NBER WORKING PAPER SERIES

INEXPENSIVE HEATING REDUCES WINTER MORTALITY

Janjala Chirakijja
Seema Jayachandran
Pinchuan Ong

Working Paper 25681
http://www.nber.org/papers/w25681

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2019

We thank Sachet Bangia, Jamie Daubenspeck, Alejandro Favela, and Caitlin Rowe for outstanding research assistance and several seminar participants for helpful comments. Research reported in this paper was supported by the National Institute on Aging of the National Institutes of Health under award number R03AG058113. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Bureau of Economic Research.

Inexpensive Heating Reduces Winter Mortality
Janjala Chirakijja, Seema Jayachandran, and Pinchuan Ong
NBER Working Paper No. 25681
March 2019
JEL No. I1,J14,Q41

ABSTRACT

This paper examines how the price of home heating affects mortality in the US. Exposure to cold is one reason that mortality peaks in winter, and a higher heating price increases exposure to cold by reducing heating use. It also raises energy bills, which could affect health by decreasing other health-promoting spending. Our empirical approach combines spatial variation in the energy source used for home heating and temporal variation in the national prices of natural gas versus electricity. We find that a lower heating price reduces winter mortality, driven mostly by cardiovascular and respiratory causes.

Janjala Chirakijja
Department of Econometrics and
Business Statistics
Monash University
900 Dandenong Road
Caulfield East
Victoria 3143
Australia
janjala.chirakijja@monash.edu

Seema Jayachandran
Department of Economics
Northwestern University
2211 Campus Dr
Evanston, IL 60208
and NBER
seema@northwestern.edu

Pinchuan Ong
Department of Economics
Northwestern University
2211 Campus Dr
Evanston, IL 60208
PinchuanOng2014@u.northwestern.edu

1 Introduction

In the US, 17% of households spend more than 10% of their income on home energy. Heating is the largest component of annual home energy consumption, despite being used for only part of the year (RECS 2009).

High heating costs impose a difficult trade-off on households: They have to keep their home uncomfortably cold to save on heating or forgo other spending to afford their high heating bill. How acute this dilemma is depends on how expensive home heating is. Through both a substitution and an income channel, a higher price of heating could be harmful to health. First, using less heating means exposure to lower ambient temperature, which has been linked to cardiovascular, respiratory, and other health problems.[1] Second, if families do not cut back usage one-for-one when the price rises, their energy bills will increase. This can lead to cutbacks in other expenditures that affect health, such as food and health care.

This paper estimates the causal effect of heating prices on mortality in the US. A large literature has documented that mortality peaks in winter and that cold weather is associated with higher mortality. Our contribution is to examine whether high home heating costs exacerbate this pattern of "excess winter mortality."

Our empirical design uses spatial variation across the US in the energy source used for home heating. Natural gas and electricity are used for heating by 58% and 30% of households, respectively. Importantly, there is considerable geographic variation across counties in whether an area relies on natural gas versus electricity. We combine this spatial variation with temporal variation in the national prices of natural gas and electricity. The ratio of the natural gas to electricity price varied substantially over the 2000 to 2010 study period, most notably due to the boom in shale production of natural gas. We leverage the fact that households in areas that rely more on natural gas for heating experienced a decline in their home heating price as a result of the shale gas boom, relative to households in areas reliant on electricity.

We find that lower heating prices reduce mortality in winter months.[2] The estimated

[1]The main hypothesized mechanisms are changes in blood pressure and blood chemistry which increase the risk of strokes, myocardial infarctions, and pulmonary embolisms, and higher infection risk due to a suppressed immune system (Crawford et al. 2003; Liddell and Morris 2010).

[2]We define "winter" in this paper as November to March, the coldest months of the year. Analyses of excess winter mortality use December to March in the UK and Europe, where those are the coldest months (Wilkinson et al. 2004). We include November because the average temperature is as low as in March in the US (see Appendix Figure A1).We also show the results using December to March.

effect size implies that the drop in natural gas prices in the late 2000s, induced largely by the boom in shale gas production, averted 11,000 winter deaths per year in the US. We also find that the effect does not just represent short-run hastening of mortality. We show that the effect, which is driven mostly by cardiovascular and respiratory causes, is robust to several checks on the specification.

Our paper contributes to the literature on the effects of cold weather on mortality (Eurowinter Group 1997; Analitis et al. 2008; Deschenes and Moretti 2009), morbidity (Ye et al. 2012), and nutrition (Bhattacharya et al. 2003; Cullen et al. 2004; Beatty et al. 2014). To our knowledge, ours is the first study to estimate the causal effect of heating prices — a plausibly important and policy-relevant mediating factor — on mortality or, more generally, on health. Previous work has found that the winter spike in mortality is stronger for people living in older housing, which tends to be poorly insulated, which is suggestive but not dispositive that indoor temperature is a mediating factor (Wilkinson et al. 2007). The studies closest to ours examine how home weatherization affects health; some studies report reductions in morbidity, and others find null results (Critchley et al. 2007; El Ansari and El-Silimy 2008; Green and Gilbertson 2008; Howden-Chapman et al. 2007). Most of these studies analyze small samples and thus lack statistical power to examine mortality or other objective health outcomes. Another related literature documents a positive association between heating subsidies for low-income families and health, usually without isolating a causal relationship (Frank et al. 2006; Grey et al. 2017). An exception is Crossley and Zilio (2018) who use a UK program's age-eligibility rule to study the effects of unconditional cash transfers for the elderly labeled as "winter fuel payments"; the payments reduce one of the two biomarkers for infection examined.

Our paper also contributes to the literature on the consequences of shale gas production, specifically on its health effects due to lower energy prices. Previous work in economics has studied local economic effects of shale gas production on job creation and wages (Feyrer et al. 2017; Jacobsen 2019), fertility (Kearney and Wilson 2018), and crime (DeLeire et al. 2014; Bartik et al. forthcoming). Shale gas also affects health through channels besides energy prices. It often displaces coal in electricity generation, lowering pollution emissions (Cullen and Mansur 2017; Fell and Kaffine 2018; Holladay and LaRiviere 2017; Knittel et al. 2015; Linn and Muehlenbachs 2018). There are also potentially large local health costs due to chemical contamination of the water supply (Jackson et al. 2014; Groundwater

Protection Council 2009; Muehlenbachs et al. 2015). Several recent papers find a link between fracking and poor birth outcomes (Casey et al. 2016; Currie et al. 2017; Hill 2018). The health harm from the toxic chemicals used is likely much larger per person affected than the health benefits from lower energy prices; however, lower energy prices affect a much larger population. Thus, the net health effect of fracking aggregated for the whole US population is ambiguous. Finally, our empirical strategy is similar to that of Myers (forthcoming) who compares households that use heating oil or natural gas in Massachusetts to study whether home energy costs are capitalized into home values.

2 Empirical strategy

We estimate the effect of heating prices on mortality. As a proxy for the heating price that an individual faces, we combine information on whether her locality uses natural gas for heating and the national prices of natural gas and electricity. This approach enables us to control for average differences across localities and time.

2.1 Estimating equations

We estimate the following equation via ordinary least squares regression to quantify the impacts of the price of heating on mortality:

$$log(m_{jt}) = \alpha + \beta ShareGas_j \times log(RelPrice_t) + \gamma_j + \tau_t + \theta Z_j \times log(RelPrice_t) + \delta X_{jt} + \epsilon_{jt} \quad (1)$$

Each observation is a county-month. The outcome $log(m_{jt})$ is the log of age-adjusted mortality in county j in month t. The key regressor is the interaction of $ShareGas_j$ — the proportion of households in the area that used natural gas for heating in the base year of 2000 — and $log(RelPrice_t)$. $RelPrice$ is the ratio of the national price of gas to electricity. When natural gas prices are higher (high $RelPrice$), areas with high $ShareGas$ face relatively higher heating prices. The hypothesis is that $\beta > 0$: A higher heating price increases mortality. County fixed effects, γ, and month-year fixed effects, τ, absorb the main effects of $ShareGas$ and $log(RelPrice)$. Throughout, we cluster standard errors by state to allow for serial correlation, as well as spatial correlation among counties in a state.

We include several control variables in our main specification. Because the study period spans the housing market boom and collapse and the Great Recession, we control for a housing price index, the unemployment rate, and the manufacturing share of local employment income. The vector X also includes air pollutants linked to mortality. We also control for area characteristics Z, specifically pre-period log median income and the share of the population over age 70, interacted with $log(RelPrice)$; these control variables help safeguard against a spurious correlation from the Great Recession (or another phenomenon with a similar temporal pattern as $log(RelPrice)$) having a differential impact on mortality across socioeconomic or demographic groups (Hoynes et al. 2012).

For the difference-in-differences estimation, we restrict the data to only winter months (when possible), when energy use is mostly for heating and most of the year's heating is consumed. We also estimate a triple difference model that uses the non-winter months as an additional comparison group, testing the prediction that the price of heating affects mortality more in winter than in other, warmer months.

Some winters or particular months in winter are colder than others, so we can also use temperature to define the third difference. We calculate for each county-month a measure of coldness, namely heating degree-days (HDD), as described in Section 3. The triple difference model using HDD is as follows:

$$\begin{aligned} log(m)_{jt} \; = \; & \alpha + \lambda_1 ShareGas_j \times log(RelPrice_t) \times HDD_{jt} + \lambda_2 ShareGas_j \times log(RelPrice_t) \\ & + \lambda_3 ShareGas_j \times HDD_{jt} + \lambda_4 log(RelPrice_t) \times HDD_{jt} + \lambda_5 HDD_{jt} \\ & + \theta_1 Z_j \times log(RelPrice_t) \times HDD_{jt} + \theta_2 Z_j \times log(RelPrice_t) + \theta_3 Z_j \times HDD_{jt} \\ & + \theta_4 ShareGas_j \times log(RelPrice_t) \times \overline{HDD}_j + \theta_5 log(RelPrice_t) \times \overline{HDD}_j \\ & + \theta_6 Z_j \times log(RelPrice_t) \times \overline{HDD}_j + \gamma_j + \tau_t + \delta X_{jt} + \epsilon_{jt} \end{aligned} \tag{2}$$

The prediction is $\lambda_1 > 0$. Note that equation (2) controls for the county's average HDD in winter, $\overline{HDD}_j$, in parallel to HDD_{jt} to adjust for systematic differences (e.g., demographics) between colder regions such as the Midwest and warmer ones such as the South. The results are similar if we omit these extra control variables, using average differences across places in the severity of their winters as additional identifying variation.

2.2 Assessing the income and substitution channels

An auxiliary outcome we examine is the average price of heating experienced by consumers. We calculate the weighted average of the local prices of natural gas and electricity, where weights are the local consumption of each energy source. A model analogous to equation (1) but using log average local price as the outcome is like the "first stage" if we were using instrumental variables estimation. We would expect $\beta = 1$ if our regressors were measured without error and if local and national average prices moved entirely in lockstep. The coefficient will be less than 1 if there is either measurement error or price variation specific to a locality, which we would expect due to local demand and regulatory factors plus a supply side that is not fully integrated across the US.

We also examine two other "1.5$^{\text{th}}$" stage outcomes to gauge the importance of the substitution and income channels. First, we examine the (log) quantity of home energy use, combining gas and electricity. When the outcome is log energy use, the coefficient β from equation (1) can be interpreted as a price elasticity. We expect it to be negative: Consumers substitute away from heating when it becomes more expensive. The data on home energy use do not disaggregate it by purpose (e.g., heating, lighting). Thus, while the variation in the price of natural gas is mainly measuring variation in a household's heating price, the outcome combines heating plus other energy uses, so the coefficient represents a lower bound magnitude for the price elasticity of heating demand. Natural gas's home use is mostly for heating (space heating and water heating), with an additional small contribution from kitchen ranges and clothes dryers. Non-heating home energy needs such as lighting, refrigeration, and air conditioning predominantly use electricity throughout the US. Home heating is also the largest home energy use, accounting for 42% of annual home energy consumption, with water heating accounting for an additional 18% (RECS 2009). Other major categories are lighting and appliances (30%), refrigeration (5%), and air conditioning (6%).

Second, we examine how higher heating prices affect expenditures on home energy use, again with the caveat that we cannot distinguish spending on heating from other energy uses (although in winter months, heating accounts for the vast majority of energy use). If households are not cutting back one-for-one when the price rises, then we expect higher energy prices to lead to higher energy bills. Of course, we cannot decompose how much of the mortality effects are due to changes in the quantity of home heating versus changes in expenditures on heating since a price change generates both effects as a bundle.

2.3 Geographic variation in heating source

Natural gas and electricity are the two most common energy sources used for home heating, used respectively by 58% and 30% of households nationwide in 2000. Importantly for our purposes, there is considerable geographic variation in energy source; in some communities, almost every household uses natural gas for heating, and in other communities, almost no one does. Figure 1 shows the share of households using natural gas as their heating source across counties, based on 2000 US Census data.

Whether a locality uses natural gas, electricity, or another heating source is, of course, not random, and various factors explain the differences. Natural gas pipelines do not extend to some parts of the US, such as Maine. Areas that are well-suited for hydroelectric power generation have low electricity costs and thus rely more on electricity. For historical reasons, much of the Northeast uses heating oil, a petroleum product, instead of gas or electricity. Importantly, the geographic differences were determined long before the study period and are highly persistent. (The correlation between a county's share using natural gas in 2000 and 2010 is 0.99). Being predetermined does not rule out that an area's heating source is correlated with other factors affecting mortality, so the analysis controls for other locality characteristics in parallel to heating source.

2.4 Temporal variation in energy prices

Figure 2 plots the national prices of natural gas and electricity over the 2000 to 2010 study period. The data source is the US Energy Information Administration. (In this figure and throughout the paper, monetary amounts are in 2016 USD.) Natural gas is one of the fuel sources used in electricity generation, so the two prices co-move, but far from in lockstep. Electricity prices changed somewhat over the time period, while natural gas prices rose and then fell much more dramatically. As a result, the relative price of natural gas to electricity rose and then fell over the period.

Natural gas prices rose from 2004 to 2005 due in part to supply disruptions from major hurricanes along the Gulf coast (Hurricane Ivan in 2004 and Hurricanes Katrina and Rita in 2005) (Brown and Yücel 2008). In addition, increased efficiency of producing electricity from natural gas boosted demand for natural gas during the early 2000s (Hartley et al. 2008). The major reason for the drop in the price of natural gas in the mid-2000s was the sharp increase in shale production of natural gas, which is also plotted in Figure 2.[3]

2.5 Home heating versus other heating

While we sometimes refer to our results as due to home heating, the analysis cannot isolate home heating from other indoor (e.g., workplace) heating. Some policy implications, such as whether to promote increased energy supply, are similar whether the channel is home heating or other indoor heating. For other policies, such as subsidies for consumer heating bills, it would be valuable to isolate heating costs at home, which our research design does not permit. A related, more minor limitation is that we cannot separate the effect of space heating versus water heating; the energy source is the same in most households (RECS 2014). Both types of heating likely affect health through similar mechanisms.

3 Data

Our analysis focuses on the contiguous US between 2000 and 2010. We exclude Hawaii and Alaska because our data source for temperature excludes them. The rest of this section describes our data sources, with further details in the appendix.

3.1 Mortality

We construct the mortality rate from restricted-use Vital Statistics microdata, specifically records for all deaths in the US, indicating the month and county of residence (and county of death), and cause of death. The data include the decedent's age, sex, race, and education level. We exclude counties with a small population over age 50, specifically those in the bottom tenth percentile of all counties, as they have few (often zero) deaths per month.

Following the literature, we age-adjust the mortality rate using population data from the National Cancer Institute's Surveillance Epidemiology and End Results program. Our main specifications examine the logarithm of the age-adjusted mortality rate, and we also report the results in levels.

We focus on causes of mortality that exhibit a high degree of excess winter mortality (EWM). Overall mortality is higher in winter than the rest of the year, but the pattern is more pronounced for some causes than others. We zero in on these causes because it is most

[3] Natural gas markets are not fully integrated globally. Pipeline capacity was a bottleneck to US exports in the late 2000s. Thus, natural gas prices fell in the US relative to other countries over this period (Hausman and Kellogg 2015).

plausible that they are exacerbated by exposure to cold and also because doing so increases statistical power. We use a data-driven approach to determine these causes. We collapse the data geographically to the entire US and estimate a regression of log age-adjusted mortality on a dummy for winter, separately for each of the National Center for Health Statistics (NCHS) 113 Selected Causes of Death. Causes with a large positive winter coefficient have more excess mortality in winter. We also estimate the model in levels to exclude minor causes that might have spuriously large coefficients. We select the causes whose $Winter$ coefficients are in the top quartile in both levels and logs. This procedure identifies 16 causes. We exclude two causes, "deaths from smoke, fire, and flames" because its increase in winter is not due to a direct physiological effect of cold; and "all other diseases" (the residual category), because it is difficult to verify the mechanism for this "cause." The remaining 14 causes fall within four alphabetic (i.e, broad) categories, and generally match the causes highlighted in the literature as exacerbated by cold (e.g., cardiovascular, respiratory). These high-EWM causes (hereafter, EWM causes) account for 61% of total mortality (and 63% of total mortality in winter). Appendix Table A1 lists the causes and their degree of EWM, and Appendix Figure A2 shows the seasonality for EWM and non-EWM causes.

3.2 Independent variables

To construct $ShareGas$, we use 2000 Decennial Census data. The Census longform asks the energy source for home heating, as does the American Community Survey (ACS), which has been fielded annually since 2005. We use the 2000 Census Summary Files that report aggregate data for each county. When our geographic unit is the Public Use Micro Area (PUMA), we construct each PUMA's value of $ShareGas$ from public use microdata.

$RelPrice$, the ratio of the price of gas to electricity, is constructed using monthly national price data from the US Energy Information Administration (EIA). We use national rather than local prices, as national prices are exogenous to a locality, while local prices are affected by local energy demand.[4] For natural gas, we use the citygate price, which is the price faced by local distribution companies. We use the citygate rather than residential price because the latter incorporates fixed costs like the monthly fee to maintain a gas connection. Hence, the citygate price better captures variation in the marginal price faced by consumers.

[4]We investigated using pre-period local differences in the use of natural gas in electricity generation as an additional source of variation. The two prices co-move more if gas and electricity markets are interconnected in this way. However, this approach does not add statistical power.

For electricity, we use the residential price, as there are no citygate data available.

The correct specification depends on the timing of consumers' response to $RelPrice$. We find that residential energy use responds to $RelPrice$ with a lag of three months. This is similar to the finding in Auffhammer and Rubin (2018) that natural gas consumption responds to residential prices with a two-month lag.[5] Consumers seem to cut back on usage only after seeing their energy bill, which typically arrives a few weeks after the billing period ends. In addition, the health effects of cutbacks in heating use or paying higher bills might not be instantaneous. Hence, we use the average of the three- and four-month lagged price to construct $RelPrice$. We obtain similar results when we use prices lagged one month less or an annual-level price series. We also estimate models that incorporate mortality effects in subsequent, post-winter months; the effect in subsequent months could be negative if deaths are hastened by a short duration ("harvesting") or positive if the mortality effects materialize with a longer delay.

The analysis also uses temperature data. We start with the PRISM dataset of daily average temperature for gridpoints across the contiguous US spaced 4 kilometers apart (PRISM Climate Group 2004). We calculate the temperature for each census block and then use population weighting to construct the average for each county or other geographic unit. Our mortality data are at the month level, so we use the daily temperature data to construct heating degree-days (HDD) for the month. HDD is a commonly used measure of coldness — or need for heating — based on the idea that heating demand is linear in temperature when temperature falls below 65°F. That is, $HDD_{jt} = \sum_{x=1}^{T} \max\{65 - tmean_{jtx}, 0\}$ where $tmean$ is the mean temperature of area j on day x of month t, and T is the number of days in month t.

Controlling for air pollution is potentially important because it is correlated with weather conditions and affects mortality (Ye et al. 2012). We use data from the Air Quality System of the US Environmental Protection Agency, aggregating the daily monitoring-station air quality indices (AQI) to the county-month level. We focus on particulate matter (2.5 micron and 10 micron, separately) and nitrogen dioxide, as these are the pollutants correlated with mortality; our results are similar if we control for all of the available AQIs.

Because a major housing market run-up and collapse occurred during the study period,

[5] By federal law, utilities must price the natural gas component of their service for cost-recovery only. Any shock to the citygate price not predicted in advance by the utilities thus influences the residential price one month later, explaining the additional one-month lag for citygate versus residential prices.

we control for a housing price index, available at the state-quarter level from the Federal Housing Finance Agency. Similarly, the Great Recession had different impacts across counties so we control for the unemployment rate (available at the county-month level from the Bureau of Labor Statistics) and the manufacturing sector share of total employee compensation (available at the state-quarter level from the Bureau of Economic Analysis). We also control for median household income and the population share age 70 and older, calculated from the 2000 Census, interacted with $log(RelPrice)$. We show robustness to varying our set of control variables.

3.3 Other dependent variables

We examine intermediate outcomes to shed light on why heating prices affect mortality. We use local residential natural gas and electricity prices to compute the average (consumption-weighted) price of home energy. We also examine residential energy use, constructed as the sum of natural gas and electricity use. Price and usage data are aggregate state-month-level data from EIA.[6]

To measure household spending on home energy, we combine 2000 Census microdata (IPUMS version) and ACS data for 2005 to 2010. This analysis is conducted at the PUMA rather than county level, as the PUMA is the finest geographic identifier provided. For computational ease, the analysis collapses the data to the PUMA-year level.

4 Results

We first present results on the "first stage" and "1.5$^{\text{th}}$ stage" outcomes of home energy prices, quantity of energy consumed, and energy bills. We then present the mortality results.

4.1 Effect of heating price on energy use and spending

We use $ShareGas \times log(RelPrice)$ as an exogenous source of variation in the home heating price faced by households. We do not have household-level data on energy prices, but we can use aggregate administrative data on residential energy prices to verify that our regressor is a good proxy for household prices.

The price variable is the weighted average price of residential natural gas and electricity prices. Each observation is a state-month. As shown in Table 1, columns 1 and 2, home en-

[6]Natural gas prices and quantities are provided on a volumetric basis. To make these data comparable across the sample, we use EIA data on the heat content of natural gas supplied to residential customers.

ergy prices are strongly positively correlated with $ShareGas \times log(RelPrice)$. In column 1, we include only state and month-year fixed effects. In column 2, we add the housing price index, unemployment rate, share of income from manufacturing, interactions of $log(RelPrice)$ with median income and the share of people over age 70, and air pollution indices.

The coefficient on $ShareGas \times log(RelPrice)$ is less than 1 for several reasons. First and foremost, the outcome is average *energy* prices, while the regressor is intended to proxy for average *heating* prices. In addition, the outcome is average prices weighted by usage, so it also incorporates any responses of usage to prices.[7] Note that to convert our mortality results into an elasticity of mortality with respect to the heating price, the relevant scale factor is 1, not the smaller coefficient estimated here; a change in $ShareGas \times log(RelPrice)$ can still be interpreted as a proportional change in the heating price faced by a household.

We next quantify how households' energy use responds to higher prices and then the impact on their energy bills. (In principle, once we know one of these numbers, we could calculate the other, but showing both is useful given that the data are available at different geographic levels and based on different samples.) We start by examining the impact on usage using EIA data, shown in Table 1, columns 3 and 4. As expected, higher prices lead to less consumption. Both the outcome and key regressor are in logs, so the coefficient represents an elasticity. The coefficient of -0.14 implies that households cut back usage quite a bit, but not one-for-one with price. To quantify the elasticity, one needs to scale the coefficient by the corresponding price-change coefficient from columns 1 and 2. We report this implied elasticity, which is around -0.35, at the bottom of the table. This elasticity is a similar magnitude as the winter natural gas demand elasticity for California estimated by Auffhammer and Rubin (2018). In Appendix Table A2, we show that the estimates based on our triple difference specification are similar.[8]

The elasticity having a magnitude smaller than 1 implies that households are spending more money on energy expenses when the heating price increases. We verify this using household-level Census/ACS data. Columns 5 and 6 of Table 1 show that the heating price shock is associated with a 24 log point increase in energy expenses. If the result is driven by changes in winter expenses, then the coefficient is an underestimate of the impact during winter months. (ACS does not release the survey month or distribution of months when

[7] Also, we construct $ShareGas$ weighting each household equally, whereas EIA's usage-weighted measure implicitly weights bigger users more. There might also be some measurement error in $ShareGas$.

[8] Appendix Table A3 reports results varying how we construct $RelPrice$, in particular using different lags.

surveys are completed. The Census asks about annual spending on energy bills.) Column 7 and 8 examine the outcome in levels: a 10% increase in the price of heating is associated with a \$5 increase in the monthly home energy bill, averaged over the year. To help interpret these magnitudes, note that the relative price of natural gas fell by 42% (54 log points) between 2005 and 2010. This price decline led to a 12% or \$315 annual decrease in energy bills for natural gas users, using the estimates in columns 6 and 8, respectively.

To summarize, we find that households reduce usage in response to an increase in their heating price, but not fully, so they also experience a meaningful increase in energy bills when the price of heating increases.

4.2 Effect of heating price on mortality

We examine the effect of heating prices on the log of the age-adjusted mortality rate, following Stevens et al. (2015).[9] We focus on "EWM causes of death," that is, causes with a pronounced peak in winter. Focusing on EWM causes provides a more honed test of the hypothesis that heating prices affect deaths due to exposure to cold and increases statistical power. We report effects on all-cause mortality in Appendix Table A5.[10]

Table 2 shows that a higher heating price increases mortality from EWM causes. Column 1 includes as regressors only county and year-month fixed effects in addition to $ShareGas \times log(RelPrice)$. Column 2 adds in our full set of control variables, listed earlier. In column 2 (our preferred specification), the elasticity of EWM mortality with respect to price is 0.057. Given that our EWM causes account for 63% of total mortality in winter, the implied elasticity of total mortality is 0.036, which is similar to the estimated elasticity of 0.030 when we directly examine all-cause mortality.[11]

The next columns show the effects disaggregated by broad category: non-viral non-respiratory infections; neurological; circulatory; and respiratory. The overall effect on EWM mortality is mainly driven by circulatory and respiratory causes. Appendix Table A6 reports results separately for each of the 14 EWM causes. The largest effect size is seen for emphysema, other chronic lower respiratory diseases, acute myocardial infarction, and pneumonia.

[9] We show robustness to using the age-adjusted mortality rate in levels in Appendix Table A4.

[10] We pursued also estimating effects on morbidity using the Heath and Retirement Study and on hospitalizations using the National Inpatient Sample, but due to the smaller sample sizes, we are underpowered to detect even elasticities considerably larger than our mortality estimates.

[11] When we estimate our main specification for non-EWM causes, the coefficient on $ShareGas \times log(RelPrice)$ is 0.006 and statistically insignificant. As the income channel should affect non-EWM mortality too, this pattern is suggestive of the importance of the substitution channel.

Interestingly, the price of heating does not seem to exacerbate influenza deaths.

The effects we estimate are not due to deaths being moved earlier by just a short duration, or "harvesting." Appendix Table A7 shows that the cumulative mortality effect is stable in magnitude when we incorporate effects in subsequent months. The cumulative effect is statistically significant when we add up to three subsequent months; there is not enough statistical power to determine at what point it becomes zero.

Appendix Tables A8 and A9 show robustness of the results to varying the definition of winter; excluding shale-gas-producing states; using residential instead of citygate, or annual instead of monthly, natural gas prices; population-weighting the regressions; dropping the Great Recession period; controlling for the state's per capita spending on the Low Income Home Energy Assistance Program (LIHEAP)[12]; and varying the main set of control variables.

We next bring in data for non-winter months to estimate triple difference models. We use either $Winter$ or HDD as the third difference. Table 2, column 7 shows that the effect of heating prices on mortality is stronger in winter than the rest of the year. Reassuringly, the coefficient on $ShareGas \times log(RelPrice)$ is close to zero: the price of heating has no effect on mortality in non-winter months.

Using HDD, we find that the price of heating increases mortality more in colder months. HDD is normalized so that a unit change is the difference between every day in the month being 65°F or above and being 32°F. A one-unit increase in HDD_{jt}, relative to the county's average winter HDD, which we control for in parallel to HDD_{jt}, leads to a 0.084 higher elasticity of mortality with respect to heating price.

The results are similar but somewhat weaker when we do not control for average HDD (see Appendix Table A10), which is consistent with previous findings that, due to adaptation (e.g., better insulated homes in colder places), atypical cold for an area is what especially affects mortality (Eurowinter Group 1997). As another reassuring placebo test, when $HDD = 0$ in this specification, the heating price does not affect mortality.

Table 3 returns to the difference-in-differences model to examine heterogeneous effects

[12] LIHEAP provides assistance with energy bills to low-income households. On average between 2001 and 2010, 4.5% of US households received LIHEAP heating assistance per year, which is 23% of households below 150% of the poverty line. LIHEAP pays eligible households a preset amount each year based on income and household size, and, depending on the state, also fuel type or the last year's utility bills. Arizona is, to our knowledge, the only state that varies the amount based on contemporaneous bills or prices (LIHEAP Clearinghouse 2010). LIHEAP state-year spending data are from the Department of Health and Human Services.

by poverty. Heating bills comprise a larger share of expenditures for the poor. For this and other reasons, we expect heating prices to have larger effects on mortality among the poor. Columns 1 to 4 each use a different poverty proxy. In column 1, the proxy is whether the county's median income is in the bottom half of the distribution across counties. Columns 2 and 3 use the proportion of households in the county that are below 150% of the federal poverty line, as either a continuous variable or an indicator for the county having a below-median proportion of households below the income threshold. Column 4 uses the decedent's education level, specifically an indicator for no high school degree. Across the board, the point estimates suggest larger effects among the poor, but the finding is only statistically significant in column 2 ($p < 0.10$) and column 3 ($p < 0.05$), which use the proportion of households below 150% of the poverty line.

Another dimension of heterogeneity we examined is sex. Table 3, column 5, reports that the mortality effects do not significantly differ by sex.

5 Conclusion

This paper finds that winter mortality is lower when the price of heating is lower. To put the estimated elasticity of all-cause mortality with respect to the price of heating of 0.03 in context, the price of natural gas relative to electricity fell by 42% between 2005 to 2010. Our findings imply that this price decline caused a 1.6% decrease in the winter mortality rate for households using natural gas for heating. Given that 58% of American households use natural gas for heating, the drop in natural gas prices lowered the US winter mortality rate by 0.9%, or, equivalently, the annual mortality rate by 0.4%. This represents over 11,000 deaths per year.

This effect size is large enough that it should not be ignored when assessing the net health effects of shale production of natural gas. The findings also highlight the health benefits of other policies to reduce home energy costs, particularly for low-income households.

References

Analitis, A., A. Biggeri, A. Hojs, B. Forsberg, E. Cadum, F. Ballester, J. Sunyer, K. Katsouyanni, L. Bisanti, M. Baccini, P. Michelozzi, P. Tiittanen, P. G. Goodman, and U. Kirchmayer (2008). "Effects of Cold Weather on Mortality: Results From 15 European Cities Within the PHEWE Project". *American Journal of Epidemiology* 168 (12), pp. 1397–1408.

Auffhammer, M. and E. Rubin (2018). *Natural Gas Price Elasticities and Optimal Cost Recovery Under Consumer Heterogeneity: Evidence from 300 million natural gas bills.* Working Paper 24295. National Bureau of Economic Research.

Bartik, A. W., J. Currie, M. Greenstone, and C. R. Knittel (forthcoming). "The Local Economic and Welfare Consequences of Hydraulic Fracturing". *American Economic Journal: Applied Economics*.

Beatty, T. K., L. Blow, and T. F. Crossley (2014). "Is There a 'Heat-or-Eat' Trade-off in the UK?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 177 (1), pp. 281–294.

Bhattacharya, J., T. DeLeire, S. Haider, and J. Currie (2003). "Heat or Eat? Cold-Weather Shocks and Nutrition in Poor American Families". *American Journal of Public Health* 93 (7), pp. 1149–1154.

Brown, S. P. and M. K. Yücel (2008). "What Drives Natural Gas Prices?" *The Energy Journal*, pp. 45–60.

Casey, J. A., D. A. Savitz, S. G. Rasmussen, E. L. Ogburn, J. Pollak, D. G. Mercer, and B. S. Schwartz (2016). "Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA". *Epidemiology* 27 (2), p. 163.

Crawford, V., M McCann, and R. Stout (2003). "Changes in Seasonal Deaths from Myocardial Infarction". *QJM* 96 (1), pp. 45–52.

Critchley, R., J. Gilbertson, M. Grimsley, G. Green, and Warm Front Study Group (2007). "Living in Cold Homes after Heating Improvements: Evidence from Warm-Front, England's Home Energy Efficiency Scheme". *Applied Energy* 84 (2), pp. 147–158.

Crossley, T. F. and F. Zilio (2018). "The Health Benefits of a Targeted Cash Transfer: The UK Winter Fuel Payment". *Health Economics* 27 (9), pp. 1354–1365.

Cullen, J. A. and E. T. Mansur (2017). "Inferring Carbon Abatement Costs in Electricity Markets: A Revealed Preference Approach Using the Shale Revolution". *American Economic Journal: Economic Policy* 9 (3), pp. 106–33.

Cullen, J. B., L. Friedberg, and C. Wolfram (2004). "Consumption and Changes in Home Energy Costs: How Prevalent is the 'Heat or Eat' Decision?" University of California, Berkeley manuscript.

Currie, J., M. Greenstone, and K. Meckel (2017). "Hydraulic Fracturing and Infant Health: New Evidence from Pennsylvania". *Science Advances* 3 (12), e1603021.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris (2008). "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States". *International Journal of Climatology* 28 (15), pp. 2031–2064.

DeLeire, T., P. Eliason, and C. Timmins (2014). "Measuring the Employment Impacts of Shale Gas Development". Unpublished manuscript, McCourt School of Public Policy, Georgetown University.

Deschenes, O. and E. Moretti (2009). "Extreme Weather Events, Mortality, and Migration". *Review of Economics and Statistics* 91 (4), pp. 659–681.

El Ansari, W. and S. El-Silimy (2008). "Are Fuel Poverty Reduction Schemes Associated with Decreased Excess Winter Mortality in Elders? A Case Study from London, UK". *Chronic Illness* 4 (4), pp. 289–294.

Eurowinter Group (1997). "Cold Exposure and Winter Mortality from Ischaemic Heart Disease, Cerebrovascular Disease, Respiratory Disease, and All Causes in Warm and Cold Regions of Europe". *The Lancet* 349 (9062), pp. 1341–1346.

Fell, H. and D. T. Kaffine (2018). "The Fall of Coal: Joint Impacts of Fuel Prices and Renewables on Generation and Emissions". *American Economic Journal: Economic Policy* 10 (2), pp. 90–116.

Feyrer, J., E. T. Mansur, and B. Sacerdote (2017). "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution". *American Economic Review* 107 (4), pp. 1313–1334.

Frank, D. A., N. B. Neault, A. Skalicky, J. T. Cook, J. D. Wilson, S. Levenson, A. F. Meyers, T. Heeren, D. B. Cutts, P. H. Casey, M. M. Black, and C. Berkowitz (2006). "Heat or Eat: The Low Income Home Energy Assistance Program and Nutritional and Health Risks Among Children Less Than 3 Years of Age". *Pediatrics* 118 (5), e1293–e1302.

Green, G. and J. Gilbertson (2008). "Warm Front, Better Health: Health Impact Evaluation of the Warm Front Scheme". *Sheffield: Sheffield Hallam University.*

Grey, C. N. B., T. Schmieder-Gaite, S. Jiang, C. Nascimento, and W. Poortinga (2017). "Cold Homes, Fuel Poverty and Energy Efficiency Improvements: A Longitudinal Focus Group Approach". *Indoor and Built Environment* 26 (7), pp. 902–913.

Groundwater Protection Council (2009). *Modern Shale Gas Development in the United States: A Primer*. Research Report. Prepared for the U.S. Department of Energy, National Energy Technology Laboratory with ALL Consulting, Oklahoma City, OK.

Hartley, P. R., K. B. Medlock, and J. E. Rosthal (2008). "The Relationship of Natural Gas to Oil Prices". *The Energy Journal*, pp. 47–65.

Hausman, C. and R. Kellogg (2015). "Welfare and Distributional Implications of Shale Gas". *Brookings Papers on Economic Activity*.

Hill, E. L. (2018). "Shale Gas Development and Infant Health: Evidence from Pennsylvania". *Journal of Health Economics* 61, pp. 134–150.

Holladay, J. S. and J. LaRiviere (2017). "The Impact of Cheap Natural Gas on Marginal Emissions from Electricity Generation and Implications for Energy Policy". *Journal of Environmental Economics and Management* 85, pp. 205 –227.

Howden-Chapman, P., A. Matheson, J. Crane, H. Viggers, M. Cunningham, T. Blakely, C. Cunningham, A. Woodward, K. Saville-Smith, D. O'Dea, M. Kennedy, M. Baker, N. Waipara, R. Chapman, and G. Davie (2007). "Effect of Insulating Existing Houses on Health Inequality: Cluster Randomised Study in the Community". *BMJ* 334 (7591), p. 460.

Hoynes, H., D. L. Miller, and J. Schaller (2012). "Who Suffers During Recessions?" *Journal of Economic Perspectives* 26 (3), pp. 27–48.

Jackson, R. B., A. Vengosh, J. W. Carey, R. J. Davies, T. H. Darrah, F. O'Sullivan, and G. Pétron (2014). "The Environmental Costs and Benefits of Fracking". *Annual Review of Environment and Resources* 39, pp. 327–362.

Jacobsen, G. D. (2019). "Who Wins in an Energy Boom? Evidence from Wage Rates and Housing". *Economic Inquiry* 57 (1), pp. 9–32.

Kearney, M. S. and R. Wilson (2018). "Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom". *Review of Economics and Statistics* 100 (4), pp. 678–690.

Knittel, C. R., K. Metaxoglou, and A. Trindade (2015). *Natural Gas Prices and Coal Displacement: Evidence from Electricity Markets*. Working Paper 21627. National Bureau of Economic Research.

Liddell, C. and C. Morris (2010). "Fuel Poverty and Human Health: a Review of Recent Evidence". *Energy Policy* 38 (6), pp. 2987–2997.

LIHEAP Clearinghouse (2010). *Targeting LIHEAP Benefits: State Strategies Based On Household Income, Energy Burden And Heating Costs*. Tech. rep. U.S. Department of Health and Human Services.

Linn, J. and L. Muehlenbachs (2018). "The Heterogeneous Impacts of Low Natural Gas Prices on Consumers and the Environment". *Journal of Environmental Economics and Management* 89, pp. 1 –28.

Manson, S., J. Schroeder, D. Van Riper, and S. Ruggles (2017). *IPUMS National Historical Geographic Information System: Version 12.0 [Database].* Minneapolis: University of Minnesota.

Muehlenbachs, L., E. Spiller, and C. Timmins (2015). "The Housing Market Impacts of Shale Gas Development". *American Economic Review* 105 (12), pp. 3633–59.

Myers, E. (forthcoming). "Are Home Buyers Myopic? Evidence From Capitalization of Energy Costs". *American Economic Journal: Economic Policy.*

PRISM Climate Group (2004). *PRISM Gridded Climate Data.* Oregon State University.

RECS (2009). *Residential Energy Consumption Survey.* Research Report. Energy Information Agency.

— (2014). *Everywhere But Northeast, Fewer Homes Choose Natural Gas as Heating Fuel.* Research Report. Energy Information Agency.

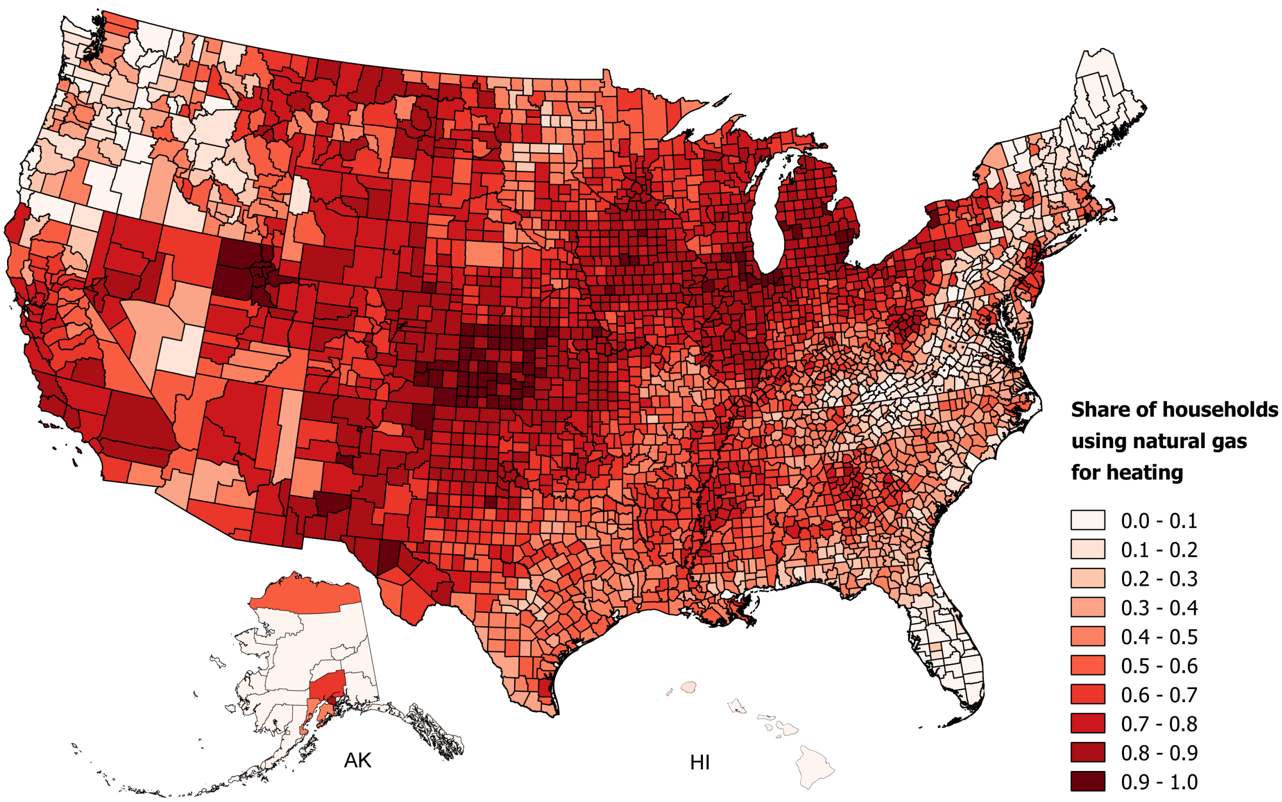
Stevens, A. H., D. L. Miller, M. E. Page, and M. Filipski (2015). "The Best of Times, the Worst of Times: Understanding Pro-Cyclical Mortality". *American Economic Journal: Economic Policy* 7 (4), pp. 279–311.

Wilkinson, P., S. Pattenden, B. Armstrong, A. Fletcher, R. S. Kovats, P. Mangtani, and A. J. McMichael (2004). "Vulnerability to Winter Mortality in Elderly People in Britain: Population Based Study". *BMJ* 329 (7467), p. 647.

Wilkinson, P., K. R. Smith, S. Beevers, C. Tonne, and T. Oreszczyn (2007). "Energy, Energy Efficiency, and the Built Environment". *The Lancet* 370 (9593), pp. 1175–1187.

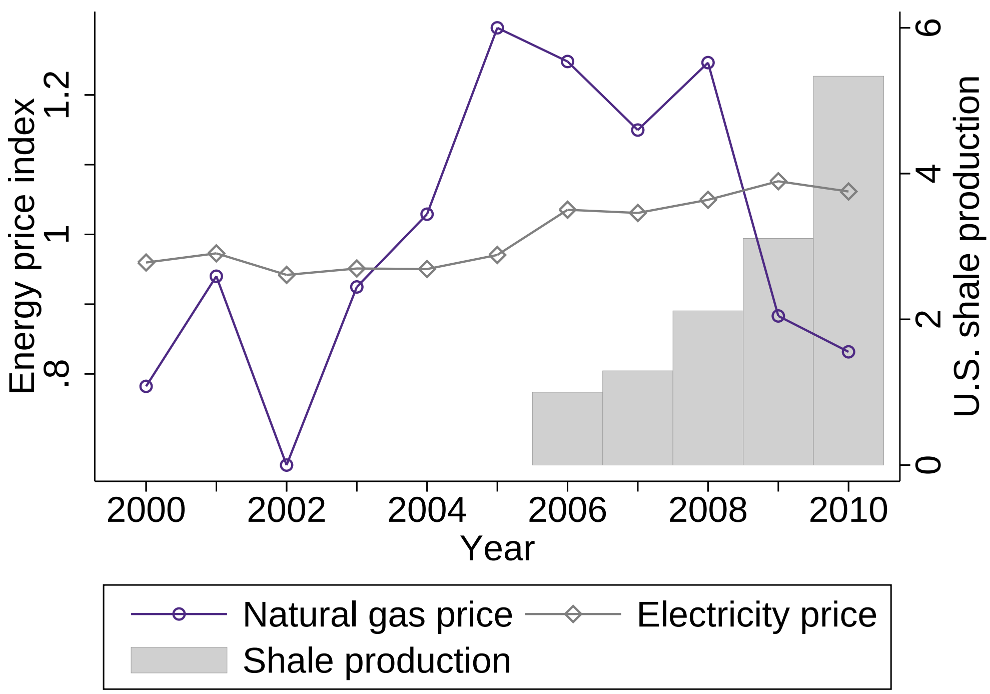
Ye, X., R. Wolff, W. Yu, P. Vaneckova, X. Pan, and S. Tong (2012). "Ambient Temperature and Morbidity: A Review of Epidemiological Evidence". *Environmental Health Perspectives* 120 (1), p. 19.

Figure 1: Share of households using natural gas for heating, by US county



Notes: The figure shows the proportion of occupied housing units in each county that report using natural gas as their main heating source. Data are from the 2000 US Census.

Figure 2: US natural gas and electricity prices, 2000 to 2010



Notes: The data series depicted with lines are the national prices of natural gas and electricity, normalized by their respective averages between 2000 and 2010 (left axis). National shale gas production in trillion cubic feet is shown as the bar chart (right axis). Data are from the US Energy Information Administration.

Table 1: Effect of heating price on energy use and energy spending

	Dependent variable:							
	Log of average electricity and gas price		Log of total energy consumption		Log of total monthly energy bill		Total monthly energy bill	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ShareGas × Log(RelPrice)	0.40***	0.39***	-0.14**	-0.14**	0.25***	0.24***	53.2***	48.7***
	[0.068]	[0.074]	[0.042]	[0.050]	[0.044]	[0.050]	[11.5]	[12.3]
Observations	2,695	2,695	2,695	2,695	14,385	14,385	14,385	14,385
Mean price/quantity	21.1	21.1	22.1	22.1	231.6	231.6	231.6	231.6
Basic fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All other controls	No	Yes	No	Yes	No	Yes	No	Yes
Implied elasticity			-0.36	-0.35				

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Columns 1 to 4: The sample comprises state-year-months in the contiguous U.S. for winter months (November–March) between 2000 and 2010. Outcomes are constructed from EIA data. Columns 5 to 8: The sample comprises PUMA-years in the contiguous U.S. included in the 2000 Census and the ACS PUMS data between 2005 and 2010. Outcomes are constructed from Census/ACS data. *Log(RelPrice)* is the log of the ratio of the US monthly citygate price of natural gas, averaged over the three- and four-month lag, to the US monthly residential price of electricity, also averaged over the three- and four-month lag. *ShareGas* is the proportion of occupied housing units in the state or PUMA in 2000 with natural gas as their main heating source. Average electricity and gas price is the state's consumption-weighted average of the residential prices of electricity and gas, in dollars per million British Thermal Units (BTUs). Total energy consumption is the state's total delivery of natural gas and electricity to residential consumers, in trillion BTUs. Total monthly energy bill is the mean monthly bill from electricity, gas and other fuels in the PUMA. *Basic fixed effects* are state and year-month fixed effects for columns 1 to 4, and PUMA and year fixed effects for columns 5 to 8. *All other controls* are the interactions of *log(RelPrice)* with the log of the median state or PUMA household income in 1999 and the share of people aged 70 and above in 2000, the state housing price index, the unemployment rate, the state's manufacturing sector share of total employee compensation, and the AQIs for $PM_{2.5}$, PM_{10}, and NO_2. Implied elasticity is the ratio of the coefficient reported in that column to the corresponding coefficient from the first two columns (the "first stage"). Monetary variables are in constant 2016 US dollars.

Table 2: Effect of heating price on mortality from EWM causes of death

	Dependent variable: Log of EWM-causes mortality rate							
	All EWM causes	All EWM causes	Group A: Non-viral, non-respiratory infections	Group G: Neurological diseases	Group I: Circulatory system diseases	Group J: Respiratory system diseases	All EWM causes	All EWM causes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ShareGas × Log(RelPrice)	0.052**	0.057***	0.021	0.016	0.054**	0.10***	-0.013	0.080**
	[0.022]	[0.021]	[0.027]	[0.034]	[0.024]	[0.024]	[0.015]	[0.039]
ShareGas × Log(RelPrice) × Winter							0.073***	
							[0.019]	
ShareGas × Log(RelPrice) × HDD								0.084***
								[0.031]
Observations	152,927	152,927	108,659	110,742	151,589	148,583	366,668	366,668
Mean mortality rate	577.6	577.6	74.16	74.01	371.8	259.8	527.8	527.8
Basic fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All other controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. between 2000 and 2010. In columns 1 to 6, the sample is restricted to winter months (November–March). Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. *Winter* is a binary variable that equals one in winter months (November to March). *HDD* is the number of heating degree-days in the county for the month, based on a threshold of 65°F, in units of °F-days divided by 1000, and scaled to a 30-day month. *Basic fixed effects* are county and year-month fixed effects. Columns 1 to 6: *All other controls* are the interactions of *log(RelPrice)* with the log of the median county household income in 1999 and the share of people aged 70 and above in 2000, the state housing price index, the unemployment rate, the state's manufacturing sector share of total employee compensation, and the AQIs for $PM_{2.5}$, PM_{10}, and NO_2. Column 7 and 8: *All other controls* is the above set plus the following: all possible two-way interactions between *ShareGas*, *log(RelPrice)*, and the triple difference variable (either *Winter* or *HDD*), unless collinear with year-month fixed effects; and the two- and three-way interactions among *log(RelPrice)*, *Winter/HDD*, and each of the log of the median county household income in 1999 and the share of people aged 70 and above in 2000. Column 8 also includes *HDD*; the interaction of the average county HDD in winter months with *log(RelPrice)*; and the three-way interactions of the average county HDD in winter months, *log(RelPrice)*, and each of *ShareGas*, the log of the median county household income in 1999, and the share of people aged 70 and above in 2000.

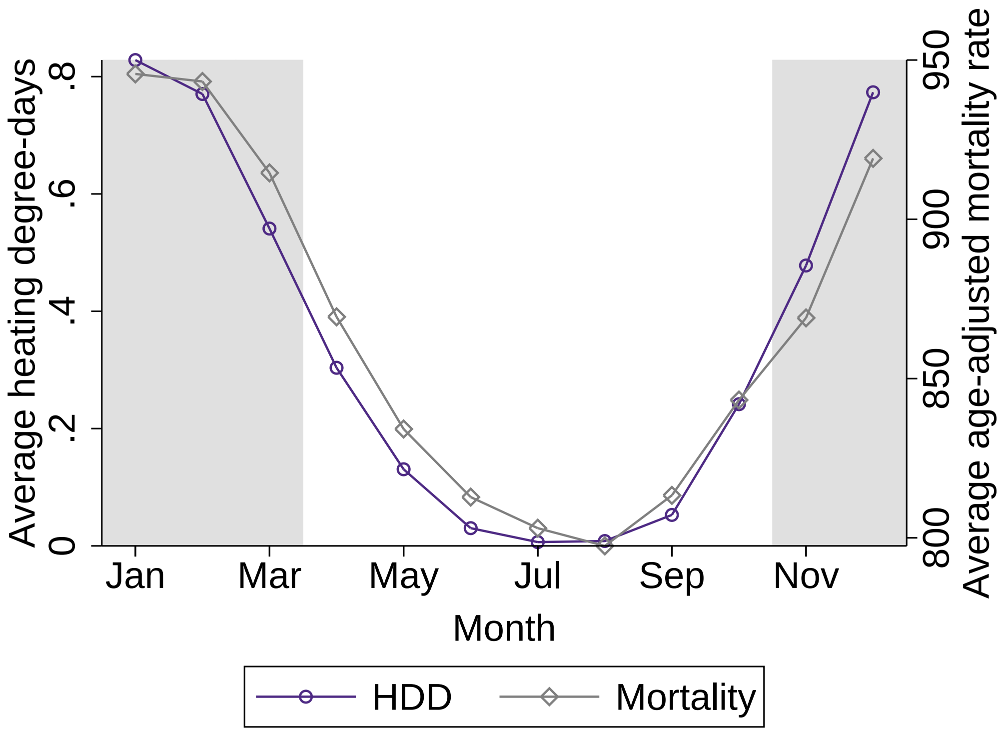
Table 3: Heterogeneous effects on mortality

	Dependent variable: Log of all-EWM-causes mortality rate Trait is:				
	Below-median county income	Proportion below 150% of poverty line	Above-median proportion below 150% of poverty line	No high school degree	Male
	(1)	(2)	(3)	(4)	(5)
ShareGas × Log(RelPrice) × Trait	0.019	0.34*	0.060**	0.038	0.013
	[0.035]	[0.20]	[0.030]	[0.045]	[0.027]
ShareGas × Log(RelPrice)	0.047**	-0.025	0.033*	0.019	0.057***
	[0.019]	[0.045]	[0.019]	[0.052]	[0.019]
Observations	152,927	152,927	152,927	284,700	300,311
Mean mortality rate	577.6	577.6	577.6	999.4	605.3
Implied effect for Trait = 1	0.07*	0.31*	0.09***	0.06	0.07**
	[0.03]	[0.16]	[0.03]	[0.06]	[0.03]

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. For columns 1 to 3, the sample comprises county-year-months in the contiguous U.S. for winter months (November–March) between 2000 and 2010. For columns 4 and 5, the sample comprises county-year-months-education group and county-year-months-sex group respectively for winter months. Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. Column 1: *Trait* is an indicator variable that equals one if the county's median household income is below the median of all counties in the sample in 1999. Column 2: *Trait* is the proportion of households in the county with income in 1999 below 150 percent of the poverty threshold. Column 3: *Trait* is an indicator variable that equals one if the proportion from column 2 is above the median of all counties in the sample. Column 4: *Trait* is an indicator variable that equals one for the subgroup that did not complete high school. Column 5: *Trait* is an indicator variable that equals one for the male population. All columns include all controls from column 2 of Table 2, the main effect for *Trait*, and the interaction of each control variable with *Trait*.

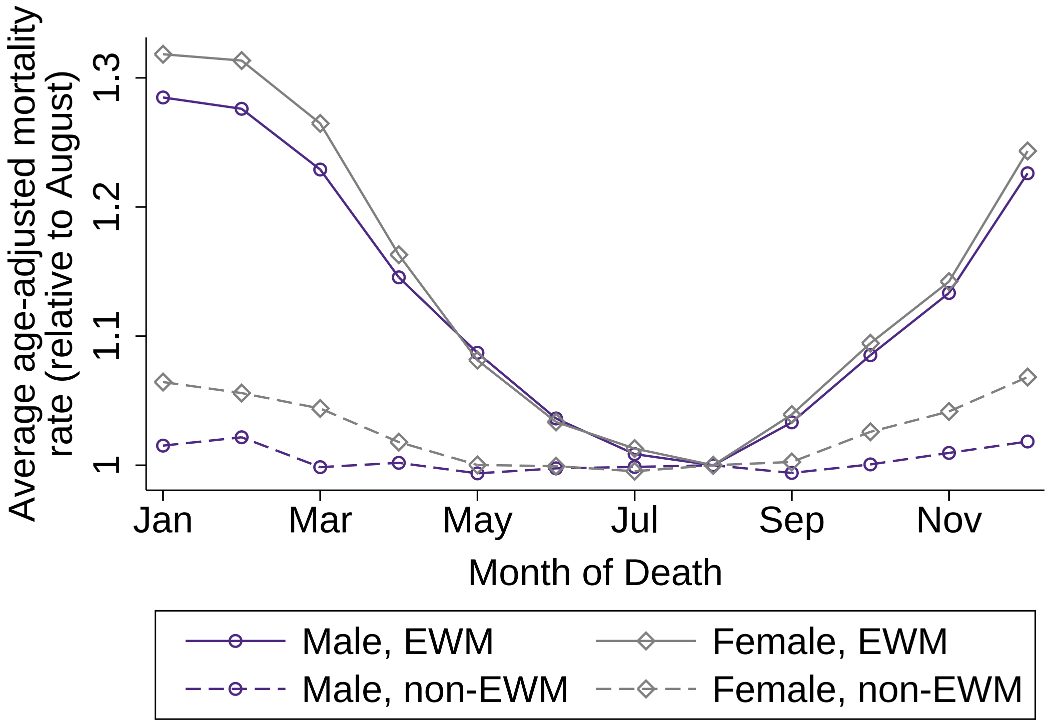
A Appendix figures and tables

Appendix Figure A1: Heating degree-days and monthly mortality in the US



Notes: Average heating degree-days (HDD) and average age-adjusted mortality rates across US counties (excluding Hawaii and Alaska) between 2000 and 2010 plotted by month. Average HDD is computed using temperature data from the PRISM Climate Group, and is based on a threshold of 65°F, in units of °F-days divided by 1000, and scaled to a 30-day month. Average age-adjusted mortality rates are computed using the NCHS mortality data and expressed per 100,000 population on an annualized basis. Months we define as winter in our analysis (November–March) are shaded in the background.

Appendix Figure A2: Seasonality in mortality for EWM and non-EWM causes



Notes: Average age-adjusted mortality rates across US counties (excluding Hawaii and Alaska) between 2000 and 2010, broken down by sex and EWM versus other causes. EWM causes are those that exhibit a strong pattern of higher mortality in winter than the rest of the year, as described in the text; see data appendix for further details. We normalize each series by its value in August (the month with the lowest all-cause mortality rate). Age-adjusted mortality rates are computed using the NCHS mortality data.

Appendix Table A1: Causes of death exhibiting high excess winter mortality

Cause of death (ICD-10 codes)	Mean monthly mortality rate	Level coefficient	Log coefficient
Septicemia (A40-A41)	0.95	0.14	0.14
Parkinson's disease (G20-G21)	0.53	0.08	0.16
Alzheimer's disease (G30)	1.92	0.36	0.18
Acute myocardial infarction (I21-I22)	4.34	0.62	0.14
All other forms of chronic ischemic heart disease (I20, I25.1-I25.9)	6.32	0.80	0.12
Heart failure (I50)	1.61	0.21	0.13
Cerebrovascular diseases (I60-I69)	4.12	0.52	0.12
Atherosclerosis (I70)	0.30	0.04	0.14
Influenza (J09-J11)	0.04	0.06	2.21
Pneumonia (J12-J18)	1.63	0.58	0.34
Emphysema (J43)	0.38	0.08	0.21
Other chronic lower respiratory diseases (J44, J47)	3.11	0.63	0.20
Pneumonitis due to solids and liquids (J69)	0.47	0.09	0.18
Other diseases of respiratory system (J00-J06, J30-J39, J67, J70-J98)	0.77	0.11	0.14
All other diseases (Residual)*	6.17	0.80	0.13
Accidental exposure to smoke, fire and flames (X00-X09)*	0.09	0.05	0.56

Notes: Mortality rates are expressed per 100,000 population and computed using the NCHS mortality data. The 75th percentile of level and log coefficient are 0.02 and 0.12, respectively. We remove *All other diseases* and *Accidental exposure to smoke, fire and flames* (marked with *) when we analyze mortality from high-EWM causes. See the data appendix for further details on the selection of high-EWM causes of deaths.

Appendix Table A2: Triple difference estimates of effects on average energy price and consumption

	Dependent variable: Log of average electricity and gas price		Dependent variable: Log of total energy consumption	
	(1)	(2)	(3)	(4)
ShareGas × Log(RelPrice) × Winter	0.36***		-0.12*	
	[0.048]		[0.050]	
ShareGas × Log(RelPrice) × HDD		0.36***		-0.091
		[0.074]		[0.067]
Observations	6,468	6,468	6,468	6,468
Mean price/quantity	25.5	25.5	16.0	16.0
Implied elasticity			-0.34	-0.25

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises state-year-months in the contiguous U.S. between 2000 and 2010. Average electricity and gas price is the state's consumption-weighted average of the residential prices of electricity and gas, in dollars per million BTUs. Total energy consumption is the state's total delivery of natural gas and electricity to residential consumers, in trillion BTUs. *ShareGas* is the proportion of occupied housing units in the state in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. *Winter* is a binary variable that equals one in winter months (November to March). *HDD* is the number of heating degree-days in the county for the month, based on a threshold of 65°F, in units of °F-days divided by 1000, and scaled to a 30-day month. Monetary variables are in constant 2016 US dollars. Implied elasticity is the ratio of the coefficient reported in that column to the corresponding coefficient from the first two columns (the "first stage"). All columns include covariates analogous to those used in columns 7 and 8 of Table 2.

Appendix Table A3: Effect of heating price on energy consumption, based on alternative *RelPrice* definitions

	Dependent variable: Log of total energy consumption						
	0 lags (contemporaneous)	1 lag	2 lags	3 lags	4 lags	5 lags	6 lags
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: RelPrice constructed based on residential gas price, at indicated lag.							
ShareGas × Log(RelPrice)	-0.18	-0.24*	-0.38*	-0.35*	-0.27	-0.17	-0.15
	[0.099]	[0.12]	[0.16]	[0.16]	[0.15]	[0.11]	[0.11]
Observations	2,695	2,695	2,695	2,695	2,695	2,695	2,695
Mean quantity	22.1	22.1	22.1	22.1	22.1	22.1	22.1
Implied elasticity	-0.23	-0.33	-0.59	-0.68	-0.58	-0.40	-0.41
Panel B: RelPrice constructed based on citygate gas price, at indicated lag.							
ShareGas × Log(RelPrice)	-0.023	-0.0014	-0.027	-0.12*	-0.15*	-0.12	-0.13
	[0.058]	[0.052]	[0.052]	[0.047]	[0.055]	[0.063]	[0.069]
Observations	2,695	2,695	2,695	2,695	2,695	2,695	2,695
Mean quantity	22.1	22.1	22.1	22.1	22.1	22.1	22.1
Implied elasticity	-0.053	-0.0035	-0.076	-0.31	-0.39	-0.38	-0.46

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises state-year-months in the contiguous U.S. for winter months (November–March) between 2000 and 2010. Total energy consumption is the state's total delivery of natural gas and electricity to residential consumers, in trillion BTUs. *ShareGas* is the proportion of occupied housing units in the state in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly residential or citygate price of natural gas in the US, lagged by the number of months indicated in each column, to the corresponding residential price of electricity. Monetary variables are in constant 2016 US dollars. All columns include all controls from Table 1. Implied elasticity is the ratio of the coefficient in that column to the corresponding coefficient for the "first stage" (not shown).

Appendix Table A4: Effect of heating price on mortality from high-EWM causes of death using mortality rate in levels

	Dependent variable: EWM-causes mortality rate							
	All EWM causes	All EWM causes	Group A: Non-viral, non-respiratory infections	Group G: Neurological diseases	Group I: Circulatory system diseases	Group J: Respiratory system diseases	All EWM causes	All EWM causes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ShareGas × Log(RelPrice)	32.1***	30.4***	0.14	-3.38	16.9**	23.4***	-2.75	27.9
	[11.3]	[11.3]	[2.80]	[2.57]	[8.35]	[6.15]	[6.51]	[22.1]
ShareGas × Log(RelPrice) × Winter							34.2***	
							[10.0]	
ShareGas × Log(RelPrice) × HDD								36.7**
								[15.4]
Observations	153,340	153,340	153,340	153,340	153,340	153,340	368,016	368,016
Mean mortality rate	576.0	576.0	52.55	53.45	367.5	251.7	525.9	525.9
Basic fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All other controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. between 2000 and 2010. In columns 1 to 6, the sample is restricted to winter months (November–March). Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. *Winter* is a binary variable that equals one in winter months (November to March). *HDD* is the number of heating degree-days in the county for the month, based on a threshold of 65˚F, in units of ˚F-days divided by 1000, and scaled to a 30-day month. *Basic fixed effects* are county and year-month fixed effects. Columns 1 to 6: *All other controls* are the interactions of *log(RelPrice)* with the log of the median county household income in 1999 and the share of people aged 70 and above in 2000, the state housing price index, the unemployment rate, the state's manufacturing sector share of total employee compensation, and the AQIs for $PM_{2.5}$, PM_{10}, and NO_2. Column 7 and 8: *All other controls* is the above set plus the following: all possible two-way interactions between *ShareGas*, *log(RelPrice)*, and the triple difference variable (either *Winter* or *HDD*), unless collinear with year-month fixed effects; and the two- and three-way interactions among *log(RelPrice)*, *Winter*/*HDD*, and each of the log of the median county household income in 1999 and the share of people aged 70 and above in 2000. Column 8 also includes *HDD*; the interaction of the average county HDD in winter months with *log(RelPrice)*; and the three-way interactions of the average county HDD in winter months, *log(RelPrice)*, and each of *ShareGas*, the log of the median county household income in 1999, and the share of people aged 70 and above in 2000.

Appendix Table A5: Effect of heating price on all-cause mortality

	Dependent variable: Log of mortality rate				Dependent variable: Mortality rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ShareGas × Log(RelPrice)	0.021	0.030*	-0.011	0.046*	29.1**	32.6**	-6.68	42.6*
	[0.016]	[0.016]	[0.0098]	[0.026]	[13.8]	[14.3]	[7.86]	[24.7]
ShareGas × Log(RelPrice) × Winter			0.042**				39.6***	
			[0.016]				[14.4]	
ShareGas × Log(RelPrice) × HDD				0.042*				41.4*
				[0.024]				[21.5]
Observations	153,296	153,296	367,905	367,905	153,340	153,340	368,016	368,016
Mean mortality rate	929.5	929.5	872.6	872.6	929.2	929.2	872.3	872.3
Basic fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All other controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. between 2000 and 2010. In columns 3, 4, 7 and 8, the sample is restricted to winter months (November–March). Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. *Winter* is a binary variable that equals one in winter months (November to March). *HDD* is the number of heating degree-days in the county for the month, based on a threshold of 65°F, in units of °F-days divided by 1000, and scaled to a 30-day month. *Basic fixed effects* are county and year-month fixed effects. Columns 1, 2, 5, and 6: *All other controls* are the interactions of *log(RelPrice)* with the log of the median county household income in 1999 and the share of people aged 70 and above in 2000, the state housing price index, the unemployment rate, the state's manufacturing sector share of total employee compensation, and the AQIs for $PM_{2.5}$, PM_{10}, and NO_2. Column 3, 4, 7, and 8: *All other controls* is the above set plus the following: all possible two-way interactions between *ShareGas*, *log(RelPrice)*, and the triple difference variable (either *Winter* or *HDD*), unless collinear with year-month fixed effects; and the two- and three-way interactions among *log(RelPrice)*, *Winter/HDD*, and each of the log of the median county household income in 1999 and the share of people aged 70 and above in 2000. Columns 4 and 8 also includes *HDD*; the interaction of the average county HDD in winter months with *log(RelPrice)*; and the three-way interactions of the average county HDD in winter months, *log(RelPrice)*, and each of *ShareGas*, the log of the median county household income in 1999, and the share of people aged 70 and above in 2000.

Appendix Table A6: Effect of heating price on mortality, by specific cause of death

Dependent variable: Log of specified disease mortality rate			
Septicemia	0.021 [0.027] {74.2}	Atherosclerosis	0.045 [0.057] {45.9}
Parkinson's disease	0.041 [0.031] {32.3}	Influenza	-0.20 [0.16] {27.3}
Alzheimer's disease	0.029 [0.037] {63.2}	Pneumonia	0.099*** [0.034] {104.9}
Acute myocardial infarction	0.10*** [0.036] {107.3}	Emphysema	0.14*** [0.047] {29.8}
Chronic ischemic heart disease	0.078** [0.030] {158.0}	Other chronic lower respiratory diseases	0.11*** [0.027] {114.2}
Heart failure	0.051* [0.026] {137.4}	Pneumonitis (solids and liquids)	0.039 [0.047] {44.4}
Cerebrovascular diseases	0.076** [0.034] {114.4}	Other respiratory diseases	0.047 [0.030] {107.4}

Notes: Each cell shows the result from a separate regression, and reports the coefficient on *ShareGas* $\times$ *log(RelPrice)*, the corresponding standard error clustered by state in square brackets, and the mean mortality rate of the specified cause in curly brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. for winter months (November–March) between 2000 and 2010. Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. All columns include all controls from column 2 of Table 2.

Appendix Table A7: Dynamic effects of heating price on mortality

	Dependent variable: Log of all-EWM-causes mortality rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Contemporaneous effect	-0.012	0.055	-0.029	-0.0043	-0.027	-0.0035
	[0.053]	[0.055]	[0.054]	[0.056]	[0.061]	[0.057]
Effect on mortality 1 month after	0.066	-0.036	0.11	0.038	0.095	0.075
	[0.053]	[0.080]	[0.094]	[0.085]	[0.10]	[0.098]
Effect on mortality 2 months after		0.046	-0.13	-0.016	-0.086	-0.11
		[0.056]	[0.10]	[0.098]	[0.11]	[0.11]
Effect on mortality 3 months after			0.13**	-0.026	0.090	0.15
			[0.063]	[0.091]	[0.11]	[0.12]
Effect on mortality 4 months after				0.043	-0.13	-0.18
				[0.046]	[0.10]	[0.14]
Effect on mortality 5 months after					0.11*	0.16
					[0.064]	[0.12]
Effect on mortality 6 months after						-0.030
						[0.059]
Observations	183,510	214,043	244,552	275,071	305,602	336,113
Cumulative effect	0.05**	0.07**	0.08**	0.03	0.05	0.05
	[0.02]	[0.03]	[0.03]	[0.03]	[0.04]	[0.04]

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. between 2000 and 2010. The sample is restricted to months November to April in column 1, November to May in column 2, November to June in column 3, November to July in column 4, November to August in column 5, and November to September in column 6. The specification used is $log(m_{jt}) = \sum_{k=0}^{K} \beta_k ShareGas_j \times log(RelPrice_{t-k}) \times MonthofEffect_k + \gamma_k ShareGas_j \times log(RelPrice_{t-k}) + Controls + \epsilon_{jt}$, where $MonthofEffect_0$ takes on a value of one in the months of November to March; $MonthofEffect_1$ takes on a value of one in the months of December to April; $MonthofEffect_2$ takes on a value of one in the months of January to May; $MonthofEffect_3$ takes on a value of one in the months of February to June; $MonthofEffect_4$ takes on a value of one in the months of March to July; $MonthofEffect_5$ takes on a value of one in the months of April to August; $MonthofEffect_6$ takes on a value of one in the months of May to September; and $Controls$ are all controls from column 2 of Table 2 and are fully interacted with the $MonthofEffect_k$ dummies. K, the total number of months after the contemporaneous effect, is 1 in column 1, 2 in column 2, and so on. The coefficients shown are β_k's, the effect of the winter price of heating on winter mortality k months after winter, after accounting for intertemporal correlation (since we estimate the β_k's jointly), and after removing the effect on mortality in irrelevant months through $MonthofEffect_k$ dummies (e.g., April is not a winter month, so is not relevant for the contemporaneous effect). *Cumulative effect* is the sum of the β_k's. All other definitions not noted above are as in column 2 of Table 2.

Appendix Table A8: Effect of heating price on mortality: Robustness checks

	Dependent variable: Log of all-EWM-causes mortality rate		
	Difference-in-differences (1)	Triple difference using winter (2)	Triple difference using HDD
Winter defined as December to March	0.048** [0.022]	0.065*** [0.021]	n/a
Winter defined as December to February	0.064** [0.028]	0.078*** [0.027]	n/a
Exclude fracking states	0.054** [0.021]	0.067*** [0.019]	0.096*** [0.030]
Use residential gas price, averaged over 2nd and 3rd lags	0.077** [0.031]	0.044 [0.036]	0.071 [0.056]
Use annual residential gas price	0.11** [0.041]	0.12*** [0.037]	0.073 [0.061]
Weight by population in 2000	0.049*** [0.016]	0.040*** [0.015]	0.050** [0.023]
Exclude Great Recession	0.035 [0.021]	0.067*** [0.020]	0.077** [0.031]
Control for Log(LIHEAP per capita)	0.056** [0.021]	0.073*** [0.019]	0.084*** [0.031]
Control for all pollutants	0.056** [0.021]	0.073*** [0.019]	0.085*** [0.031]

Notes: Each cell shows the result from a separate regression, and reports the coefficient on *ShareGas* × *log(RelPrice)* (column 1), *ShareGas* × *log(RelPrice)* × *Winter* (column 2), or *ShareGas* × *log(RelPrice)* × *HDD* (column 3). The corresponding standard error, clustered by state, is shown in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Each row shows a change in specification compared to columns 2, 7, and 8 of Table 2. Row 1, column 1: The sample additionally excludes November. Row 1, column 2: The sample excludes November, and uses December to March as winter months in the triple difference. Row 2, column 1: The sample additionally excludes November and March. Row 2, column 2: The sample excludes November and March, and uses December to February as winter months in the triple difference. Rows 1 and 2, column 3: The triple difference using HDD is the same as in the main specification since HDD is defined independently of winter. Row 3: The sample additionally excludes Arkansas, Louisiana, North Dakota, Oklahoma, Pennsylvania, Texas, and West Virginia, which are the states with significant production of shale natural gas. Row 4: *RelPrice* is constructed as the ratio of the monthly residential price of natural gas in the US, averaged over the two- and three-month lag, to the corresponding residential price of electricity. Row 5: *RelPrice* is constructed as the ratio of the annual residential price of natural gas in the US to the corresponding residential price of electricity. Row 6: The regression is weighted by the county population in 2000. Row 7: The sample excludes months between December 2007 and June 2009, inclusive. This is the period of the Great Recession as defined by the NBER Business Cycle Dating Committee. Row 8: The specification additionally includes the log of total LIHEAP assistance funds per capita in the state-fiscal year. Row 9: The specification additionally includes the AQIs of carbon monoxide, ozone, and sulfur dioxide as control variables. All other definitions not noted are as in columns 2, 7, and 8 of Table 2 respectively.

Appendix Table A9: Effect of heating price on mortality: Robustness to excluding control variables

	Dependent variable: Log of all-EWM-causes mortality rate		
	Difference-in-differences (1)	Triple difference using winter (2)	Triple difference using HDD (3)
Exclude housing price index	0.064***	0.074***	0.086***
	[0.019]	[0.019]	[0.031]
Exclude unemployment rate	0.055**	0.072***	0.083***
	[0.021]	[0.019]	[0.031]
Exclude manufacturing share	0.055**	0.072***	0.084***
	[0.021]	[0.019]	[0.031]
Exclude Log(Income) × Log(RelPrice)	0.039*	0.066***	0.075**
	[0.022]	[0.019]	[0.031]
Exclude Share70+ × Log(RelPrice)	0.057**	0.072***	0.083***
	[0.021]	[0.020]	[0.031]
Exclude all pollution controls	0.060***	0.075***	0.084***
	[0.021]	[0.019]	[0.031]
Exclude Share70+ × Log(RelPrice) and Log(Income) × Log(RelPrice)	0.043*	0.068***	0.075**
	[0.023]	[0.020]	[0.031]
Exclude unemployment rate, manufacturing share, and housing price index	0.060***	0.073***	0.085***
	[0.019]	[0.019]	[0.031]

Notes: Each cell shows the result from a separate regression, and reports the coefficient on *ShareGas* × *log(RelPrice)* (column 1), *ShareGas* × *log(RelPrice)* × *Winter* (column 2), or *ShareGas* × *log(RelPrice)* × *HDD* (column 3). The corresponding standard error, clustered by state, is shown in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Each row shows a change in specification compared to columns 2, 7, and 8 of Table 2, specifically the exclusion of the control variable(s) indicated in the first column and, where applicable, two-way and three-way interactions that include that variable. All other definitions not noted are as in columns 2, 7, and 8 of Table 2 respectively.

Appendix Table A10: Triple difference mortality estimates using alternative outcomes and specifications

	Dependent variable: Log of mortality rate			
	All causes	All EWM causes	Group I EWM: Circulatory system diseases	Group J EWM: Respiratory system diseases
	(1)	(2)	(3)	(4)
Panel A: Triple difference using winter.				
ShareGas × Log(RelPrice)	-0.011	-0.013	0.0093	-0.0080
	[0.0098]	[0.015]	[0.019]	[0.021]
ShareGas × Log(RelPrice) × Winter	0.042**	0.073***	0.045**	0.11***
	[0.016]	[0.019]	[0.021]	[0.028]
Panel B: Triple difference using HDD.				
ShareGas × Log(RelPrice)	0.046*	0.080**	0.089**	0.079*
	[0.026]	[0.039]	[0.042]	[0.040]
ShareGas × Log(RelPrice) × HDD	0.042*	0.084***	0.060*	0.10**
	[0.024]	[0.031]	[0.035]	[0.039]
Panel C: Without controlling in parallel for average winter HDD.				
ShareGas × Log(RelPrice)	-0.0063	-0.0052	0.016	0.0010
	[0.011]	[0.018]	[0.022]	[0.021]
ShareGas × Log(RelPrice) × HDD	0.037*	0.055*	0.030	0.078**
	[0.021]	[0.028]	[0.032]	[0.037]
Observations	367,905	366,668	362,930	353,692
Mean mortality rate	872.6	527.8	343.5	232.7

Notes: Standard errors clustered by state in brackets. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The sample comprises county-year-months in the contiguous U.S. between 2000 and 2010. Mortality rates are age-adjusted mortality rates expressed as annual deaths per 100,000 population; see data appendix for further details. *ShareGas* is the proportion of occupied housing units in the county in 2000 with natural gas as their main heating source. *Log(RelPrice)* is the log of the ratio of the monthly citygate price of natural gas in the US, averaged over the three- and four-month lag, to the corresponding residential price of electricity. *HDD* is the number of heating degree-days in the county for the month, based on thresholds of 65°F, in units of °F-days divided by 1000, and scaled to a 30-day month. All columns in panels A and B include the covariates from columns 7 and 8 respectively of Table 2. Panel C excludes from this set of control variables the two-way and three-way interactions based on the county's average HDD in winter months.

B Data appendix

Appendix Table B1 lists the data source for each of our outcome and independent variables. The following sections provide further description of our data sources and the construction of variables used in this paper.

B.1 Mortality rate

B.1.1 Data source

The data source for mortality is Vital Statistics records, specifically restricted-use "mortality files with all county geographical information" obtained from the National Center for Health Statistics (NCHS). These mortality files include a record for every death certificate filed in the United States during the study period. Each record includes a single underlying cause of death, up to twenty additional multiple causes, month of death, and demographic data, including the deceased's age, gender, race, Hispanic origin, education, county of residence and county of death. The definition of the underlying cause of death follows that of the World Health Organization (WHO): the disease or injury which initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury. Causes of death are classified using the Tenth Revision of the International Classification of Disease (ICD-10) during the 2000 to 2010 study period.

We compute mortality rates by county, classifying individuals by their county of residence. We restrict our analyses to the contiguous US throughout the paper. We account for substantial county boundary changes by aggregating counties to a larger stable unit.[13] Specifically, we combine Adams, Broomfield, Boulder, Jefferson, and Weld counties in Colorado; Prince George's and Montgomery in Maryland; Craven and Carteret in North Carolina; Franklin and Gulf in Florida; Bedford and Bedford City in Virginia; Alleghany and Clifton Forge in Virginia; Augusta and Waynesboro in Virginia; Prince William and Manassas Park in Virginia; Southampton and Franklin in Virginia; and York and Newport News in Virginia.

In addition, when analyzing county-level data, we exclude counties whose population aged 50 and over in 2000 are in the lowest decile of the full sample to reduce noise from mortality rates of counties with small population and missing observations when we use the logarithm of the mortality rate.

B.1.2 Calculating age-adjusted mortality rate

To calculate mortality rates, we use population data from the National Cancer Institutes's Surveillance Epidemiology and End Results (Cancer-SEER) program. These data give yearly county population estimates by age group, sex, race, and Hispanic origin.[14] For 2005, we use the SEER's adjusted set of population estimates that takes into account population shifts due to Hurricanes Katrina and Rita.

We use these population estimates to calculate both crude and age-adjusted mortality rates, expressed per 100,000 population. The crude mortality rate at county-year-month level is the total number of deaths in that county in that year-month divided by its population estimate in that year. The age-adjusted mortality rate is a weighted average of the crude

[13] Information on substantial county boundary changes was taken from the Census Bureau's website.

[14] We use vintage 2014 population estimates. The data and documentation are available at `https://seer.cancer.gov/popdata/`.

mortality rates across age categories, where the shares of each age category in the whole U.S. population are used as weights.[15] We use the age distribution of US population in 2000 (the "U.S. 2000 standard population") published by SEER as weights in the calculation of age-adjusted mortality rates. All mortality rates in the paper are expressed on an annual basis obtained by multiplying the month-level mortality rates by (365/number of day in that month).

B.1.3 Selection of causes of deaths

We use a data-driven approach to select causes of deaths that exhibit significant "excess winter mortality" (EWM), or higher mortality in winter months than in other months.

We use the NCHS's 113 Selected Causes of Death, which represent groupings of detailed ICD-10 codes, as the mutually exclusive set of causes of death. To measure the degree of EWM for each cause, we construct an observation for each month in the 2000 to 2010 period (132 observations) and calculate the total deaths in the US, by cause, in that month. For each cause separately, we run a regression of the number of deaths due to that cause (i.e., as the underlying cause of death) on a *Winter* dummy, which equals 1 for November to March, and year fixed effects. A similar set of regressions is estimated with the logarithm of deaths as the outcome instead of the level. We then select causes whose *Winter* coefficient is in the top quartile among all causes of deaths in both levels and logs (i.e., above 0.12 for logs and 0.02 for levels). We use both levels and logs of mortality because we want to select causes that are both common and have a strong degree of excess winter mortality.

We exclude two causes from the data-driven list of excess winter mortality causes: first, *Accidental exposure to smoke, fire and flames*, since accidental deaths that are not a physiological result of exposure cold differ from our focus, and second, *All other diseases* (the residual category), since it is difficult to verify the mechanism for this "cause." Appendix Table A1 reports *Winter* coefficients in levels and logs and average monthly crude mortality rate for each of the selected causes. The final selected list includes the following fourteen causes of death, with their ICD-10 codes in brackets. These causes can be further grouped into four broader cause groups.

- **Group A:** Non-viral, non-respiratory infections
  - Septicemia (A40-A41)
- **Group G:** Neurological diseases
  - Parkinson's disease (G20-G21)
  - Alzheimer's disease (G30)
- **Group I:** Circulatory system diseases
  - Acute myocardial infarction (I21-I22)
  - All other forms of chronic ischemic heart disease (I20, I25.1-I25.9)
  - Heart failure (I50)

[15] We use the following 19 age categories: under 1 year, 1-4 years, 5-9 years, ..., 80-84 years, and 85 years and over.

  - Cerebrovascular diseases (I60-I69)
  - Atherosclerosis (I70)

- **Group J:** Respiratory system diseases
  - Influenza (J09-J11)
  - Pneumonia (J12-J18)
  - Emphysema (J43)
  - Other chronic lower respiratory diseases (J44, J47)
  - Pneumonitis due to solids and liquids (J69)
  - Other diseases of respiratory system (J00-J06, J30-J39, J67, J70-J98)

B.2 Home energy price and usage

All energy prices and consumption data come from monthly series published by the U.S. Energy Information Administration (EIA), available at the state and national level. The data are based on samples of firms supplying natural gas or electricity to residential consumers, and include some processing by the EIA to account for non-response.[16]

The raw data express quantities in kilowatt-hours for electricity and cubic feet for natural gas. To allow comparison between energy types, we convert these quantities to British Thermal Units (BTU). The conversion is straightforward for electricity. For natural gas, we apply estimates of the heat content of natural gas delivered to residential consumers for each state and year using the company-level data available in the EIA's Natural Gas Annual Respondent Query System. For these estimates, we drop firms reporting heat content values of 0 or above 2,500 BTU per cubic feet, and weight the reported heat content for each firm by the volume of gas supplied to residential consumers.[17] We also apply two manual edits. First, five state-year observations are missing residential consumer heat content data for all firms; we use the all-consumers heat content for these five observations. Second, the dominant firm in Arkansas is missing heat content data for 2001; we use the average of its report in 2000 and 2002 instead.

Lastly, to aid interpretation of monetary units, we deflate all prices in this paper—including the prices of natural gas and of electricity—to 2016 prices using the Bureau of Labor Statistics's (BLS) Consumer Price Index (CPI-U).

B.3 Home energy bills

For data on energy bills, we use Census 2000 5-Percent Public Use Microdata Sample (PUMS) files combined with the 2005 to 2010 American Community Survey (ACS) PUMS files. The Census/ACS data are available on an annual basis, and the finest geographic identifier is the Public Use Microdata Area (PUMA). We aggregate the microdata to obtain mean monthly energy bill for each PUMA for the year 2000 and 2005-2010. The relevant question in the Census 2000 is "What are the annual costs of utilities and fuels for this house, apartment, or mobile home?", broken down into different types of utilities and fuels.

[16]Response to the survey is required by law, and hence such non-response should not be a large problem.

[17]Heat contents typically range between 900 and 1,200 BTU per cubic feet.

In the ACS, households are asked how much these bills cost them last month (for electricity and gas) and last 12 months (for other fuels). We exclude households whose energy bills are included in their rent or condominium fees.

B.4 Home heating sources (*ShareGas*)

Our identification strategy uses cross-sectional variation in heating sources across geographic areas, represented by the variable *ShareGas*, the proportion of occupied housing units in each state, county, or PUMA in 2000 that indicates that gas is their main heating source. Gas refers to both utility gas from underground pipes serving the neighborhoods and bottled, tank or LP gas. The data source for *ShareGas* at the state and county level is the 2000 Decennial Census of Population and Housing Summary Files, which provide aggregate data at each of the aforementioned geographic levels.[18] The relevant Census question from which *ShareGas* is derived is "Which fuel is used most for heating this house, apartment, or mobile home?" We account for substantial county boundary changes by aggregating counties to a larger stable unit, as described in the mortality data section. *ShareGas* of the larger stable unit is the weighted average of *ShareGas* of each county that makes up the unit, with the weights being the county's share of total population in 2000 of the larger unit.[19]

The data source for *ShareGas* at the PUMA level is the Census 2000 5-Percent PUMS files. Our sample consists of occupied housing units and excludes group quarters and vacant units. We aggregate information at the household level to obtain the proportion of housing units in each PUMA that indicates that gas is their main heating source. We account for the change in PUMA definition due to Hurricane Katrina by aggregating three PUMAs in Louisiana into one stable unit across our sample period[20].

B.5 Relative price of natural gas to electricity (*RelPrice*)

Another key variable in our analysis is *log(RelPrice)*, the log of the relative price of natural gas to electricity in the US. We use the same data described in Section B.2 for this. The electricity price for the denominator is the national counterpart of that described previously: the national monthly price of electricity supplied to residential consumers.

One candidate for the natural gas price is the national monthly price of natural gas delivered to residential consumers, which is computed by dividing the reported revenue of local distribution utilities by the associated sales volume. The relevant survey question that the EIA uses defines revenue as "gross revenues including any and all demand charges, commodity charges, taxes, surcharges, adjustments or other charges billed for gas delivered"; consequently, fixed charges that utilities frequently include (e.g. basic monthly customer charges that do not depend on volumes) are included. However, we expect consumers to respond to the variable (i.e., usage-dependent) component of prices, not the fixed charge component. In the data, since the fixed charges are averaged over a smaller volume in summer, the residential price spikes in summer (Appendix Figure B1).

Because of this, we use the national monthly price of natural gas at the citygate instead. The citygate price is the price faced by local distribution utilities (companies that sell gas

[18] Table H40 of Summary Files 3. Available at American FactFinder (`https://factfinder.census.gov`)

[19] County population data come from the 2000 Census of Population and Housing Summary File 1.

[20] The Census Bureau merged 3 PUMAs (Louisiana 01801, 01802, and 01905) into 1 PUMA (77777) for 2006 and later

to residential consumers); hence it captures variation due to natural gas prices and excludes fixed charges to residents. In addition, utilities are required by federal law to price gas on a cost-recovery basis.[21] This means that absent forecast errors by the utilities, citygate prices should capture the variation in gas price perfectly. With forecast errors, utilities are legally required to return unexpected profits or losses made on natural gas to consumers by adjusting the future months' prices downwards or upwards. Citygate prices are not available for electricity.

When looking at annual outcomes in the Census/ACS data, we use the annual versions of the price variables. These are based on a separate survey of the universe of firms in the US, but are otherwise identical to the monthly versions.

B.6 Heating degree-days

To compute the number of heating degree-days (HDD), we use daily gridded temperature data for the contiguous US (4 kilometers by 4 kilometers resolution) from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) data developed and maintained by the PRISM Climate Group at Oregon State University (PRISM Climate Group 2004).[22] The PRISM data incorporate the current knowledge of U.S. spatial climate patterns, including elevation and prevailing wind patterns, and are the official spatial climate datasets of the U.S. Department of Agriculture (Daly et al. 2008).

To obtain HDD for each county-month, we first compute the geographic average daily mean temperature of each Census 2000 block group. For each block group, we take a simple average of all grid points within, or on the boundary of, the block group. We then compute HDD for each month for each block group, based on

$$HDD_{it} = \sum_{x=1}^{T(t)} \max\{threshold - tmean_{ix}, 0\} \tag{3}$$

where HDD_{it} is the HDD of block group i in month t, *threshold* is a temperature threshold (set at 65°F, following convention), $tmean_{ix}$ is the mean temperature of block group i on day x, and $T(t)$ is the number of days in month t. Next, we compute each county's HDD for the month by taking the average of the block groups within the county, weighted by the population in Census 2000. Finally, we scale HDD to a 30-day month, and divide by 1,000, to yield an average monthly measure of coldness. Block group geographic and population data come from the National Historical Geographic Information System (NHGIS; Manson et al. 2017).

B.7 Median household income and population share age 70+

Data for county and state median household income and fraction of people age 70 and above are from the 2000 Decennial Census of Population and Housing Summary Files. Data at the PUMA level are from the Census 2000 5-Percent PUMS files. Both variables are derived from the Census using the same approach as described above for *ShareGas*.

[21] Note that they may still charge a markup on distribution of gas, which is more difficult for the state to monitor.

[22] The specific dataset used is version D1 of the AN81d dataset, retrieved February 2017, from `http://prism.oregonstate.edu`.

B.8 House price index

State house price index used in the paper is the quarterly seasonally-adjusted purchase-only house price index, available from the Federal Housing Finance Agency (FHFA).

B.9 Unemployment rate

We use the Bureau of Labor Statistics' county-month unemployment rate as a control variable. A few county-level observations are missing due to Hurricane Katrina; we use the state unemployment rate for these observations, and include a dummy for affected observations in regressions. When analyzing the Census/ACS, we compute PUMA unemployment rate directly from the microdata.

B.10 Manufacturing share of the economy

We use the Bureau of Economic Analysis's state-quarter personal income data when controlling for manufacturing share of total employee compensation (meant to proxy for share of the economy). A few observations (fewer than 0.5%) are missing; we impute these observations by interpolation. Quarterly data are then matched to the appropriate time period.

B.11 Air pollution data

The data source for air pollution is daily station-level data from the US Environmental Protection Agency (EPA) Air Quality System (AQS).[23] The AQS data contain daily air quality indices (AQIs) for carbon monoxide, nitrogen dioxide, ozone, particulate matter (2.5 and 10), and sulfur dioxide; data for some pollutants for some stations are missing. We construct monthly AQIs for each geographic unit of analysis, and then aggregate to the appropriate time period.

To construct monthly AQIs for each county in our sample for analysis of mortality data, we use a mix of procedures. For the first procedure, we compute each pollutant's AQI at the Census 2000 block group level, and then aggregate to the county level, weighting by population. We compute the AQI for each block group as the average of all AQI measurements taken within a month at all stations within 100 kilometers, weighted by the inverse of the squared distance to the station. County AQI is then the population-weighted average of block group AQIs. Block group and population data come from the NHGIS.

The above procedure (setting a distance threshold and computing the AQI) is standard in the literature using stations data, but it produces many missing observations. We patch missing data using a second procedure. Specifically, if a county has more than 50 percent of its population not assigned a pollutant AQI value in any month, we use a second procedure to compute its AQI values for all months. For these counties, for each month, we compute AQI based on the nearest five stations with available measurements, weighted by the inverse of the squared distance between the station and the county centroid. This guarantees that all counties have a pollution AQI measure for all months in consideration.

Appendix Table B2 shows, for each pollutant, the breakdown of the procedure used to compute AQI in the sample of counties used in our analysis of mortality data.

[23] The EPA provides several ways of accessing the data. We use the pre-generated data files, accessed February 2017.

We construct monthly AQIs at the PUMA or state level for use in analyzing the other outcome variables analogously. We then aggregate to the relevant time period for analysis. This differs depending on the outcome variable. When looking at the effects on home energy price, usage, and mortality, we aggregate to the monthly level directly. When looking at the effects on home energy bills, we additionally account for the time structure in the ACS. ACS documentation indicates that around half of data for each month are collected two months prior through computer-assisted personal interviewing, and around half are collected contemporaneously through mail or the internet. In addition, we account for possible lags in bill payment. For this, we assume that two-thirds of interviewees in a certain month report the bill paid in the previous month, and one-third report the bill paid two months ago (because the most recent bill might not have been paid yet). Hence, when aggregating monthly AQIs for a particular year when analyzing the Census/ACS, we include data from up to four previous months (September the previous year gets one-sixth of a month's worth of weight, etc.).

B.12 Independent variables used in heterogeneity analysis

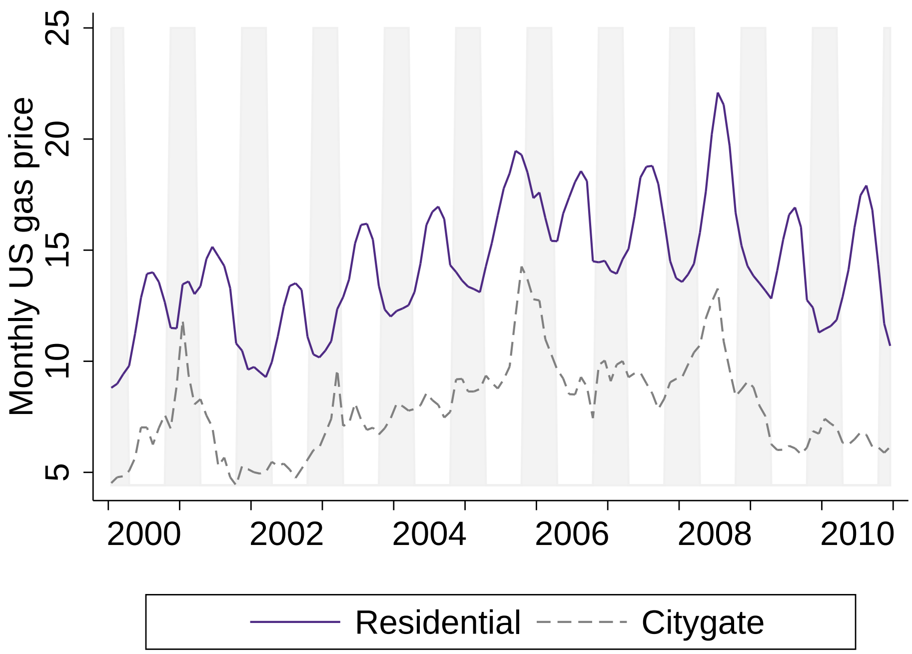
Variables used in analysis of heterogeneity in the effect of heating price on mortality are from the following sources:

- **Poverty rate:** Data on proportion of households in the county with income below 150% of the poverty level is from the 2000 Decennial Census of Population and Housing Summary Files.
- **Education:** Data on the deceased's education level is provided in the mortality files. We drop deaths that occur before the age of 25, with censored education level, for this analysis. To compute age-adjusted mortality rates by education level, we use Census/ACS population data, since the SEER data does not contain a breakdown by education level. We interpolate proportions for the years in which no population data exist (2001 to 2004).
- **Sex:** Data on the deceased's sex is provided in the mortality files.

B.13 LIHEAP data

As a robustness check, we use data on Low Income Home Energy Assistance Program (LIHEAP) spending from the U.S. Department of Health and Human Services' LIHEAP Data Warehouse. The data are based on mandatory reports from states for each fiscal year, and are available at the state-fiscal year level starting in fiscal year 2001 (i.e. since October 2000). For the nine months in our sample without LIHEAP data, we impute an arbitrary value for LIHEAP per capita and include a dummy for affected observations in the regressions.

Appendix Figure B1: National price of natural gas over time



Notes: Price in dollars per million BTU. Gray regions are winter months (November–March).

Appendix Table B1: Data sources

Data	Data source	Geographic identifier	Temporal identifier
Dependent variables			
Mortality rate	Vital Statistics Mortality Files	County	Month
Average home energy price	Energy Information Administration (EIA)	State	Month
Home energy usage	Energy Information Administration	State	Month
Home energy bill	Census; American Community Survey (ACS)	PUMA	Year
Independent variables			
Home heating energy type	Census	Census tract	Year
Energy prices	Energy Information Administration	State	Month
Temperature	PRISM	Grid point[a]	Day
Median household income	Census	Census tract	Year
Fraction of people aged 70 & above	Census	Census tract	Year
House price index	Federal Housing Finance Agency	State	Quarter
Air pollution	Environ. Protection Agency Air Quality System	Pollution monitor	Day
Unemployment rate	Bureau of Labor Statistics	County	Month
Manufacturing share of economy	Bureau of Economic Analysis	State	Quarter
LIHEAP assistance funds	Department of Health and Human Services	State	Fiscal year

[a] 4 km by 4 km resolution

Appendix Table B2: Frequency of the two interpolation procedures used for calculating AQIs

	CO	NO_2	O_3	$PM_{2.5}$	PM_{10}	SO_2
Based on distance threshold	1,177	1,048	1,096	2,231	1,762	1,512
Based on nearest 5 stations	1,616	1,745	1,697	562	1,031	1,281
Total counties	2,793	2,793	2,793	2,793	2,793	2,793