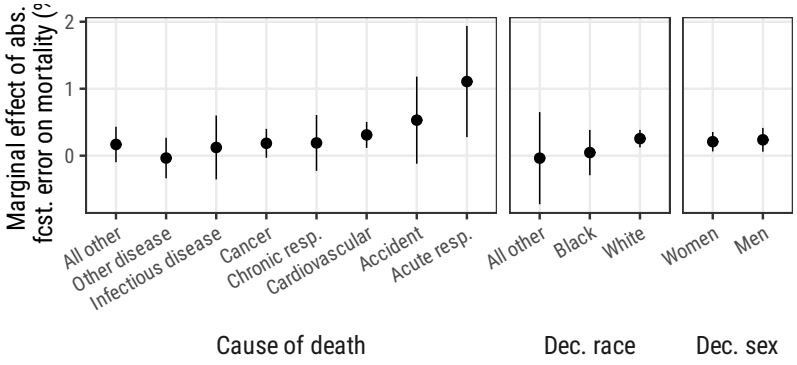
To analyze heterogeneity and mechanisms, we use two variations of the baseline estimating equation. One type of heterogeneity analysis varies the dependent variable—replacing it with mortality for specific subpopulations. Otherwise the estimating equation is identical to equation (2). We report heterogeneity by the each of the characteristics of the deceased individual that are recorded in the CDC mortality records.

The second set of estimating equations interacts temperature, forecast error, or both with a heterogeneity measure observed at the county level. These estimating equations take the form

$$y_{ct} = \sum_{\ell=0}^{L} \left[f_\ell(T_{c,t-\ell}) + g_\ell(e_{c,t-\ell}) + h_\ell(e_{c,t-\ell}) Z_{ct} + \delta_\ell \text{prec}_{c,t-\ell} \right] + X_{ct}\gamma + \alpha_{cm} + \rho_t + \varepsilon_{3,ct} \quad (4)$$

where all variables are the same as in equation (2) except for the addition of a heterogeneity measure Z_{ct} that is included in X_{ct} and the interaction between the heterogeneity measure and the forecast error function. The effect of the heterogeneity term can be assessed by comparing the g and h estimates. For example, if a given term attenuates the effect of forecast errors, the marginal effect of g would be larger than the marginal effect of g and h combined.

Figure 10: Heterogeneity by Cause of Death and Demographics



Notes: The figure shows the 4-day cumulative percent reduction in the mortality rate from a 1°C reduction in forecast absolute error based on thirteen separate models fit using Equation (4) on the baseline data. The forecast error function is a quadratic. The dimension of heterogeneity is indicated below each figure. The circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level.

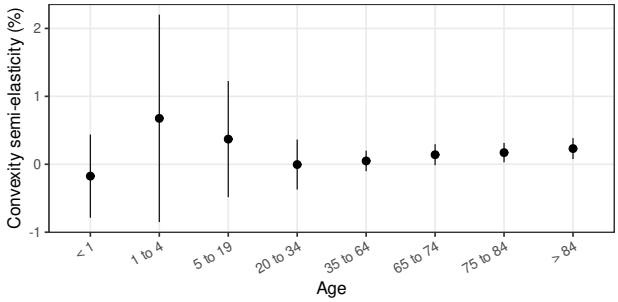
Figure 10 shows heterogeneity results along 3 different dimensions. Each of these dimen-

sions are variables that are available in the CDC mortality records, and we vary the baseline estimating equation by replacing the mortality rate with the specific mortality for the given group. The left panel breaks down value by cause of death. In terms of point estimates, the three causes that are most strongly associated with forecasts are acute respiratory failure, accidents, and cardiovascular disease. The strong associations with respiratory and cardiovascular deaths is consistent with the findings on leading causes of death from temperature exposure (Deschênes and Moretti, 2009). The higher association with accidents could be due to avoidance behavior engaged in by individuals to try to reduce their exposure to extreme weather. The other causes of death have more modest associations.

The middle panel compares forecast value by the race of the deceased. The effect of forecasts on mortality is substantially greater for white individuals than for all other individuals. Notably, the point estimates indicate that forecasts have close to zero effect on mortality for Black individuals and have a slightly negative value for individuals of all other races. These results suggest that different groups either are differentially aware of forecasts, differentially trust forecasts, differentially value mortality reductions, or differentially have scope to act on forecasts. The first two possibilities suggest benefits from improved communication and outreach, whereas the fourth suggests benefits from loosening constraints on adaptation.

Finally, the right panel decomposes forecasts' effect by the sex of the deceased individual. Across both men and women, the effect of forecasts is nearly indistinguishable.

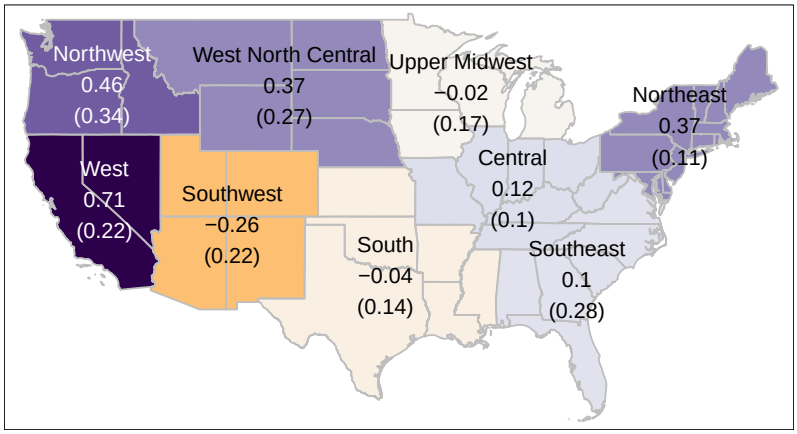
Figure 11: Heterogeneity by Age



Notes: The figure shows the 4-day cumulative percent reduction in the mortality rate from a 1°C reduction in forecast absolute error based on separate models fit using Equation (4) on the baseline data. The forecast error function is a quadratic. The dimension of heterogeneity is indicated below the figure. The circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level.

Figure 11 looks at effects by the age of the deceased. The most precise, strong effects come from older individuals,

Figure 12: Spatial Heterogeneity in Forecast Effect



Notes: The figure shows the 4-day cumulative percent reduction in the mortality rate from a 1°C reduction in forecast absolute error based on Equation (4) estimated across NOAA climate regions. The forecast error function is a quadratic. Darker purple colors indicate stronger benefits from forecast improvements, while lighter, orange colors indicate lower benefits. The value in parentheses is the standard error, calculated using the delta method. Confidence intervals for the estimates can be found in Figure A8

Figure 12 shows heterogeneous effects of forecasts across NOAA climate regions.[33] Forecasts have the largest effects on mortality in the West, Northwest, West North Central, and Northeast regions, with the effects in the West and Northeast being particularly precise. Point estimates in the Central, Southeast, Upper Midwest, and South regions are close to zero. Although not significantly different from zero, the Southwest is notable for a substantial negative forecast effect. Taken literally, the point estimate would indicate that reductions in forecast absolute error would lead to worse mortality outcomes in the region.

6 Conclusion

Routine weather forecasts are a widely used, highly sophisticated prediction product that most people interact with on a daily basis. Despite the ubiquity of weather forecasts, the number of people who rely on them, and the global effort involved in their production, surprisingly little is known about their economic value. Existing estimates are based on out-of-date stated-preference surveys or modeling exercises that might not capture real-world behavior. This paper provides the first revealed preference estimates of the value of daily weather forecasts.

We show that whether improving forecasts' accuracy reduces mortality is theoretically

[33]The information is also presented in the same format as Figure 10 in Figure A8.

ambiguous, depending on the convexity of mortality risk in forecast errors and thereby on the form of adaptation undertaken. Using the universe of mortality events and weather forecasts for a twelve-year period in the U.S., we show that forecasts are effective at helping people avoid mortality. Across the full temperature distribution, halving the standard deviation of day-ahead forecasts' errors would provide benefits of $18 billion per year in avoided mortality. This value stands out when we consider the $1.1 billion annual budget of the U.S. National Weather Service and even the $5 billion annual budget of its parent agency, the National Oceanic and Atmospheric Administration.

The majority of the value from improving forecasts' accuracy comes from days with relatively mild weather because these days are especially frequent. However, days with extreme cold and heat also provide substantial value, despite extremely cold days representing only 9% of days and extremely hot days representing only 2% of days. The value of improvements on extremely cold days derives from their current forecasts being especially noisy, and the value of improvements on extremely hot days derives from mortality being especially convex in forecast errors. This convexity suggests that adaptation is particularly important on extremely hot days. As global warming increases temperatures in the U.S., high-quality weather forecasts will be even more important for avoiding excess mortality. However, the scope for improvements is limited by the high quality of existing forecasts. Future work should analyze the benefits of improving accuracy at longer horizons. It should also investigate the value of forecast-based adaptation facilitated by the public sector.

Our analysis identifies the value of more accurate forecasts through the effects of idiosyncratic forecast errors on mortality. When we value improved accuracy, we implicitly hold agents' responses to forecasts fixed. This approach may be sensible for marginal changes in forecast quality, but nonmarginal changes should eventually change the way people use forecasts. Our analysis suggests that people do in fact act on forecasts more in counties where forecasts tend to be of higher quality. Our estimates may therefore represent a lower bound on the mortality value of increased accuracy. Future work should use variation in forecast quality to identify how people would respond to improved forecasts.

References

Anderson, B. G. and M. L. Bell (2009). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology 20*(2), 205–213.

Andersson, H. and N. Treich (2011). The value of a statistical life. In A. de Palma, E. Quinet, and R. Vickerman (Eds.), *A Handbook of Transport Economics*. Elgar.

Bakkensen, L. (2016). Public versus private risk information: Evidence from U.S. tornadoes. Working paper, University of Arizona.

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.

Barwick, P. J., S. Li, D. Rao, and N. B. Zahur (2018). The morbidity cost of air pollution: evidence from consumer spending in china. Technical report, National Bureau of Economic Research.

Bauer, P., A. Thorpe, and G. Brunet (2015). The quiet revolution of numerical weather prediction. *Nature 525*(7567), 47–55.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources 50*(2), 317–372.

Carleton, T. A., A. Jina, M. T. Delgado, M. Greenstone, T. Houser, S. M. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. B. Nath, et al. (2020). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. Technical report, National Bureau of Economic Research.

Chapman, R. E. (1992, July). Benefit-cost analysis for the modernisation and associated restructuring of the National Weather Service. NISTIR 4867, National Institute of Standards and Technology.

CIESIN (2017). U.S. Census grids 2010. https://sedac.ciesin.columbia.edu/data/collection/gpw-v4. Accessed: 2018-08-30.

Craft, E. D. (1998). The value of weather information services for nineteenth-century great lakes shipping. *American Economic Review 88*(5), 1059–1076.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics 3*(4), 152–185.

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

Drèze, J. (1962). L'utilité sociale d'une vie humaine. *Revue Franĉaise de Recherche Opérationnelle 23*, 93.

EPA (2021). Mortality risk valuation. https://www.epa.gov/environmental-economics/mortality-risk-valuation. Accessed: 2021-12-10.

Fishback, P. V., W. Troesken, T. Kollmann, M. Haines, P. W. Rhode, and M. Thomasson (2011, May). Information and the impact of climate and weather on mortality rates during the Great Depression. In G. D. Libecap and R. H. Steckel (Eds.), *The Economics of Climate Change: Adaptations Past and Present*, pp. 131–167.

Freebairn, J. W. and J. W. Zillman (2002). Economic benefits of meteorological services. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling 9*(1), 33–44.

Gasparrini, A., B. Armstrong, and M. G. Kenward (2010). Distributed lag non-linear models. *Statistics in medicine 29*(21), 2224–2234.

Gasparrini, A., Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz, A. Tobias, S. Tong, J. Rocklöv, B. Forsberg, et al. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The lancet 386*(9991), 369–375.

Glahn, H. R. and D. P. Ruth (2003). The new digital forecast database of the National Weather Service. *Bulletin of the American Meteorological Society 84*(2), 195–202.

Heutel, G., N. H. Miller, and D. Molitor (2021). Adaptation and the mortality effects of temperature across us climate regions. *Review of Economics and Statistics 103*(4), 740–753.

Hooke, W. H. and R. A. Pielke Jr. (2000). Short-term weather prediction: An orchestra in search of a conductor. In D. Sarewitz, R. A. Pielke Jr., and R. Byerly (Eds.), *Prediction: Science, Decision Making and the Future of Nature*, pp. 61–83. Island Press.

Jones-Lee, M. (1974, July). The value of changes in the probability of death or injury. *Journal of Political Economy 82*(4), 835–849. Publisher: The University of Chicago Press.

Katz, R. W. and J. K. Lazo (2011, July). Economic value of weather and climate forecasts. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Kruttli, M., B. Roth Tran, and S. W. Watugala (2019, May). Pricing Poseidon: Extreme weather uncertainty and firm return dynamics. SSRN Scholarly Paper ID 3284517, Social Science Research Network, Rochester, NY.

Lemoine, D. (2021). Estimating the consequences of climate change from variation in weather. Working Paper 25008, National Bureau of Economic Research.

Letson, D., D. S. Sutter, and J. K. Lazo (2007, August). Economic value of hurricane forecasts: An overview and research needs. *Natural Hazards Review 8*(3), 78–86.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, et al. (2006). North american regional reanalysis. *Bulletin of the American Meteorological Society 87*(3), 343–360.

Meza, F. J., J. W. Hansen, and D. Osgood (2008, May). Economic value of seasonal climate forecasts for agriculture: review of ex-ante assessments and recommendations for future research. *Journal of Applied Meteorology and Climatology 47*(5), 1269–1286.

Miller, B. M. (2018, January). The not-so-marginal value of weather warning systems. *Weather, Climate, and Society 10*(1), 89–101.

Millner, A. and D. Heyen (2021). Prediction: The long and the short of it. *American Economic Journal: Microeconomics 13*(1), 374–398.

Missirian, A. (2020). Yes, in your backyard: Forced technological adoption and spatial externalities. *Working Paper*, 1–64.

Morss, R. E., J. K. Lazo, B. G. Brown, H. E. Brooks, P. T. Ganderton, and B. N. Mills (2008, March). Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bulletin of the American Meteorological Society 89*(3), 335–346.

Myrick, D. T. and J. D. Horel (2006). Verification of surface temperature forecasts from the National Digital Forecast Database over the Western United States. *Weather and Forecasting 21*(5), 869–892.

National Research Council (1998). *The Atmospheric Sciences: Entering the Twenty-First Century*. Washington, DC: The National Academies Press.

National Research Council (2010, July). *When Weather Matters: Science and Services to Meet Critical Societal Needs*. Washington, DC: The National Academies Press.

National Research Council (2012). *The National Weather Service Modernization and Associated Restructuring: A Retrospective Assessment*. Washington, DC: The National Academies Press.

NOAA (2021). NOAA Blue Book FY2022. Technical report, National Oceanic and Atmospheric Administration.

NOAA (2021). NOAA FY2021 budget blue book summary. Date accessed: 2021-03-26.

Pielke, R. and R. E. Carbone (2002, March). Weather impacts, forecasts, and policy: An integrated perspective. *Bulletin of the American Meteorological Society 83*(3), 393–403.

PRISM Climate Group (2004). http://prism.oregonstate.edu.

Rosenzweig, M. R. and C. R. Udry (2019). Assessing the benefits of long-run weather forecasting for the rural poor: Farmer investments and worker migration in a dynamic equilibrium model. Technical report, National Bureau of Economic Research.

Schlenker, W. and M. J. Roberts (2006). Nonlinear effects of weather on corn yields. *Review of agricultural economics 28*(3), 391–398.

Schlenker, W. and M. J. Roberts (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences 106*(37), 15594–15598.

Shrader, J. (2021). Expectations and adaptation to environmental risks. Working paper.

Simon, N. B., C. Dockins, K. B. Maguire, S. C. Newbold, A. J. Krupnick, and L. O. Taylor (2019). Policy brief—what's in a name? a search for alternatives to "vsl". *Review of environmental economics and policy 13*(1), 155–161.

Stratus Consulting Inc. (2002). Economic value of current improved weather forecasts in the U.S. household sector. Technical report.

Sutter, D. and B. T. Ewing (2016). State of knowledge of economic value of current and improved hurricane forecasts. *Journal of Business Valuation and Economic Loss Analysis 11*(1), 45–64.

Toth, Z. and R. Buizza (2019). Weather forecasting: What sets the forecast skill horizon? In *Sub-Seasonal to Seasonal Prediction*, pp. 17–45. Elsevier.

Viscusi, W. K. (1993). The value of risks to life and health. *Journal of Economic Literature 31*(4), 1912–1946. Publisher: American Economic Association.

Weinberger, K. R., A. Zanobetti, J. Schwartz, and G. A. Wellenius (2018). Effectiveness of national weather service heat alerts in preventing mortality in 20 us cities. *Environment international 116*, 30–38.

WMO, W. Bank, GFDRR, and USAID (2015). Valuing Weather and Climate: Economic Assessment of Meteorological and Hydrological Services. WMO 1153, World Meteorological Organization, World Bank, Global Facility for Disaster Reduction and Recovery, United States Agency for International Development, Geneva, Switzerland.

Appendix for online publication

A Data Processing Details

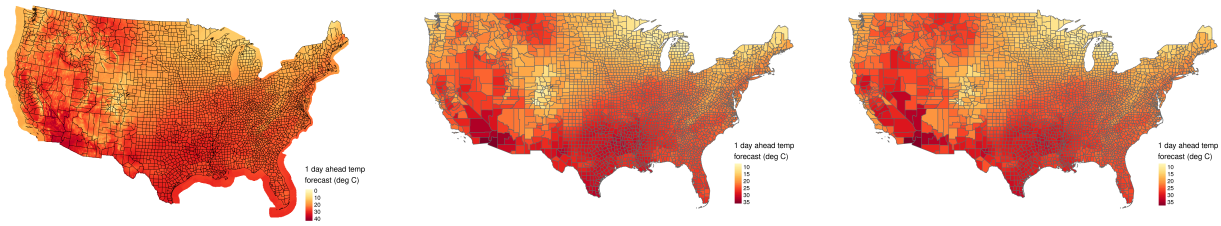
The raw mortality data from the CDC NCHS MCOD files report the day and county of each vital event. All events in a county on a day are added together to generate the county-level number of daily deaths. The deaths are translated into a death rate by dividing by annual county population, as described in Section 3.2.

The day-county structure of the mortality data motivates our processing choices for the PRISM weather data and NDFD forecast data. Both datasets originally provide daily observations on a consistent, high-resolution spatial grid across the U.S. For the forecast data, there are multiple potential observations per day. Forecast models are run and results are reported multiple times per day. For the minimum and maximum temperature forecasts we focus on, the major model runs occur at 12UTC and 00UTC. Based on feedback from National Weather Service meteorologists, we examine the 12UTC forecast which are available in the early morning for locations in the U.S. and typically form the basis of the morning forecast on local news. We aggregate the spatial grid to the county level using the following procedure:

First, for each county, we find the spatial points that fall inside the geographic boundary of the county, using 2010 county TIGER/Line shapefile from the Census. Given the high resolution of the underlying datasets, all counties in our sample contain multiple grid points.

Second, we assign a weight to each grid point based on 2010 population grids from CIESIN (2017). The CIESIN grids are at a roughly 1km resolution, which is higher than either the weather or forecast grid resolutions. Therefore, we use bilinear resampling to reproject the the population grid to match that of the weather or forecast grids.

Figure A1: Comparison of Example Raw Gridded Forecast Data and County-level Data



(a) Gridded Raw Data (b) County Area Average (c) Population-weighted

Notes: The maps show the raw, gridded forecast data in panel (a) and the corresponding county-level area and population-weighted average forecasts in panels (b) and (c) respectively. The maps are for one day and forecast horizon: the 1-day-ahead forecast for September 9, 2006. The forecast shown in the figure is the 00UTC forecast of hourly temperature. In practice, we focus on the 12UTC forecast of minimum and maximum temperature.

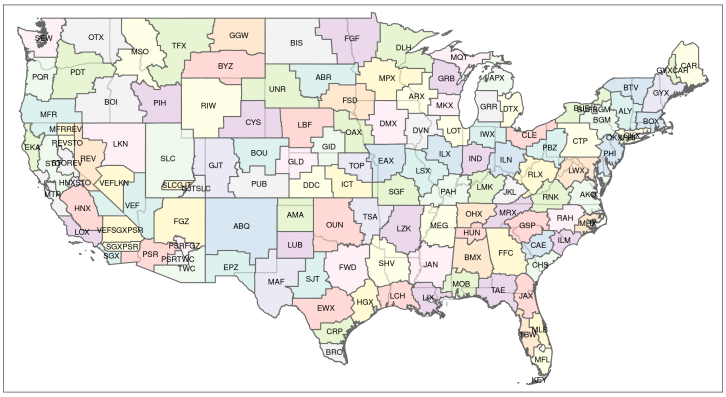
Third, we calculate population-weighted average values for each weather or forecast observation within the county. The end result is a daily, population-weighted spatial average of the maximum temperature, minimum temperature, total precipitation, dewpoint temperature, maximum temperature forecast for 1 to 6 days ahead, and the minimum temperature forecast for 1 to 6 days ahead (the NWS issues forecasts out to 7 days, but given our choice of the 12UTC forecast, the 7-day-ahead minimum temperature forecast is not available). Comparison of an example gridded forecast data and the corresponding county-level data is shown in Figure A1.

Fourth, we correct errors in the forecast data. The NDFD data undergo minimal error checking (such checks are the responsibility of local Weather Forecast Offices). In particular, from one forecast horizon to the next, there are a small number of observations that have a change in forecast value of exactly -17.4999 degrees. These changes occur only at one forecast horizon, so we use adjacent forecast horizons to interpolate the erroneous value. We do the same for positive forecast errors that occur for just a single forecast horizon and are greater than 25 degrees in absolute value (in the primary results, we Windsorize the forecast errors, so this data cleaning step does not affect the estimates).

Fourth, we match the timing conventions in the forecast and weather data. The NWS typically uses a noon to noon UTC convention for daily temperature forecasts. Minimum temperatures are forecasted for the nighttime (midnight UTC day to noon UTC day t or 7 p.m day $t-1$ to 7 a.m. day t EST). Maximum temperatures are forecasted for the daytime (noon UTC to midnight UTC). PRISM also typically follows this timing convention, but not as strictly. To match the timing conventions between the two datasets, for maximum and minimum temperatures separately, we regress realized temperature on the day t 1-day-ahead forecast and the day $t-1$ 1-day-ahead forecast. For maximum temperature, we find that the day t forecast is sufficient (the day $t-1$ forecast does not predict the realization conditional on the day t forecast). For minimum temperature, we find that both days' forecasts are predictive, with the day t forecast being about twice as predictive as the day $t-1$ forecast. We therefore construct a time-corrected day t minimum temperature forecast that is the weighted average of the original day t and day $t-1$ forecasts with weight 2/3 on the day t forecast and 1/3 on the day $t-1$ forecast. The time-corrected forecast does exhibit forecast sufficiency.

After creating the daily, county-level dataset, we merge counties with identifiers for their NOAA County Warning Area (CWA). CWAs are collections of counties, and the local NWS Weather Forecasting Office (WFO) is responsible for generating forecasts for the CWA. The map of counties and CWAs is shown in Figure A2.

Figure A2: Map of NOAA County Warning Areas (CWAs)



Notes: The map shows (in colored areas with black outlines) the geographic boundaries of County Warning Areas (CWAs), the collection of counties for which a given NWS Weather Forecasting Office is responsible for creating forecasts. State borders are shown in gray, thinner lines. CWAs are typically composed of one or more counties and can cross state borders. There are 116 CWAs in the continental U.S. Some counties are part of multiple CWAs, and in those cases, we assign the county a CWA ID composed of each CWA that it is in. The end result is 1 to 1 mapping of all continental U.S. counties to 130 CWAs or CWA groups.

B Counterfactual Approximation Quality

In Section 2.1, we derive marginal conditions for forecast value. These conditions are second-order approximations to the value of a change in forecast error distribution. The value of a discrete change in error distribution is

$$\mathbb{E}[V] = \int h_e(\hat{T}, e) p^c(e) \, \mathrm{d}e - \int h_e(\hat{T}, e) p(e) \, \mathrm{d}e \tag{5}$$

where $p(e)$ is the probability density of forecast errors and $p^c(e)$ is a counterfactual density. For instance, it could be the density with a 50% smaller standard deviation.

The approximation will be accurate if the distribution of errors is close to normal or if the mortality hazard function is approximately quadratic in forecast errors. The approximation is practically useful because it is faster to compute. Table A1 compares the estimated counterfactual across all realized temperatures using both the approximation and a nonparametric estimate based on numerically evaluating the two integrals in Equation (5). One can see that the approximation accurately reproduces the results from the nonparametric estimator in this setting.

Note that this comparison is done by estimating versions of Equation (2) with only 4 total lags of forecast error, temperature, and precipitation to speed up computation. The estimates values are four-day cumulative effects by will differ from the estimates presented in the body of the paper based on longer lag models.

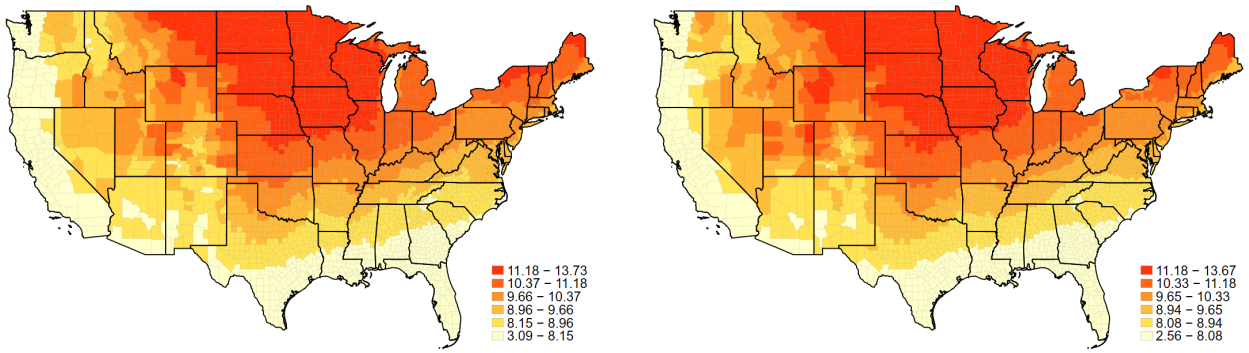
Table A1: Comparison of Counterfactual Approximation and Nonparametric Calculation

Method	(1) Marginal change in s.d.	(2) % reduction in s.d.
Nonparametric	37.192	19.676
Approximation	37.661	19.684

Notes: The table compares estimates of the counterfactual value of forecast error changes calculated using a second-order approximation given in Section 2 (Column 1) and a nonparametric estimate based on moment conditions (Column 2). The counterfactual values differ from the results shown in the body of the paper (first row of Table 3) because they are based on estimating only a 4-day lag model to speed computation.

C Additional Figures and Tables

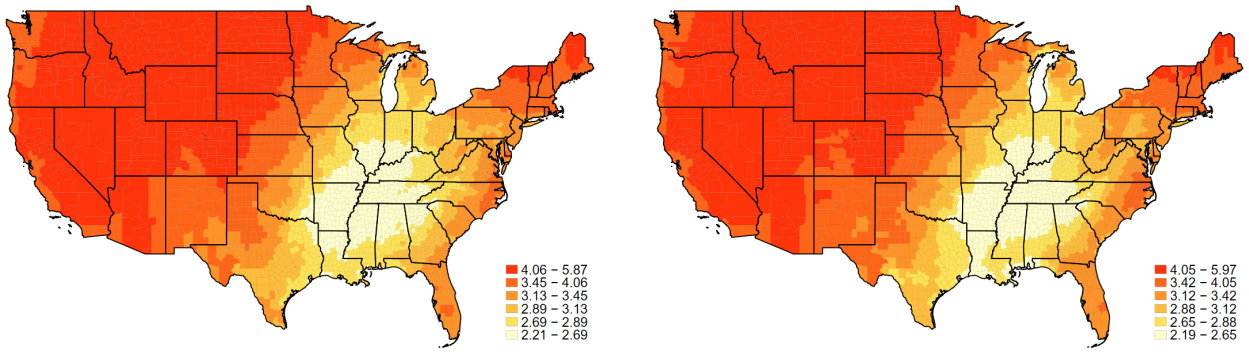
Figure A3: Unconditional Variation in Temperature and Day-ahead Forecast



(a) St. dev. of average temperature (b) St. dev. of 1-day-ahead forecast

Notes: The maps show the standard deviation of the unconditional average temperature (left panel) or the 1-day-ahead forecast of average temperature (right panel). For an indication of the identifying variation conditional on controls, compare these maps to the maps in Figure A4.

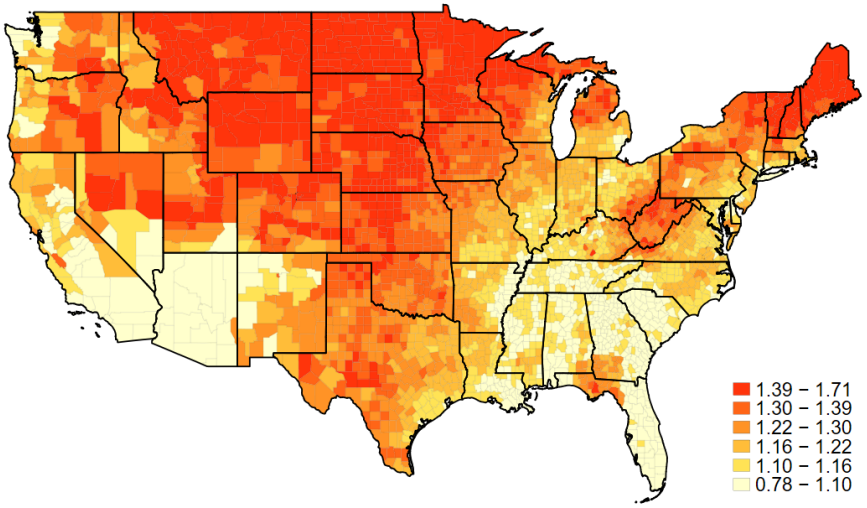
Figure A4: Residual Variation in Temperature and Day-ahead Forecast



(a) St. dev. of residual average temperature (b) St. dev. of residual 1-day-ahead forecast

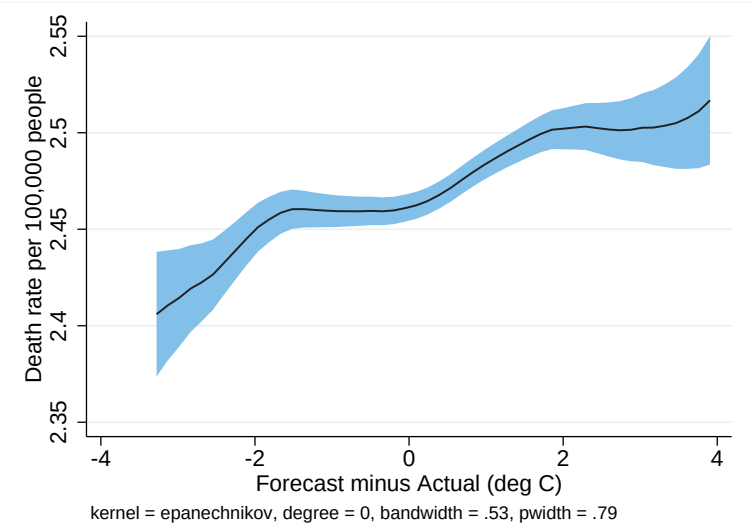
Notes: The maps show the standard deviation of the residuals from a regression of average temperature (left panel) or the 1-day-ahead forecast of average temperature (right panel) on all of the controls in the baseline regression specification (see Equation **??**).

Figure A5: Spatial Variation in Forecast RMSE (Conditional on Baseline Controls)

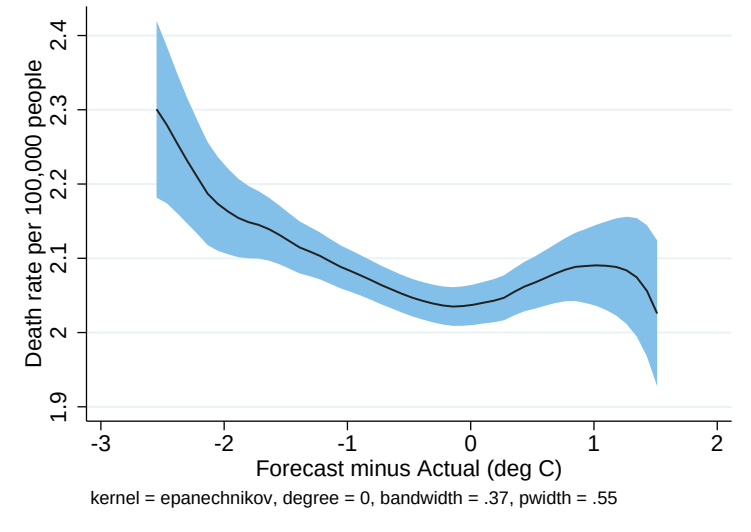


Notes: The map shows the root mean squared error of the 1-day-ahead forecast for each county in the continental U.S. over the sample period. Redder values indicate higher average RMSE and yellower values indicate lower values. The values are all conditional on the baseline fixed effects and other control variables.

Figure A6: Raw Data Relationship Between 1-Day Ahead Forecast Error and Mortality for Days with Hot and Cold Realized Temperatures



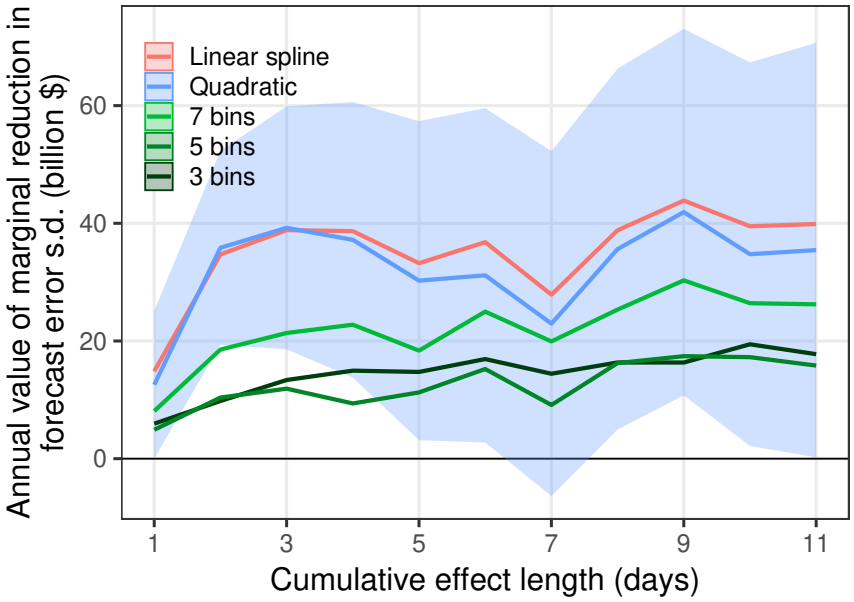
(a) Realized temperature ≤ 5°C



(b) Realized temperature ≥ 30°C

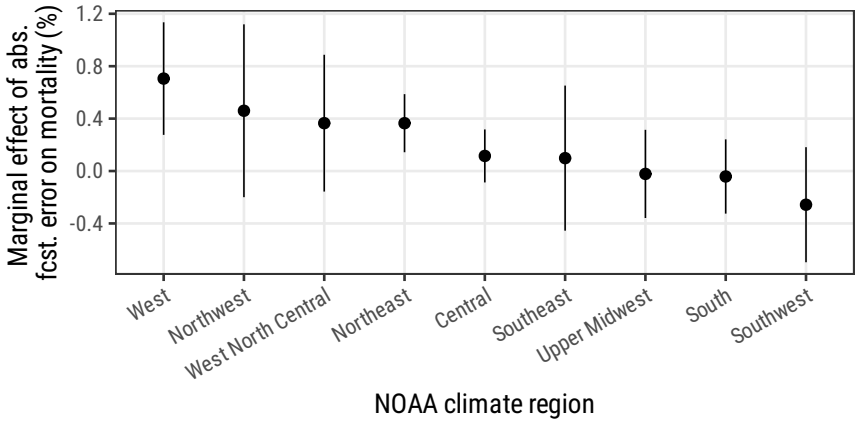
Notes: The figures show estimates of the raw-data relationship between forecast error from the 1-day ahead forecast and the daily mortality rate. The fitted lines are based on local polynomial regressions with an Epanechnikov kernel and the default, plugin bandwidth. The shaded areas are 95% confidence intervals but should be interpreted with caution because they treat all observations as i.i.d. Clustered inference is presented in Section 5.5. Panel (a) shows the relationship when the actual temperature turns out to be cold (≤ 5°C), and Panel (b) shows the relationship when the actual temperature turns out to be hot (≥ 30°C).

Figure A7: Comparing Value Estimates: Binned Forecast Error Specifications



Notes: The figure shows estimates of the cumulative value of a 1°C reduction in mean absolute forecast error. Each line is based on a single regression estimated using Equation 2 fit to the baseline data. The estimates are translated into economic value by calculating the counterfactual value of a 1° reduction in forecast absolute error multiplied by the VSL, population, and days per year. The red line estimates a quadratic relationship between mortality and forecast error, the blue line uses a linear spline with a knot at 0, and the green lines use progressively more granular quantile bins. To help assess the significance of the diffrences between the methods, the blue area shows the 95% confidence interval for the linear spline specification, calculated using the delta method, based on standard errors clustered at the CWA level.

Figure A8: Effect of Forecast Error Increase on Mortality: Regional Heterogeneity



Notes: The figure shows the 4-day cumulative percent increase in the mortality rate from a 1°C increase in forecast error standard deviation based on a model fit using Equation (4) on the baseline data. The forecast error function is a quadratic. The dimension of heterogeneity is indicators for each NOAA climate region in the Continental U.S. The circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level. A map showing these effects spatial can be found in Figure 12.