Accurate Weather Forecasts Become More Critical with Climate Change (Preliminary and Incomplete)

Laura Bakkensen, Derek Lemoine, Manuel Linsenmeier, Jeffrey G. Shrader, and Stephan Thies*

October 14, 2024

Abstract

People use weather forecasts to limit mortality from extreme temperatures, but relying on weather forecasts leaves them more exposed to errors in those forecasts. We show that the U.S. population especially relies on forecasts on hot days. Inaccurate forecasts cause 1,900 more deaths per year as climate change increases the frequency of hot days. However, if forecasts improve at their historical rate, then total mortality from forecast errors will fall by 1,500 lives per year over the century. Disentangling the effects of predicted and surprising heat, we project that climate change will have an insignificant effect on temperature-related mortality by the end of the century if forecasting does not improve but that it will significantly decrease temperature-related mortality if forecast accuracy improves at its historical rate. In ongoing work, we elicit experts' opinions about the future improvement in forecast accuracy.

*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); Linsenmeier: Climate School, Columbia University (email: mpl2157@columbia.edu); Shrader: School of International and Public Affairs, Columbia University and IZA (email: jgs2103@columbia.edu); Thies: School of International and Public Affairs, Columbia University and CEP, London School of Economics (email: stephan.thies@columbia.edu). The authors gratefully acknowledge NSF grant SES-2315555 for funding. All errors are our own.

1 Introduction

Extreme temperatures are already an important cause of mortality [1, 2, 3, 4]. Climate change is projected to increase mortality from hot days and to reduce mortality from cold days [5]. To date, such projections do not account for how the forecastability of hot and cold days may change over the century. When people know extreme weather is coming, they can use that information to adapt their plans to the weather [6, 7, 8, 9, 10]. People already use accurate weather forecasts to reduce mortality, especially on hot days [11]. And while forecasts have improved substantially over the last decades [12], forecast accuracy could still be improved in many regions even with current technology [13], and new investments [14, 15] and advancements in machine learning [16, 17] promise improvements even at the bleeding-edge. We show that by increasing the frequency of hot days, climate change makes future mortality especially sensitive to the quality of forecasting systems. Forecast errors cause more mortality under climate change, but policies that continue historical improvements in forecasting systems can reduce forecast errors by enough to offset that increase.

Figure 1 illustrates why it is important to disentangle the effects of well-forecasted weather from surprising weather. The three sets of people experience the same extreme heat, but only people of type A had received an accurate forecast of that extreme heat. Instead, people of type B had received a forecast of more typical heat, and people of type C had received a forecast of even more extreme heat. People of type B have the highest risk of death because they were not prepared for the extreme heat.

Today, there are more people of type B than of types A and C (top row). When we select for days that turn out to be extremely hot, we are selecting for days with the highest realized temperatures, which will also tend to be days that overshot their forecasts ex post. Days that undershot their forecasts tended to be cooler, so we see few people of type C. Regressing mortality on temperature without accounting for forecasts has two problems. First, it cannot tell us how many deaths are due to forecast errors rather than to high temperatures per se. Second, it estimates an effect of extreme heat that reflects the experience of people of type B, which can cause problems when projecting the effects of future climate change.

Climate change makes extreme heat more common. In the middle row, extreme heat becomes well-predicted on average, as it is no longer so extreme relative to the distribution of temperature. There are now as many people of type C as of type B, and there are also more people of type A, with accurate forecasts. A temperature-mortality relationship estimated from the top row, which is dominated by people of type B, would not correctly project the mortality rate in the middle row, with a larger share of types A and C. Moreover, prior work has shown that people rely on forecasts less in cold weather [11], so the tendency to

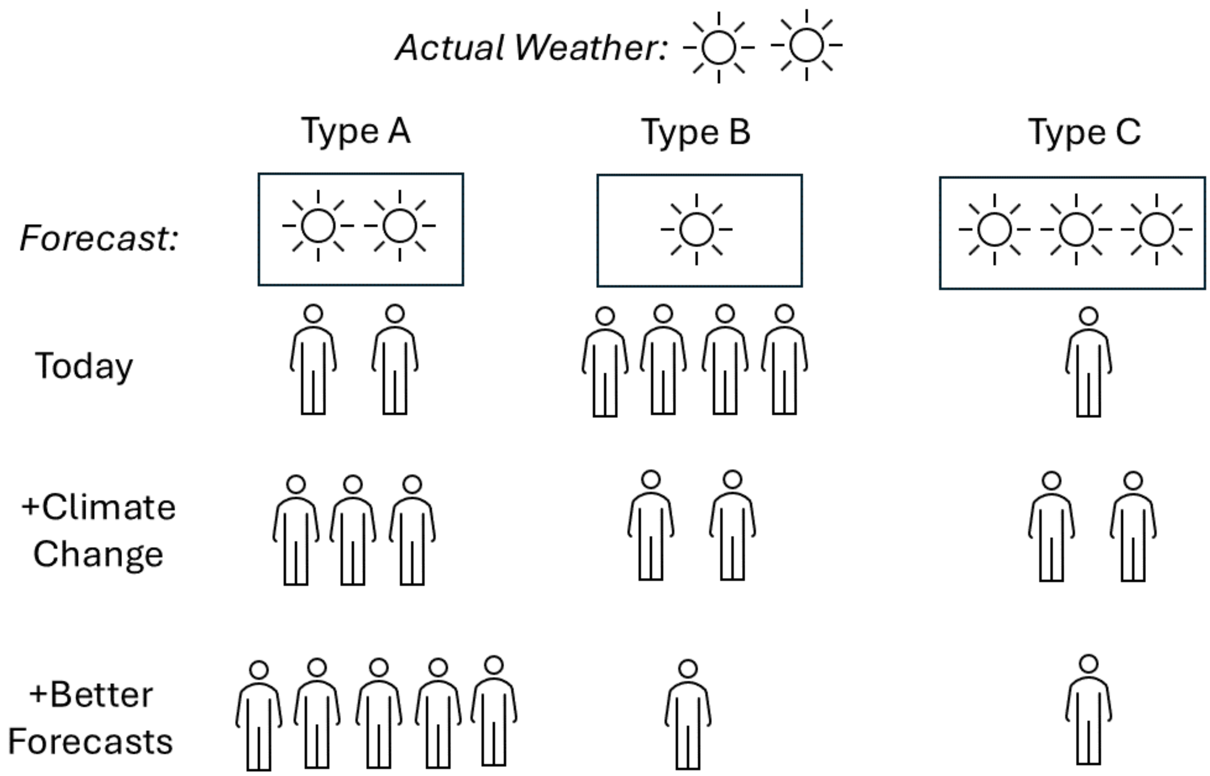

Figure 1: Extreme Weather May Be Forecasted Differently Today Versus in the Future Before a day with extreme heat, today some people receive accurate forecasts (type A), today many people receive too-mild forecasts (type B), and today few people receive too-extreme forecasts (type C). With climate change (middle row), extreme heat becomes more typical and there are equal types with too-mild and too-extreme forecasts. If, in addition, forecasts improve (bottom row), then even fewer people receive too-mild forecasts. If we estimate the effect of extreme temperatures using today's data but without disentangling forecasts from temperatures, then we misrepresent the effects of extreme temperatures in a world with climate change and, potentially, improved forecasts.

overestimate the effects of climate change on mortality will be greater for hot days than for cold days. We project mortality in a warming world by jointly estimating the effects of realized temperatures and forecast errors.

In addition, forecasts may continue improving over time. In such a world (bottom row), there would be relatively few people of type B—and thus fewer people who do not get to prepare for extreme heat. A projection based on the large number of people of type B in the top row could substantially overstate mortality from extreme heat. Moreover, people may come to rely on forecasts more as they improve. Then there are not only more people of type A but they also use their accurate forecasts more effectively, whereas people of type B become more exposed to their bad forecasts when they rely on them more. The net effects will depend on how many people shift from type B to type A and on how the consequences of weather and forecast errors both change when people come to rely on forecasts more. We project the benefits of forecasting improvements by jointly estimating the effects of realized temperatures and forecast errors.

We estimate how the effects of temperature on mortality vary with the 1-day-ahead forecast of that temperature (see Methods; note: we may change this to 3-day-ahead soon for consistency why the expert elicitation). We use a flexible specification that permits the relation between mortality, temperature, and forecasts to differ based on how hot or cold the temperature is and on the sign and magnitude of the forecast error. To this end, we combine data on monthly mortality in U.S. counties from 2005–2017 (soon to be updated through 2021) with daily data on temperature and day-ahead forecasts. We use county fixed effects to absorb persistent differences in counties over time, month-of-year fixed effects to absorb common health or economic shocks across counties, and county-month-specific time trends in order to absorb differences in counties' demographic and economic trajectories. We therefore estimate how temperatures that are unusual for that county and month affect excess mortality, and how that effect varies with the forecasts of those temperatures.

We also permit the effects of temperature and forecasts to vary with a location's characteristics that might affect how people use forecasts and adapt to temperatures. First, people in hotter climates may be better adapted to hot weather. Second, people in climates with more variable weather may invest in becoming resilient to weather and may rely on forecasts more. Third, people in locations that tend to have more accurate forecasts may choose to rely on those forecasts more. We examine whether each type of heterogeneity matters, holding the others constant.

2 Accurate Forecasts Reduce Mortality from Heat

We first show that excess mortality responds to both temperature realizations and temperature forecast errors. These results are shown in Figure 2 which reports results from a simplified specification that does not allow heterogeneity in the mortality-temperature-forecast relationship based on a location's characteristics. The left panel shows the estimated effect on mortality of various combinations of temperature (horizontal) and forecast errors (vertical). The yellow line in the middle depicts the average forecast error at each daily temperature. For moderate temperatures, this line tracks the zero forecast error line, indicating that these temperatures are accurately forecasted on average. At more extreme temperatures, the average forecast error becomes nonzero, with forecasts tending to be too warm (positive average error) when the weather is cold and forecasts tending to be too cool (negative average error) when the weather is hot. These average errors do not result from forecasters skewing their forecasts, as SI Figure S1 shows that the mean error is zero when we condition on the forecast rather than the temperature. In other words, forecasts are unbiased. Instead, these average errors reflect the forces described in the top row of Figure 1. When we study days that turned out to be extremely hot or cold, we are in part selecting for days in which forecasts were too mild and selecting against days in which forecasts were overly extreme.

The orange lines in Figure 2 depict a one standard deviation forecast error. The standard deviation is largest (around 1.25°C) on the coldest days, indicating that temperature forecasts tend to be less accurate on cold days. The standard deviation declines to a minimum of around 0.75°C near 30°C daily average temperature (the 99th percentile of days in our sample), indicating that temperature forecasts tend to be most accurate on relatively hot days.

The shading in Figure 2 reflects the estimated effect on mortality, from blue for a small effect to yellow for a large effect. At a given temperature (i.e., along a given vertical line), mortality is minimized around the average forecast error and thus around fairly accurate forecasts. One might think that whether inaccurate forecasts increase or decrease mortality depends on whether they are too extreme or too moderate, but we find that, at any given temperature, inaccurate forecasts in either direction tend to increase mortality. The effect is, however, asymmetric, with too-warm forecasts (positive forecast errors) especially hazardous in cold temperatures and too-cool forecasts (negative forecast errors) especially hazardous in hot temperatures.

The right panel of Figure 2 plots the effect of each temperature when perfectly forecasted (orange) and averaged over forecast errors of 1 standard deviation in either direction (red)

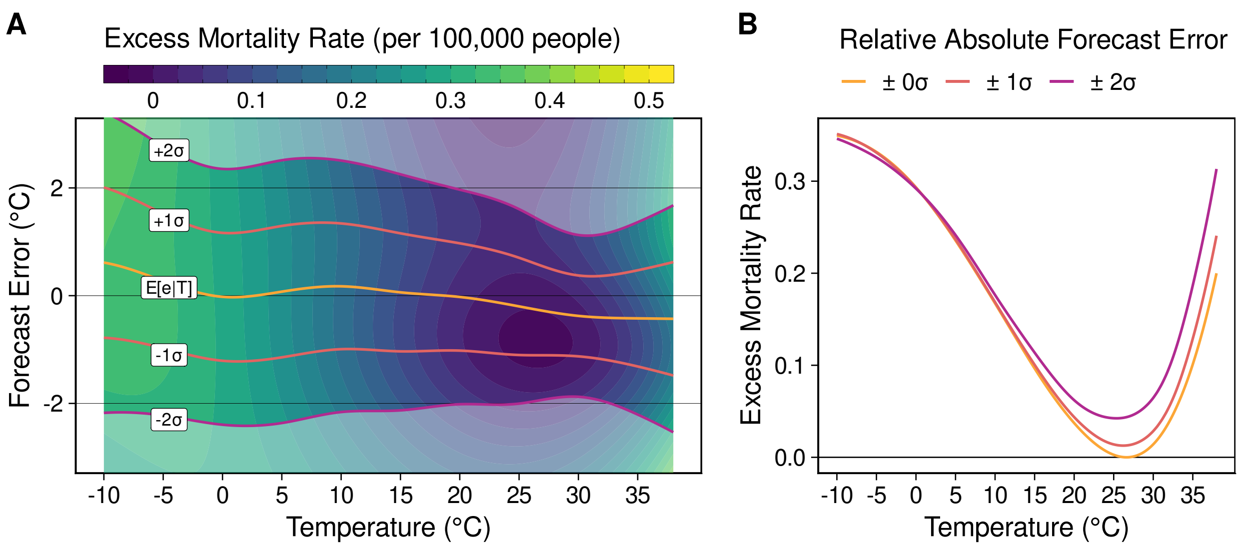

Figure 2: Mortality is affected by temperature and forecast errors. The figure shows the effect of realized temperature and forecast errors on the daily excess mortality rate (per 100,000 people). Panel A shows estimated excess mortality rate for every combination of realized temperature and forecast error. Lines colored from black to orange inform about the distribution of errors conditional on temperature and indicate 0, 1, and 2 standard deviations. Panel B plots estimated excess mortality rate against temperature when fixing absolute forecast errors to one or two conditional standard deviations. The curves correspond to a movement along the colored lines in panel A.

and of two standard deviations in either direction (purple). Averaging over either type of mistake reflects an ex ante assessment, in which we know that errors will happen but judge errors to be equally possible in either direction [11]. If too-moderate forecasts were harmful but too-extreme forecasts were beneficial, then averaging over positive and negative forecast errors may suggest effects that are similar to the case with accurate forecasts. However, because we find that errors of both types increase mortality, averaging over errors does increase mortality. We find that errors increase mortality by more in hot weather than in cold weather. And the magnitude of errors matters more in hot weather than in cold weather.

The mortality-minimizing temperature is at an average daily temperature of 25.7°C for well-forecasted temperatures, and is slightly lower for poorly forecasted temperatures. This temperature is fairly warm for the U.S., being at the 87th percentile of the distribution of days in our sample. A large proportion of mortality on days with this temperature or slightly warmer temperatures derives from inaccurate forecasts rather than the heat itself. In the U.S., cold days kill people at a high rate whether well-forecasted or not, whereas, for all but the very hottest days, it is surprising heat that kills rather than the heat itself.

3 People Use Forecasts Differently in Different Locations

Figure 3 shows how the mortality-temperature-forecast relationship depends on a location's climate and on the quality of its forecasts. The left column depicts the effect of perfectly forecasted temperatures on mortality. The reported p-values test for the significance of that dimension of heterogeneity across all temperatures.

Panel A1 shows that hot days cause fewer deaths in locations with hot climates and that cold days cause more deaths in those same locations. These differences are significant at the 1% level ($p = 0.001$). The mortality-minimizing temperature is 20.2°C in cold counties (the 66th percentile of U.S. temperatures) and 28.4°C in hot counties (the 95th percentile of U.S. temperatures). The dampened sensitivity of hot locations to hot days and the exaggerated sensitivity of hot locations to cold days reinforce prior work [5, 18]. These differences are consistent with adaptation to local climate, whether through biophysical channels or through conscious investments in public policy responses or private actions like housing choices, air conditioning, clothing, and lifestyles.

Panel B1 shows that mortality at a given temperature is lower in locations with high variability of day-to-day temperature ($p = 0.088$), and especially so at hot temperatures. These differences are consistent with people investing in weather resilience where the weather is more fickle. Panel C1 shows that mortality tends to also be lower in hot weather for locations

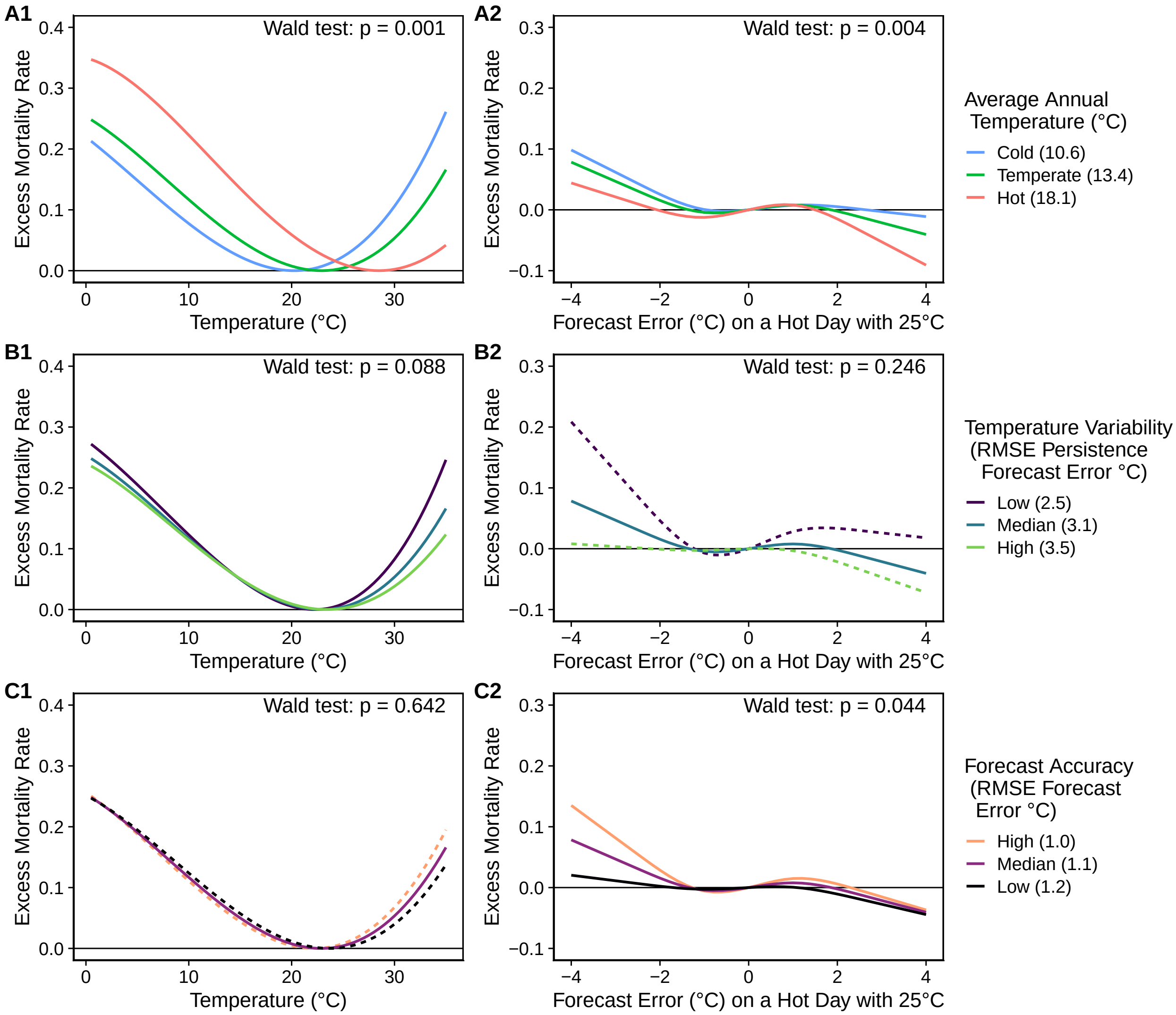

Figure 3: Mortality responses to temperatures and forecast errors are heterogeneous across counties. The figure shows cross-county heterogeneity in the estimated temperature mortality response (column 1) and the forecast error mortality response at 25°C (column 2) with respect to annual average temperatures (row A), temperature variability (row B), and forecast accuracy (row C). Each panel shows the predicted change in the response when setting one of the three variables to their first, second, and third quartiles while holding the other two at their respective medians. Wald test p-values indicate the significance of heterogeneity along each of the explored dimensions. Solid (dashed) lines indicate significant (insignificant) heterogeneity at the 10% level. In the first column, mortality rates per 100,000 people are expressed in excess of the respective minimum mortality rate. In the second column, mortality rates are expressed in excess of mortality at a temperature of 25°C and a forecast error of 0°C.

that tend to have less accurate forecasts, although this difference is both small and statistically insignificant ($p = 0.642$). This difference is again consistent with people investing to become more resilient to the weather where they find the weather to be less predictable.

The right column of Figure 3 depicts the relationship between forecast errors and mortality on a relatively hot day, with an average temperature of 25°C (84th percentile of U.S. days in the sample). Here, positive forecast errors (i.e., too-hot forecasts) tend to have negligible effects on mortality and perhaps even a beneficial effect on mortality. However, negative forecast errors (i.e., too-cool forecasts) tend to substantially increase mortality.

Panel A2 shows that a too-cool forecast is especially harmful in cool counties ($p = 0.004$). This result is consistent with people in those counties needing to prepare for heat. It is a priori unclear how day-to-day variability of temperature should affect the impact of forecast errors: on the one hand, people may rely less on forecasts when they invest in weather resilience, but on the other hand, forecasts of a given quality could add more value where the weather is hard to predict from recent experience. Panel B2 shows that the first effect dominates, as forecast errors cause much more mortality where temperatures are fairly stable from day to day. The point estimates suggest large effects, but they are not statistically significant ($p = 0.246$).

For a given level of day-to-day variability, Panel C2 shows that mortality from forecast errors is greater where forecasts tend to be more accurate ($p = 0.044$). This effect could reflect various differences between counties with more and less accurate forecasts or could reflect intentional investment in forecasting system upgrades. However, it is also consistent with people choosing to rely on forecasts more when they learn that forecasts tend to be good.

Putting the pieces together, we have seen that well-forecasted hot days are especially harmful in cold places with stable weather. We have also seen that forecast errors tend to be most harmful on hot days, and especially so where hot days are unusual (in cold locations with stable weather) and forecasts are typically accurate.

4 Climate Change Makes Inaccurate Forecasts More Hazardous

We now use the estimated response of mortality to temperature and forecast errors to project mortality from climate change in the U.S. These projections hold income, demographics, and technologies constant, as they maintain the mortality-temperature-forecast response surface estimated from 2005–2017 data (soon to be extended through 2021).

To project future temperatures we rely on gridded daily temperature output from the GFDL-

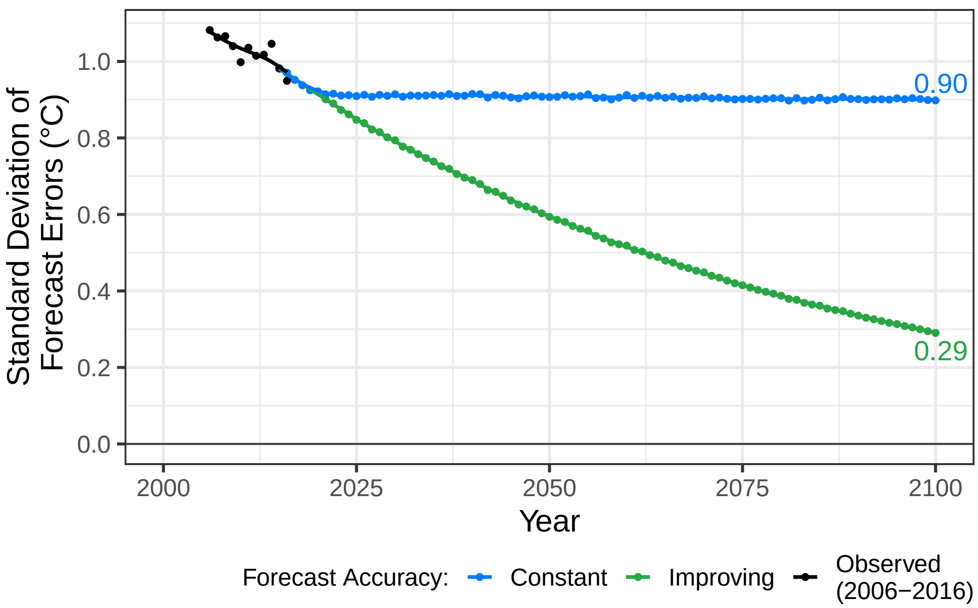

Figure 4: Temperature forecast accuracy could improve or stagnate in the future. The figure shows the annual average forecast accuracy observed in the US between 2006 and 2016 (black) as well as two potential future scenarios for the development of forecast accuracy. The improving scenario (green) assumes that the observed trend continues until the end of the century. The constant scenario (blue) holds forecast accuracy at its 2020 level. (*In work in progress, we survey professional weather forecasters to obtain their anticipated improvements.*)

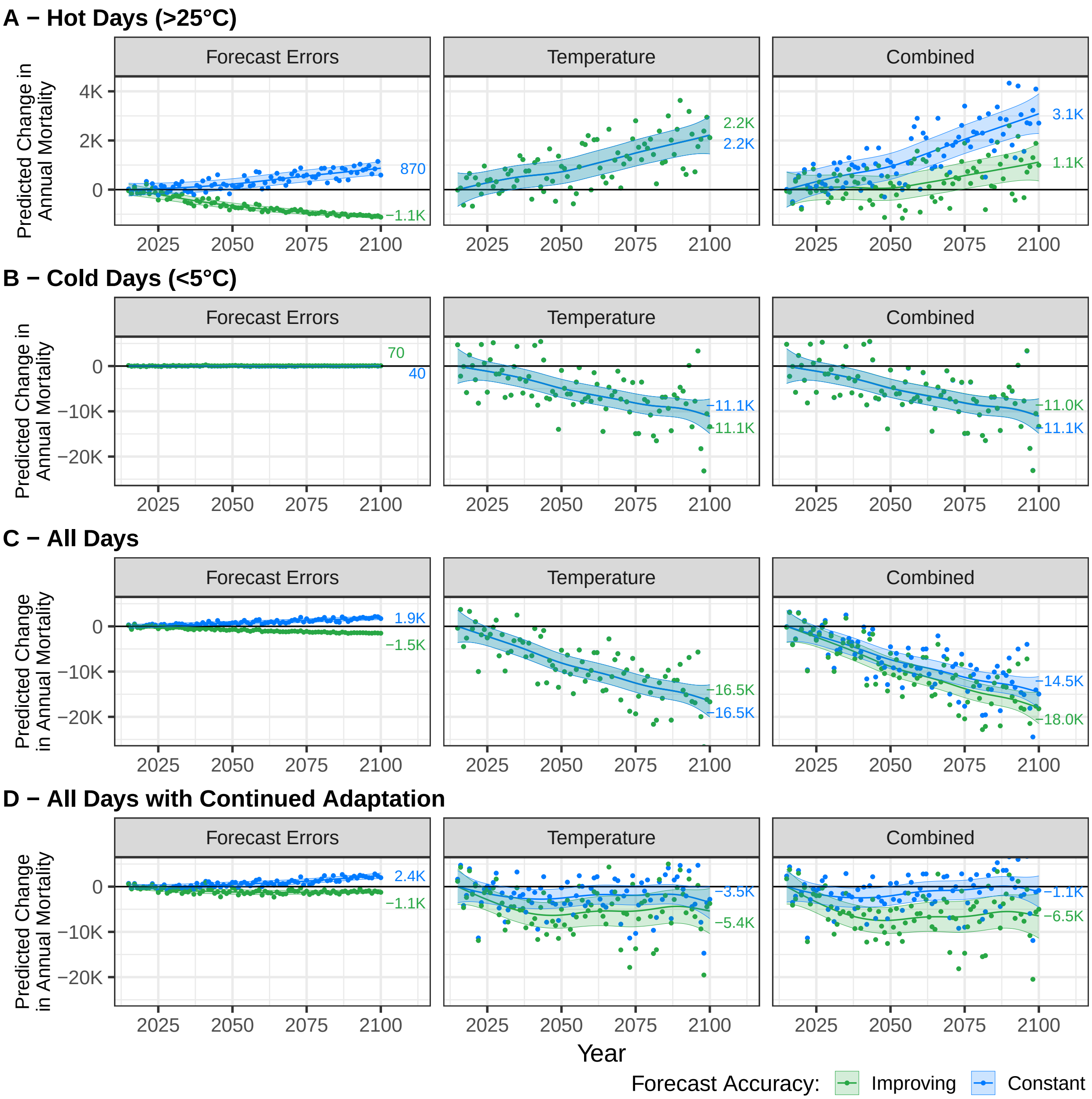

Figure 5: Forecast errors become more important with climate change. The figure shows the relationship between the future evolution of forecast accuracy and mortality from climate change in the U.S. It contrasts a scenario of constant forecast accuracy (blue) with a scenario of continued improvements in forecast accuracy (green). In each panel, column one displays changes in mortality solely due to forecast errors, column two shows changes in predicted mortality solely due to changes in temperatures, and column three displays the combination of the two previous effects. Panel A shows effects on hot days with temperatures above 25°C, panel B shows results for cold days below 5°C, and panel C displays the aggregate effect across all days of the year. Panels A-C hold the mortality-temperature-forecast relationship constant at current counties' climates and current forecast accuracy. Panel D is like panel C, except permitting county characteristics to evolve due to climate change (see text for details).

ESM4 model under SSP2-4.5 aggregated to the county level (see section 6.3 for details). Figure S7 in the appendix shows that the U.S. grows hotter over time, but its daily temperatures do not systematically become more or less variable.

We consider two scenarios for forecast accuracy (see Figure 4). First, we assume that forecasts remain as accurate as they are now. In this case, average forecast accuracy does improve slightly because the U.S. becomes hotter and forecasts are currently more accurate in heat. Second, we assume that the rate of improvement from 2006 to 2016 continues. In this case, forecasts are dramatically better by the end of the century. (In work in progress, we survey forecasters to obtain their anticipated improvements.)

Figure 5 projects the evolution of U.S. temperature-related mortality over the century. It splits mortality into two categories: mortality that is solely due to temperature forecast errors (column 1), and mortality that is solely due to temperatures deviating from the minimum mortality temperature (column 2). Their sum is in column 3.

Panel A shows changes in mortality on hot days. Temperature-driven mortality increases on these days as they become more frequent and more extreme. In addition, mortality due to forecast errors increases if forecast quality remains constant, because people rely on forecasts on hot days. However, if forecasts improve over time at their historical rate, we project that deaths from forecast errors will decline. Combining the forecast error and temperature effects, we project that around 2,000 annual temperature-related deaths can be avoided on hot days by the end of the century if forecasts continue to improve.

In contrast, Panel B of Figure 5 projects very different effects on cold days. First, temperature-driven mortality declines on cold days as they become less frequent and less extreme. Second, forecast errors do not matter much on cold days, because people do not rely on forecasts in the cold. As a result, improvements in forecast quality do not substantially affect mortality from forecast errors in the cold.

Combining the effects of hot days, cold days, and all other days (panel C), we project a net reduction in temperature-driven mortality, which is driven by changes in cold and cool days. However, we also project that mortality from forecast errors will increase over time if forecast quality is constant but fall over time if forecast quality improves at its historical rate. These effects are driven by warm and hot days. Combining each type of mortality, we project that climate change would reduce annual mortality by 14,500 lives if forecasts remain at their current quality and by 18,000 lives if forecasts improve at their historical rate, as the lives saved from reduced cold and cool days dominate other effects.

Panels A-C of Figure 5 allow the mortality-temperature-forecast relationship to vary around the country based on current counties' climates and current forecast accuracy but do not account for how these county characteristics will evolve with climate change. In essence, they permit populations to have adapted to current conditions but do not permit further adaptation as the climate changes. Panel D of Figure 5 permits new adaptation over the century. Here the effects of temperature and forecast errors change with counties' climates and with the accuracy of counties' forecasts. Mortality from forecast errors increases to 2,400 lives per year in the constant accuracy case. This increase is driven by people relying on forecasts more as counties become hotter, both directly and because hotter counties tend to have more accurate forecasts. However, mortality from forecast errors still declines over the century in the improving-accuracy case.

Mortality from temperature again falls over the century, but not by as much as before. As countries become hotter, the remaining cold and cool days become more harmful, which limits the benefits of warming. Mortality from well-forecasted temperatures is now sensitive to the forecast quality scenario, because mortality from well-forecasted temperatures tends to be lower where forecasts are more accurate (note: this may be a statistical artifact that will not persist with our next round of work).

Combining deaths form temperature and forecast errors, our central estimate projects 1,100 fewer deaths per year by the end of the century in the constant-accuracy scenario. However, our confidence intervals cannot rule out that climate change will have no effect on mortality. In contrast, in the improving-accuracy scenario, we project 5,500 fewer deaths per year by the end of the century and this effect is statistically significant.

5 Discussion

By increasing the frequency of hot days on which people use forecasts to mitigate temperature hazards, climate change will make accurate forecasts more beneficial and inaccurate forecasts more harmful. We project that deaths in the U.S. from inaccurate forecasts will increase unless forecast quality improves sufficiently fast. Past improvements in forecast quality have come via substantial public funding for earth observations, data sharing networks, modeling advances, and computational resources [**?**]. Further funding could be critical to limiting mortality from forecast errors.

The quality of forecasts is currently unevenly distributed around the world. In particular, regions that tend to be poorer and hotter tend to have worse forecasts, partly due to a lack of investment in weather observation infrastructure [13]. These same regions are also projected to have an especially strong increase in mortality from climate change [5]. Our results suggest that accurate forecasts may be especially useful—and forecast errors especially costly—in these regions. Investing in weather observation infrastructure and forecasting capacity may be an effective form of climate adaptation funding, and one that would have spillover benefits for the rest of the world by improving forecasts elsewhere.

Our projections allow people to depend on forecasts more as forecasts become better, but they do not account for how new technologies may increase dependence on forecasts. In recent years, widespread diffusion of smartphones has expanded access to quality forecasts. By increasing use of forecasts, these types of technologies will further increase the benefits of accurate forecasts and also amplify the hazards from inaccurate forecasts. By ignoring this factor, our projections may understate the sensitivity of mortality to forecast errors over the coming decades.

On the other hand, our analysis does not account for how having access to the long-run forecast that is scientific modeling of climate change may stimulate public and private actors to invest in resilience. Our heterogeneity analysis shows that people are less exposed to extreme temperatures and to inaccurate forecasts where temperatures are highly variable from day to day. We have suggested that these differences could result from investments in weather resilience. Our projections may overstate the sensitivity of mortality to forecast errors over the coming decades by ignoring investments based on expectations of climate change.

Recent policies have pushed investments in forecasts beyond the one-day-ahead horizon we study here (note: may soon be three-day), including at the seasonal scale [14, 15]. Actors throughout the economy do pay attention to forecasts at the seasonal horizon [8, 19]. Future work should study how mortality depends on these longer-run forecasts and on how these longer-run forecasts affect sensitivity to the short-run forecasts we study here. It may be that forecast at different horizons substitute for each other because better short-run forecasts enable actors to correct course after mistaken long-run forecasts [20], but it may also be that actors make long-run investments based on long-run forecasts only when they know that they can make such course corrections. As seasonal outlooks and multi-week forecasts mature and diffuse, future work should explore their benefits for mortality in order to guide policy investments.

6 Methods and Materials

6.1 Setting and Data

6.1.1 U.S. Weather and Weather Forecasts

Our two primary explanatory variables are daily average temperature and forecasted temperature. Temperature forecasts come from the U.S. National Weather Service (NWS) National Digital Forecast Database (NDFD). This database contains the forecasts that users see on the NWS website (weather.gov) and which are the result of integrating numerical weather predictions, other computational processing, and expert judgment from NWS meteorologists. We focus on the day-ahead forecast and on daily minimum and maximum temperature point forecasts, from which we calculate daily average temperature by taking the simple average of the two measures. We use the noon UTC forecast run. 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 database. We interpolate these observations using the nearest available forecast. The NDFD stores the forecasts on a consistent spatial grid with resolution of 2.5km or 5km, depending on the time period. We aggregate the forecasts to the county level by taking the population-weighted average, based on the 2010 population grids from [21].

For weather realizations, we use PRISM (Parameter-elevation Regressions on Independent Slopes) Climate Group data [22]. PRISM combines weather station observations with a regression-based interpolation procedure that accounts for 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. We aggregate the gridded measures to the county level using the same procedure as the forecasts. We calculate a day's average realized temperature and average forecasted temperature by averaging the day's minimum and maximum temperature.

6.1.2 Mortality

The primary outcome we study is mortality. Mortality data come from the CDC's WONDER Online Database (https://wonder.cdc.gov/). It contains records of all vital events that occurred in the U.S. from 2005 to 2017, recorded at the county-month level. Data for counties with a small number of vital events are suppressed by the CDC due to privacy concerns. See Figure S3 for a map of suppressed counties. For counties with missing mortality rates we impute values using the monthly state-wide mortality rate. Estimation results do not change significantly when dropping suppressed values from the estimation. We have also estimated the relationship between mortality, forecasts, and temperature at the monthly level using complete counts of vital events from the CDC's restricted-access database. The results are comparable (See Section B for details).

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.

6.1.3 Climate Projections

Projections of future temperature come from the Climate Model Intercomparison Project, version 6 (CMIP6) Global Fluid Dynamics Laboratory (GFDL) model. We consider gridded daily temperature projections within the Continental U.S. from 2015 through 2100. We aggregate the data to the county-level using 2010 population grids from [21].

6.2 Econometric Model

The econometric model relates mortality to temperature and forecast errors while controlling for potentially correlated weather and other spatial or temporal confounders.

We flexibly model the relationship between mortality, temperature, and forecast errors by fitting a tensor product spline basis to approximate the function $f(T_{cd}, e_{cd})$, where T_{cd} is realized temperature and e_{cd} is realized forecast error (defined as the forecasted temperature minus the temperature realization) in county c on day d. After the spline transformation of temperatures and forecast errors, we express the model linearly with $f(T_{cd}, e_{cd}) = X'_{cd}\beta$. On the temperature dimension, we use a fourth-order polynomial (matching recent work estimating the relationship between mortality and temperature [5]). We subtract county-specific means from forecast errors to adjust for possible measurement error or misalignment between the PRISM realized temperature and the NDFD forecast data.[1] The demeaned forecast errors enter according to a natural cubic spline with internal knots at -1 and +1°C error and boundary knots at -2.5 and +2.5°C error (so that the relationship is assumed to be linear beyond those values). These values roughly correspond to 1 and 2 standard deviations of forecast error. To allow for a different relationship between mortality and forecast errors at different temperatures, we incorporate a quadratic interaction term between the natural cubic spline of forecast errors and temperatures. Overall, the procedure results in a vector X_{cd} with entries that are each functions of T_{cd} and e_{cd}.

[1] Results are robust to using alternative demeaning techniques such as county × month or county × month × temperature bin specific means and to using raw forecast errors. Subtracting county-specific means is our preferred specifications as it results in the highest precision.

We aggregate this vector to the monthly level by summing over the days within year-month t: $X_{ct} = \sum_{d \in t} X_{cd}$. We estimate a common effect across the days within a month. Our specification accounts for harvesting across days within a month.

The estimating equation is

$$m_{ct} = X'_{ct}\beta_c + g(P_{ct}) + \lambda_{cm} + \alpha_{cm} \times t + \rho_t + \varepsilon_{ct},$$

where m_{ct} is county c's mortality rate in year-month t and

$$\beta_c = \bar{\beta} + \gamma_{AvgT} \cdot \mathrm{AvgT}_c + \gamma_{VarT} \cdot \mathrm{VarT}_c + \gamma_{FcstAcc} \cdot \mathrm{FcstAcc}_c.$$

We isolate excess mortality by controlling for location and time fixed effects. The λ_{cm} are county-month fixed effects. For each month m, these absorb all location-specific factors that are fixed over the sample period, including geography, infrastructure, and governance. These also absorb all location-specific and persistent seasonal effects. The county-month trend α_{cm} absorbs consistent evolution in a county's mortality rate, within each season. Year-month fixed effects ρ_t partial out all confounders common across counties within a year-month, including national economic or health patterns. We also control for measures of above or below median rainfall ($g(P_{ct})$). Altogether, we estimate mortality in excess of what would be expected based on county, season, time period, and rainfall.

The ε_{ct} is an error term. We cluster all standard errors at the Weather Forecast Office (WFO) level in order to account for unobserved correlation within a WFO over time and within a WFO's counties in a given year-month.

The β_c capture cross-county heterogeneity in the mortality response to temperature and forecast errors. The simpliifed model without heterogeneity contyains only $\bar{\beta}$. The full model also contains coefficients on $AvgT_c$ for annual average temperature, $VarT_c$ for temperature variability, and $FcstAcc_c$ for the average temperature forecast accuracy in county c. We calculate annual average temperatures as the simple mean across all daily temperatures in a county between 2006 and 2016. Temperature variability is measured through the root mean squared error (RMSE) of persistence temperature forecasts, i.e. how much do temperatures on day t deviate on average from temperatures on the previous day $t-1$: $\mathrm{VarT}_c = \sqrt{(1/N)\sum_t (T_{ct} - T_{ct-1})^2}$.[2] Finally, we calculate the average forecast accuracy in

[2] We have chosen this measure as it provides a good comparison for the accuracy of day-ahead temperature forecast errors. In an extension, we have also experimented with the standard deviation of temperature around the multi-year average temperature on the day of year as an alternative measure for temperature variability. This measure is highly correlated with the RMSE of persistence forecasts and results are robust to the alternative specification.

county c as the RMSE of day-ahead forecast errors. A large RMSE of forecast errors implies that temperature forecast are inaccurate in county c while small RMSEs imply accurate forecasts.

We note that average temperatures, temperature variability, and forecast accuracy are all correlated. This is visualized in Appendix Figure S2 and also apparent in the maps in Figure S5. For example, temperature variability in the US is generally lower in warmer counties which also leads to forecasts being more accurate in warm areas. The heterogeneity analysis identifies the orthogonal impact of of temperature, forecast errors, and mortality

The average relationship between daily temperatures, forecast errors, and mortality is identified by the exogeneity of weather and forecast errors once we net out location and time fixed effects and rainfall. However, the heterogeneity in that relationship is not causally identified and instead only provides correlational evidence. Confounders such as different age distributions across counties might be responsible for some of the heterogeneity that we attribute to differences in average temperatures, temperature variability, and forecast accuracy.

6.3 Extrapolating Forecast Errors and Projecting Mortality

To project future mortality as a function of temperature and forecast errors, we need to simulate future one-day-ahead forecast errors given climate projections on daily average temperatures. Analyzing past forecast errors on the period from 2005-2017 reveals that forecast error distributions are not Normal conditional on temperature, have an average skewness slightly larger than 0, and exhibit substantial excess kurtosis at high temperatures. Forecast errors also have non-zero means and differential standard deviations depending on temperature, location, and month. We further observe a strong decreasing annual trend in the standard deviation of forecast errors across all ranges of the temperature distribution. When resampling future forecast errors, we aim to preserve all of these moments but allow the standard deviation of forecast errors to either stay constant beyond the year 2020 (constant forecast accuracy scenario) or continue to decrease at the observed historical rate (improving forecast accuracy scenario).

To achieve this, we implement the following procedure, which we plan to supplement with evidence from an expert elicitation of professional weather forecasters in the future.

1. Using ordinary least squares and population weights, we estimate the mean of forecast errors conditional on temperature, county, and month in the observed 2005-2017 sample:
$$\mu_{cd} := E[e_{cd}|T_{cd}, c, t] = h_{\mu}(T_{cd}) + \lambda_{cm}.$$

Here, h_μ denotes a natural cubic spline of temperature with boundary knots at -15°C and 35°C, and internal knots evenly spaced in 5°C steps. λ_{cm} denotes county × months fixed effects.

2. Given estimated means and using a Gamma generalized linear model with a logistic link function and population weights, we compute the variance of forecast errors conditional on temperature, weather forecasting office, month, and year in the observed 2005-2017 sample:

$$\sigma_{cd}^2 := E[(e_{cd} - \mu_{cd})^2 | T_{cd}, c, t, \text{year}] = \exp\left(h_\sigma(T_{cd}) + \lambda_{\text{WFO},m} + \delta \cdot \text{year}\right).$$

   Again, h_σ denotes a natural cubic spline of temperature with the same knots as h_μ, $\lambda_{\text{WFO},m}$ denotes weather forecasting office × month fixed effects, and δ denotes the estimated annual trend in forecast error variance. We use a Gamma generalized linear model with logistic link to prevent future predicted forecast variances from becoming negative. We utilize WFO×month fixed effects as opposed to county×month fixed effects to avoid overfitting, as weather forecasting offices generally cover multiple counties.

3. We compute standardized forecast errors $\tilde{e}_{cd}$ from observed forecast errors in the 2005-2017 sample using the respective estimated mean and variance:

$$\tilde{e}_{cd} := \frac{e_{cd} - \hat{\mu}_{cd}}{\hat{\sigma}_{cd}}.$$

   The standardized errors have mean zero and unit variance, though their higher order moments can still depend on temperature, location, or month.

4. To match these higher order moments conditional on temperature, we sample a hypothetical normalized forecast error $\tilde{e}_{cd}^{(b)}$ in the following way: We determine in which 5° temperature bin the projected future temperature under climate change falls. We then sample from all historical normalized errors $\tilde{e}_{cd}$ that were observed on a day with a temperature in the same bin.

5. Given resampled normalized errors, we can compute an imputed forecast errors in county c on day d as

$$e_{cd}^{(b)} = \hat{\mu}_{cd} + \hat{\sigma}_{cd} \cdot \tilde{e}_{cd}^{(b)}.$$

   In the improving accuracy scenario, we use continued estimated annual trends $\hat{\delta} \cdot \text{year}$. In the constant accuracy scenario, we set the year variable to 2020 for all years beyond 2020. The resulting average trends across all counties are displayed in Figure 4.

The procedure ensures that we match means and standard deviations of forecast errors conditional on temperature, location, and month. When assessing the resampling procedure on the observed past weather data from 2005-2017, it also performs fairly well in matching higher order moments of forecast errors conditional on temperature as shown in Figure S4.

Given future temperatures T_{cd} and the respective imputed forecast errors $e_{cd}^{(b)}$ for each county and day from 2015 to 2100, we predict monthly mortality rates $\hat{m}_{ct}^{(b)}$ using the model outlined in section 6.2. We compute uncertainty in estimates by carrying through both estimation uncertainty and resampling uncertainty.

References

[1] Anderson, B. G. & Bell, M. L. Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology* **20**, 205–213 (2009). URL `https://www.jstor.org/stable/20485691`.

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

[3] Gasparrini, A. *et al.* Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet* **386**, 369–375 (2015).

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

[5] Carleton, T. *et al.* Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics* **137**, 2037–2105 (2022).

[6] Bakkensen, L. Public versus private risk information: Evidence from U.S. tornadoes. Tech. Rep., University of Arizona (2016).

[7] Schlenker, W. & Taylor, C. A. Market expectations of a warming climate. *Journal of Financial Economics* **142**, 627–640 (2021). URL `https://www.sciencedirect.com/science/article/pii/S0304405X21001495`.

[8] Shrader, J. Improving climate damage estimates by accounting for adaptation. *Working Paper* (2023). URL `http://jeffreyshrader.com/papers/forecasts_and_adaptation.pdf`.

[9] Anand, V. The value of forecast improvements: Evidence from advisory lead times and vehicle crashes. *Available at SSRN 4206910* (2022).

[10] Song, Y. The value of weather forecasts: Labor responses to accurate and inaccurate temperature forecasts in China. Job market paper (2022).

[11] Shrader, J. G., Bakkensen, L. & Lemoine, D. Fatal errors: The mortality value of accurate weather forecasts. *National Bureau of Economic Research* **w31361**, 41 (2023).

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

[13] Linsenmeier, M. & Shrader, J. G. Global inequalities in weather forecasts (2023). URL https://osf.io/7e2jf.

[14] Rep. Lucas, F. D. Weather Research and Forecasting Innovation Act of 2017 (2017). URL https://www.congress.gov/bill/115th-congress/house-bill/353.

[15] ECMWF. ECMWF Strategy 2021–2030. Tech. Rep. (2020). URL https://www.ecmwf.int/en/about/what-we-do/strategy. Doi: 10.21957/s21ec694kd.

[16] Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **0**, eadi2336 (2023). URL https://www.science.org/doi/10.1126/science.adi2336. Publisher: American Association for the Advancement of Science.

[17] Lopez-Gomez, I., McGovern, A., Agrawal, S. & Hickey, J. Global Extreme Heat Forecasting Using Neural Weather Models. *Artificial Intelligence for the Earth Systems* **2** (2023). URL https://journals.ametsoc.org/view/journals/aies/2/1/AIES-D-22-0035.1.xml. Publisher: American Meteorological Society Section: Artificial Intelligence for the Earth Systems.

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

[19] Lemoine, D. & Kapnick, S. Financial markets value skillful forecasts of seasonal climate. *Nature Communications* **15**, 4059 (2024). URL https://doi.org/10.1038/s41467-024-48420-z.

[20] Millner, A. & Heyen, D. Prediction: The long and the short of it. *American Economic Journal: Microeconomics* **13**, 374–398 (2021). URL https://www.aeaweb.org/articles?id=10.1257/mic.20180240&&from=f.

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

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

Supplementary Material

A Additional Figures and Tables

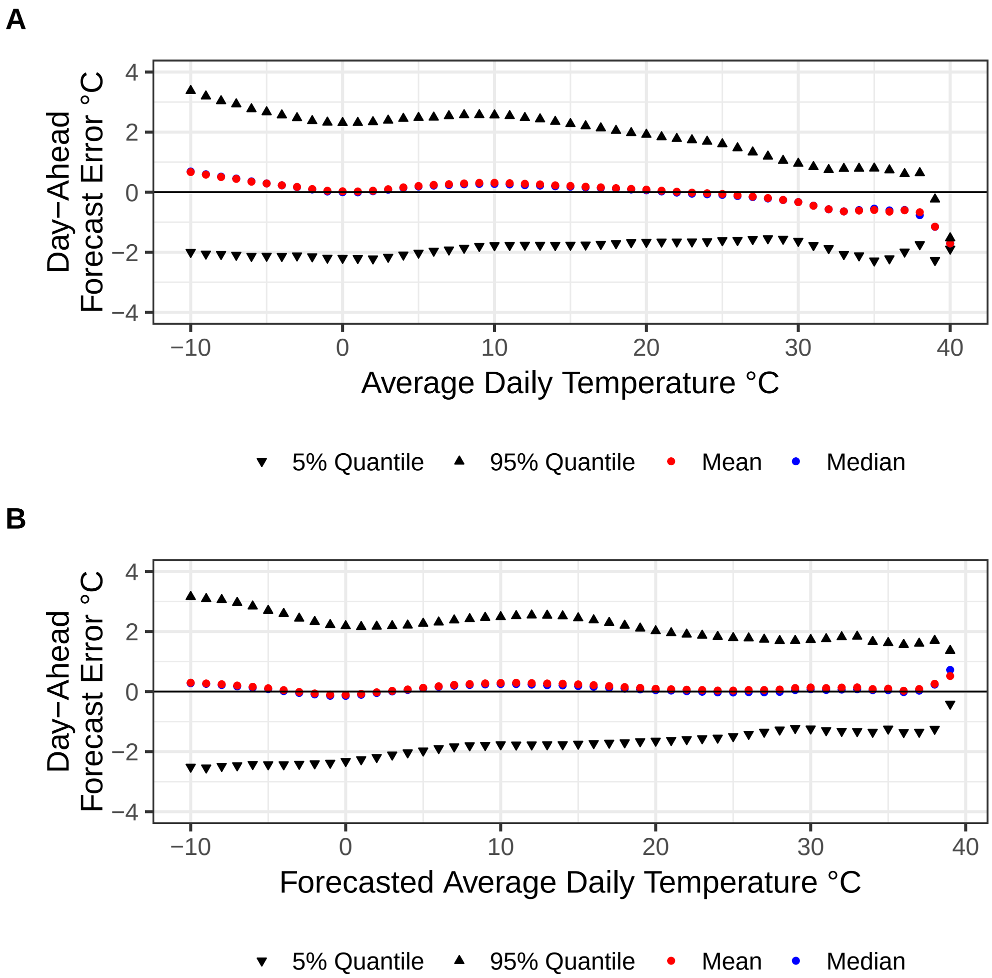

Figure S1: Forecast errors are relatively unbiased conditional on forecasted temperatures but biased conditional on realized hot and cold temperatures. The figure shows the distribution of day-ahead temperature forecast errors conditional on realized temperature (panel A) and forecasted temperatures (panel B). While forecast errors are unbiased conditional on forecasts (ex-ante), they are biased conditional on very hot and very cold temperatures (ex-post).

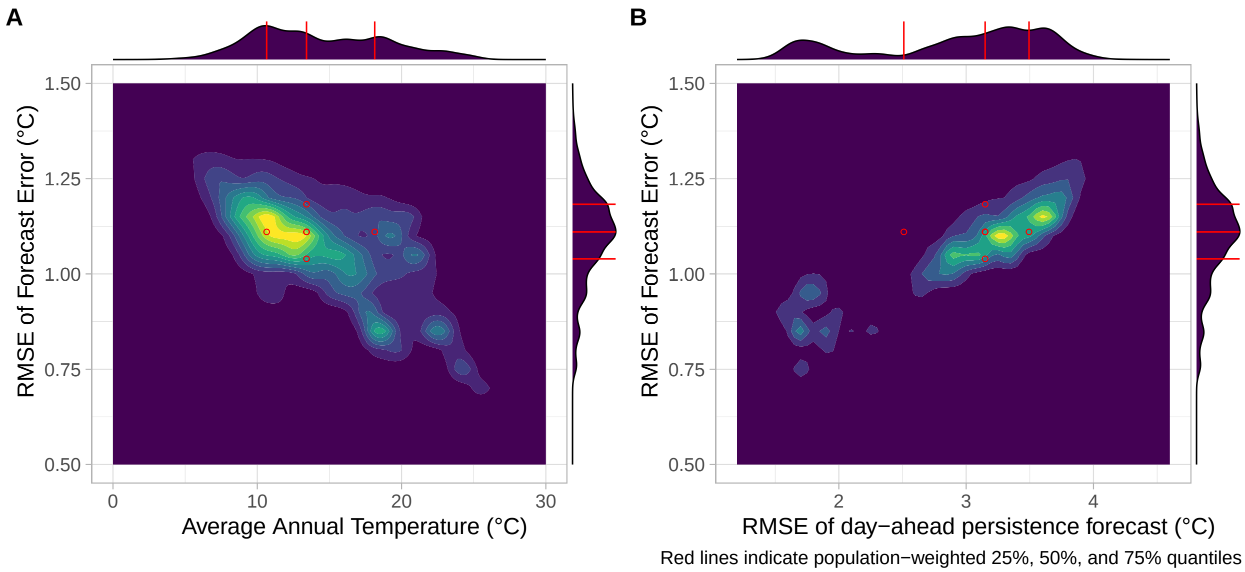

Figure S2: Temperature forecast are more accurate in counties with higher average and less volatile temperatures. The figure shows the joint distribution of forecast accuracy and average annual temperature (panel A) and the joint distribution of forecast accuracy and temperature variability (panel B) on the county-level. Red points and lines indicate the 25%, 50%, and 75% quantiles at which heterogeneity is evaluated in figure 3.

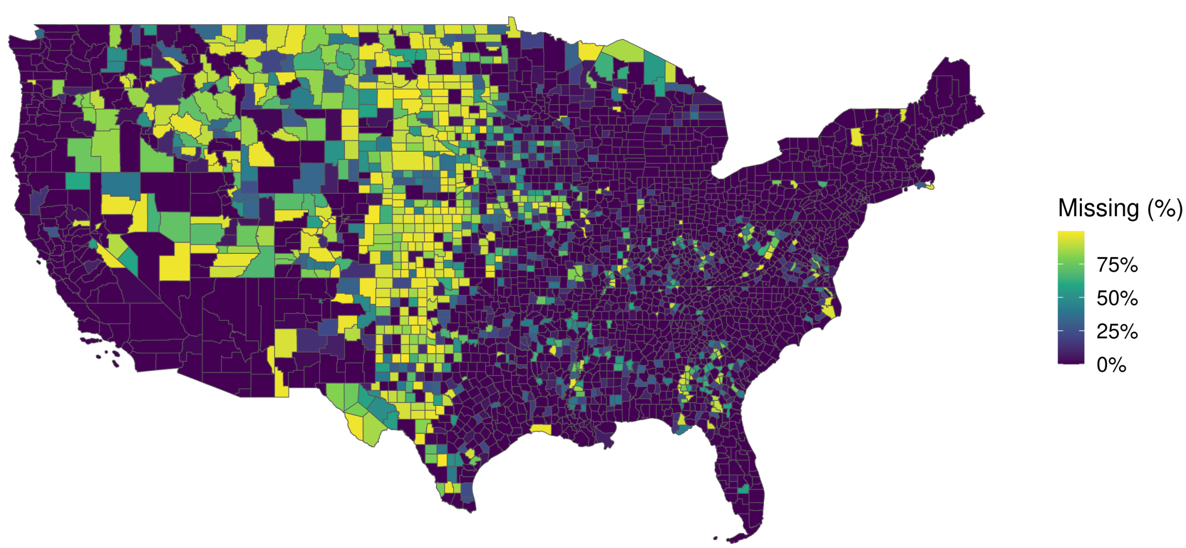

Figure S3: Percentage of suppressed values in CDC Wonder data The map shows the percentage of suppressed and missing monthly deaths in the CDC wonder database for the period between 2005 and 2017.

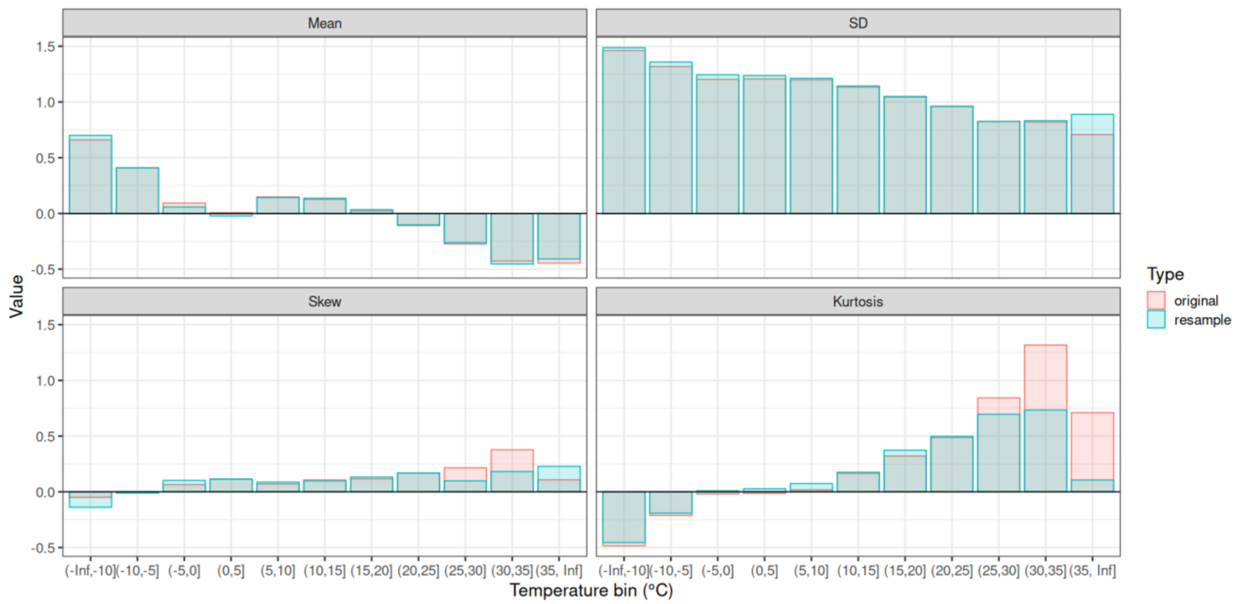

Figure S4: Resampled forecast errors match observed forecast error moments conditional on temperature. Based on the 2005-2017 daily temperature and forecast error data, the figure shows observed population weighted moments of forecast errors (red) as well as moments of resampled hypothetical forecast errors (blue) conditional on realized temperature.

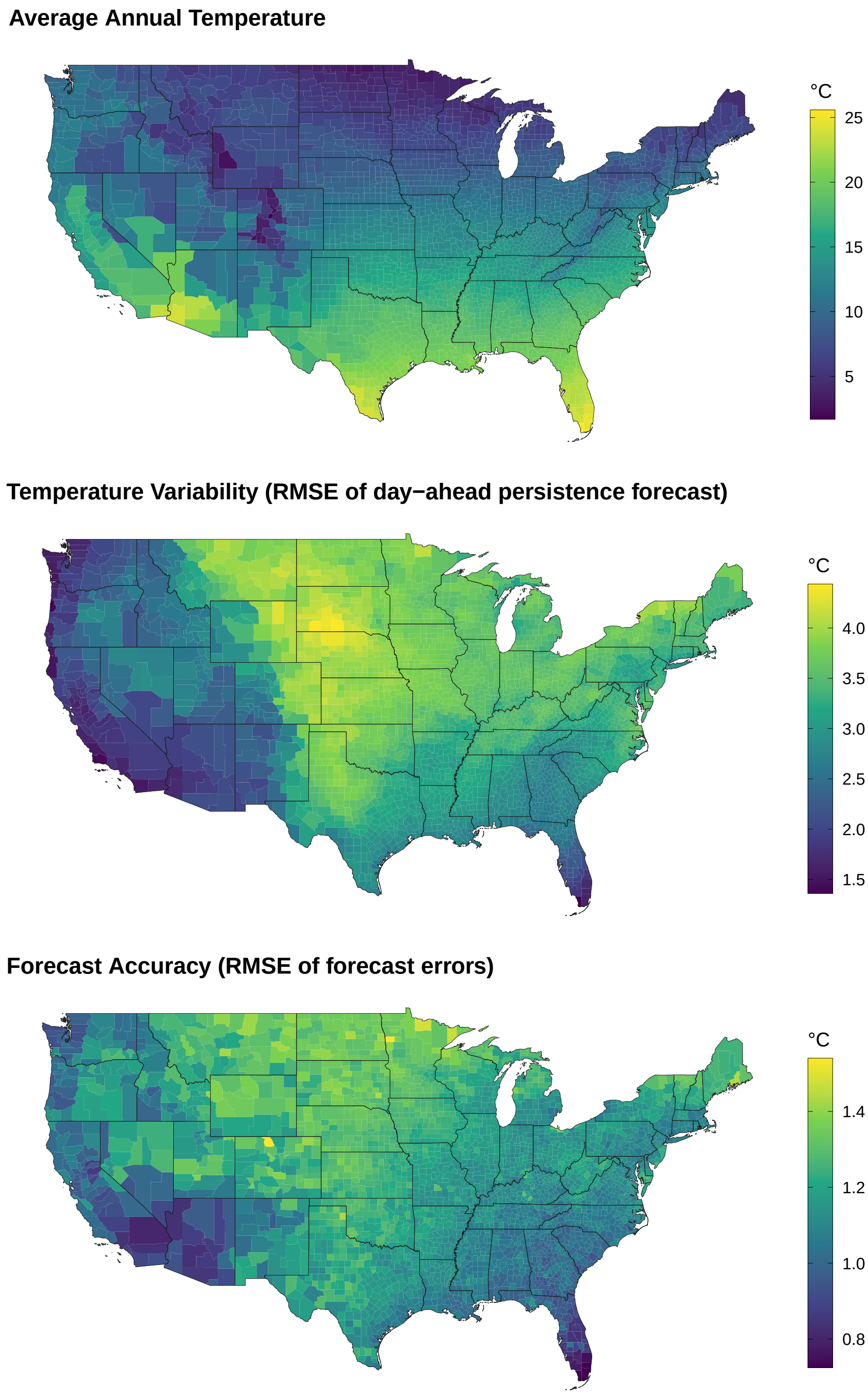

Figure S5: Mapping county-level covariates. The maps display the distribution of average annual temperatures, temperature variability, and forecast accuracy across US counties. The three covariates are used in the heterogeneity analysis.

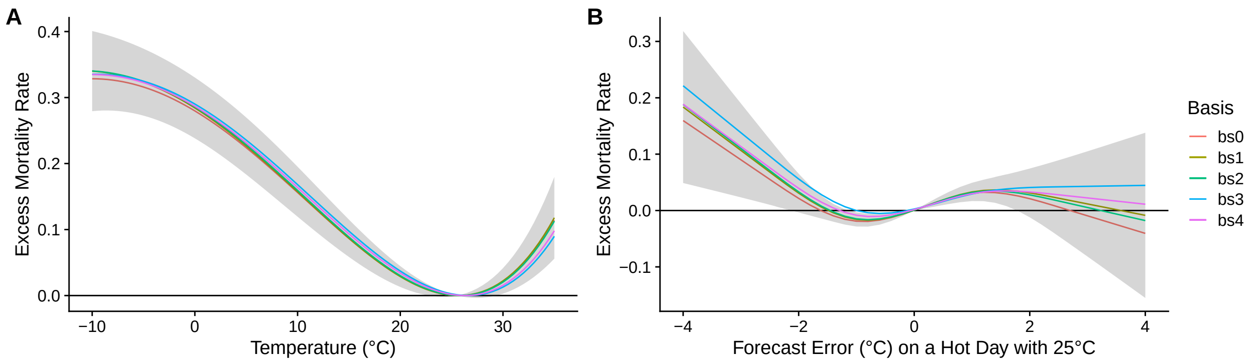

Figure S6: Robustness of estimated mortality response to different bias corrections. The figure displays the robustness of the estimated mortality response when using different bias correction techniques for forecast errors. Whether we let forecast errors enter the response function directly (bs0) or subtract county (bs1), county × month (bs2), county × month × temperature bin (bs3), or county × temperature bin (bs4) specific error means before forecast errors enter the response function, does not impact results substantially. Our preferred specification corrects for county-specific means (bs1) and its confidence intervals are displayed in grey.

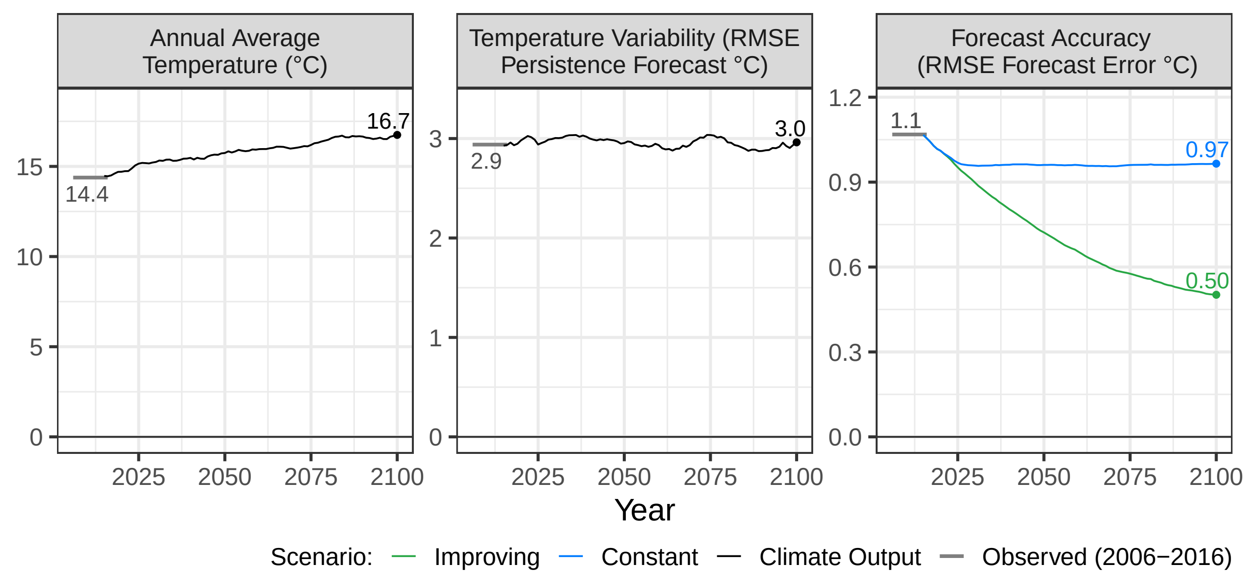

Figure S7: Projections of average US county characteristics under climate change. The figure displays projections for average annual temperatures and day-to-day temperature variability under SSP2-4.5 using GFDL-ESM4. It also shows projected forecast accuracy as measured through the RMSE of forecast errors under two potential scenarios: constant accuracy (blue) and improving accuracy (green).

B Corroborating the WONDER results

Table S1: Effect of Temperature on Mortality With and Without Accounting for Forecasts

	(1) log death rate	(2) log death rate
Temperature	-0.00302***	-0.00840**
	(0.000799)	(0.00403)
Temperature^2	-0.0000273	0.00299***
	(0.0000353)	(0.00105)
Forecast		0.00544
		(0.00372)
Forecast^2		0.00290***
		(0.000939)
Temperature × Forecast		-0.00591***
		(0.00198)
Rain 25th percentile	0.00220*	0.00214*
	(0.00116)	(0.00117)
Rain 75th percentile	-0.00291**	-0.00289**
	(0.00111)	(0.00108)
N	7,203	7,203
Clusters	130	130

Notes: The table shows results of a single estimation of Equation (**??**) on the restricted-access mortality sample. The dependent variable is the natural log of the average over the month of the daily mortality rate per 100,000 people. Forecasts are the 1-day-ahead temperature forecast. The model includes year-month fixed effects and county-by-month-of-year fixed effects interacted with a linear time trend. Weighted by annual county population. Standard errors, clustered at the county warning area-level, are below each estimate. Significance: * $p < .10$, ** $p < .05$, *** $p < .01$.