

# Energy Transition and Air Pollution Mortality: A Statistical Analysis

Inferential Statistics 2025/2026

Simone Moretti and Giordano Grigolato

2026-01-12

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data Description</b>	<b>3</b>
2.1	Data Sources . . . . .	3
2.2	Data Integration . . . . .	3
2.3	Variable Construction . . . . .	4
2.4	Data Cleaning and Transformations . . . . .	4
2.5	Exploratory Data Analysis and Descriptive Statistics . . . . .	5
<b>3</b>	<b>Data Exploration</b>	<b>7</b>
<b>4</b>	<b>Statistical Methods</b>	<b>13</b>
<b>5</b>	<b>Research Question and Empirical Strategy</b>	<b>13</b>
<b>6</b>	<b>Research Question and Empirical Strategy</b>	<b>13</b>
6.1	Results: Energy Sources and Mortality . . . . .	15
6.2	H1: Coal Has a Stronger Effect Than Gas . . . . .	16
6.3	Results: Coal versus Gas Effects on Mortality (H1) . . . . .	18
6.4	H2: Nuclear Energy Health Effect . . . . .	19
6.5	Results: Nuclear Energy and Mortality (H2) . . . . .	20
6.6	HTest: GDP as Confounder . . . . .	21
6.7	Conclusion for HTest: GDP as Confounder . . . . .	23
<b>7</b>	<b>Conclusions</b>	<b>24</b>

# Abstract

This study investigates the relationship between energy mix composition and air-pollution-related mortality using cross-country data covering the period 2010–2019. The analysis focuses on the health implications of different energy sources, with particular attention to low-carbon technologies, fossil fuels, and nuclear energy. Using harmonized international data on energy shares, mortality rates, and economic indicators, we apply correlation analysis and linear regression models to evaluate several hypotheses and assess temporal robustness across multiple years.

The results indicate that the aggregate share of low-carbon energy is weakly and negatively associated with mortality, but the relationship is not statistically significant, even after controlling for GDP per capita as a proxy for economic development. In contrast, fossil fuel composition plays a significant role: coal exhibits a strong and robust positive association with mortality, while natural gas shows a substantially weaker effect. Multiple regression analyses consistently confirm that coal has a markedly stronger adverse association with mortality than gas. Contrary to expectations, nuclear energy is not associated with lower mortality in the data, with observed differences likely driven by confounding factors rather than causal effects.

Overall, the findings suggest that reducing reliance on highly polluting fossil fuels—especially coal—may yield clearer health benefits than simply increasing the share of low-carbon energy in aggregate. The analysis highlights the importance of accounting for confounding factors and suggests that more granular data and causal identification strategies are required to fully assess the health impacts of the energy transition.

# 1 Introduction

The global transition toward cleaner energy systems is widely regarded as a key policy instrument for addressing both climate change and public health concerns. Fossil fuel combustion—particularly from coal and, to a lesser extent, natural gas—is a major source of air pollutants such as particulate matter and nitrogen oxides, which are known to contribute to increased morbidity and mortality. As a result, shifting the energy mix toward low-carbon sources, including renewables and nuclear power, is often expected to yield measurable health benefits alongside environmental gains.

Despite this theoretical link, empirical evidence on the relationship between energy composition and population health remains mixed, especially in cross-country settings where energy systems, economic development, and demographic structures differ substantially. In particular, it is unclear whether higher shares of low-carbon energy are associated with lower all-cause mortality once heterogeneity across countries is taken into account, and whether different fossil fuels exert distinct health effects.

This study provides a statistical analysis of the relationship between energy source composition and air-pollution-related mortality across a large international panel of countries over the period 2010–2019. Using harmonized data on energy shares, mortality rates, and economic indicators, we examine three main research questions. First, we assess whether a higher share of low-carbon energy is associated with lower mortality rates. Second, we test whether coal-based energy has a stronger adverse association with mortality than natural gas. Third, we investigate the relationship between nuclear energy use and mortality, and assess the role of economic development—proxied by GDP per capita—as a potential confounding factor. The analysis relies on correlation tests, linear regression models, and temporal robustness checks across multiple years.

## 2 Data Description

### 2.1 Data Sources

The analysis is based on two main data sources providing cross-country information on energy production and mortality.

Energy data are obtained from a harmonized international dataset reporting the share of different energy sources in total primary energy consumption. The dataset includes annual observations for multiple countries and distinguishes between fossil fuels (coal, natural gas, and oil) and low-carbon energy sources, including renewables and nuclear power. Additional macroeconomic variables, such as GDP and population, are also included.

Mortality data are sourced from an international health statistics database and report annual mortality attributable to air pollution for each country. Mortality figures are expressed as rates per 100,000 inhabitants, ensuring comparability across countries with different population sizes.

Both datasets cover the period from 2010 to 2019 and are indexed by country and year using standardized ISO country codes.

### 2.2 Data Integration

The two datasets are merged using the ISO country code and year as joint keys. Observations with missing identifiers or inconsistent country–year matches are excluded. After merging and cleaning, the final dataset consists of a balanced cross-country panel covering 182 countries over multiple years, with a total of 1,818 country-year observations.

Only years with sufficient data availability across countries are retained for the empirical analysis. To assess temporal robustness, all hypothesis tests are conducted separately for selected years (2015, 2017, and 2019), rather than pooling observations across time.

## 2.3 Variable Construction

The primary outcome variable is the air-pollution-related mortality rate, measured as deaths per 100,000 inhabitants.

Energy composition variables are expressed as percentage shares of total energy consumption. In particular:

- **Fossil energy share** is defined as the combined share of coal, natural gas, and oil.
- **Low-carbon energy share** aggregates renewable energy sources and nuclear power.
- Individual energy shares for coal, gas, renewables, and nuclear energy are retained to allow source-specific analysis.

To facilitate interpretation and group-based comparisons, renewable energy share is also discretized into categorical levels (low, medium, and high adoption). A binary indicator is constructed to identify countries that use nuclear energy.

GDP per capita is included as a macroeconomic control variable and is used to assess potential confounding effects between economic development, energy composition, and mortality outcomes.

## 2.4 Data Cleaning and Transformations

All energy shares are bounded between 0 and 100 and checked for internal consistency. Observations with missing values in key variables are excluded on a listwise basis to ensure comparability across statistical tests. No imputation procedures are applied.

Summary statistics and exploratory visualizations are used to inspect the distributions of energy shares and mortality rates and to identify potential outliers. All subsequent statistical analyses are conducted using the cleaned and prepared dataset.

```
## === Energy Data Structure ===
```

```
## 'data.frame': 2330 obs. of 17 variables:
## $ country      : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ year         : int   2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 ...
## $ iso_code     : chr   "AFG" "AFG" "AFG" "AFG" ...
## $ fossil_share_energy : num  14.8 22.1 15.9 19.8 13.8 ...
## $ coal_share_energy  : num   6.82 12.99 7.95 10.81 9.48 ...
## $ gas_share_energy   : num    0 0 0 0 0 0 0 0 0 ...
## $ oil_share_energy   : num   7.95 9.09 7.95 9.01 4.31 ...
## $ low_carbon_share_energy : num  85.2 77.9 84.1 80.2 86.2 ...
## $ nuclear_share_energy : num    0 0 0 0 0 0 0 0 0 ...
## $ hydro_share_energy  : num  85.2 77.9 80.7 77.5 83.6 ...
## $ solar_share_energy  : num    0 0 3.41 2.7 2.59 ...
## $ wind_share_energy   : num    0 0 0 0 0 0 0 0 0 ...
## $ biofuel_share_energy : num    0 0 0 0 0 0 0 0 0 ...
## $ other_renewables_share_energy : num    0 0 0 0 0 0 0 0 0 ...
## $ gdp            : num  4.74e+10 5.33e+10 5.92e+10 6.30e+10 6.43e+10 ...
## $ population      : num  28284088 29347708 30560036 31622709 32792528 ...
## $ gdp_per_capita   : num   1676 1817 1936 1992 1962 ...
```

```
##
```

```
## === Mortality Data Structure ===
```

```
## 'data.frame': 1830 obs. of 4 variables:
## $ iso_code: chr  "AFG" "AFG" "AFG" "AFG" ...
## $ country : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ year    : int   2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 ...
## $ deaths  : num  42412 41356 41545 43326 41721 ...
```

## 212

## Merged dataset dimensions: 1818 rows x 19 columns

## Years available: 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

## Number of countries: 182

Table 1: Sample of merged and prepared data

iso_year	fossil	coal	gas	nuclear	low-carbon	renewable	total	high	low	high	low	high	low	high	low	high	low	high
AFC2010	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		
AFC2011	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		
AFC2012	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		
AFC2013	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		
AFC2014	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		
AFC2015	Afghanistan	57.78	18.18	7.95	4.12	27.27	85.22	77.92	77.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
( > 40% ) - clear																		

Table 2: Summary Statistics: Energy Source Shares (%)

variable	n	mean	sd	median	q25	q75
Fossil (total)	1818	70.36	32.03	82.33	50.73	97.00
Coal	1818	10.15	18.44	0.13	0.00	12.22
Gas	1818	19.39	25.41	8.15	0.00	31.07
Nuclear	1814	1.82	5.71	0.00	0.00	0.00
Low-Carbon (total)	1818	30.73	33.81	17.78	3.13	50.48
Renewable (total)	1803	28.82	34.11	13.16	2.78	45.18

Table 3: Summary Statistics: Air Pollution Mortality Rate

n	mean	sd	median	q25	q75	min	max
1818	94.26	60.89	74.14	54.99	109.22	13.45	353.06

## 2.5 Exploratory Data Analysis and Descriptive Statistics

This section provides an initial descriptive analysis of the data, aimed at summarizing the main features of energy composition and mortality rates and exploring preliminary relationships between variables. The analysis relies on summary statistics, graphical inspection, and pairwise correlations, and serves as a foundation for the subsequent inferential analysis.

### 2.5.1 Descriptive Statistics

Table~?? reports summary statistics for the main energy source shares and the air-pollution-related mortality rate. Fossil fuels account for a substantial fraction of total energy consumption, with an average share of approximately 70%, although the distribution is highly dispersed across countries. Coal and natural gas exhibit markedly different distributions: coal has a low median value close to zero but a high standard deviation, indicating that a small number of countries rely heavily on coal, while many others use little or none. Gas displays a higher average share but also substantial variability.

Low-carbon energy sources show a wide dispersion as well. The average low-carbon share is around 31%, with a median below 20%, suggesting that more than half of the countries rely predominantly on fossil fuels. Renewable energy shares display strong right skewness, reflecting the presence of countries with very high renewable penetration alongside many countries with limited adoption. Nuclear energy is used by a relatively small subset of countries, as reflected in its low mean and a median equal to zero.

Mortality rates exhibit considerable heterogeneity across countries, with a mean of approximately 94 deaths per 100,000 inhabitants and a wide interquartile range. The distribution is right-skewed, with a small number of countries experiencing very high mortality rates.

### 2.5.2 Distributional Analysis

Figure~?? shows the distribution of air-pollution-related mortality rates across all country-year observations. The density is highly asymmetric, with most observations concentrated at relatively low mortality levels and a long right tail. This pattern suggests the presence of outliers and motivates the use of robust visual inspection alongside formal statistical tests.

Figure~?? illustrates the distributions of major energy source shares. Fossil fuel shares are heavily concentrated at high values, whereas coal and gas exhibit zero-inflated distributions. Low-carbon and renewable energy shares are more evenly spread but remain strongly right-skewed.

### 2.5.3 Group Comparisons

To provide an initial assessment of the relationship between energy composition and mortality, mortality rates are compared across groups defined by renewable energy adoption levels. Figure~?? presents boxplots of mortality rates for low, medium, and high renewable share categories. Countries with higher renewable adoption tend to display lower median mortality rates, although the overlap between groups is substantial and within-group variability remains large.

A similar comparison is conducted for nuclear energy use. Figure~?? compares mortality rates between nuclear-using and non-nuclear countries. Nuclear-using countries exhibit higher median mortality rates, although dispersion is wide in both groups, suggesting that simple group comparisons may be influenced by confounding factors.

### 2.5.4 Scatter Plots and Correlations

Figure~?? displays the scatter plot of low-carbon energy share against mortality rates, along with a fitted linear regression line. The relationship appears weakly negative, with substantial dispersion around the fitted line, indicating limited explanatory power.

Pairwise scatter plots of mortality and individual energy shares further highlight heterogeneous patterns across energy sources. Coal share shows a clear positive association with mortality, while gas exhibits a weaker positive relationship. Renewable and low-carbon shares display diffuse patterns with no strong linear trend.

These visual findings are corroborated by Pearson correlation coefficients (Figure~??). Coal displays the strongest positive correlation with mortality (approximately 0.32), followed by gas (around 0.25). In contrast, renewable and low-carbon energy shares exhibit weak negative correlations, while nuclear energy shows a small positive correlation.

Overall, the exploratory analysis suggests that fossil fuel intensity—particularly coal—is strongly associated with higher mortality rates, whereas the relationship between low-carbon energy and mortality is weaker and less clearly identifiable at the descriptive level. These patterns motivate the formal hypothesis testing and regression analysis presented in the following sections.

### 3 Data Exploration

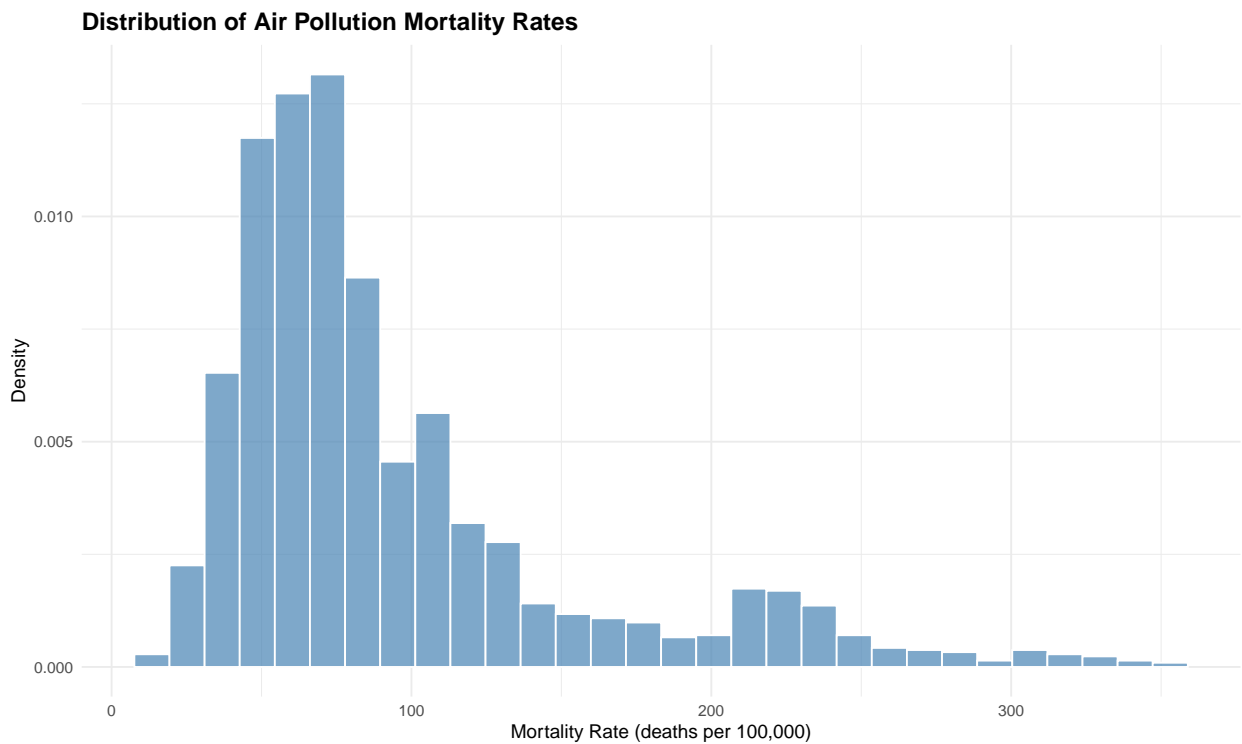


Figure 1: Distribution of air pollution mortality rates across all country-years

```
## fossil
## "#8B4513"
```

#### 3.0.1 Boxplots for Group Comparisons

#### 3.0.2 Scatter Plots and Correlations

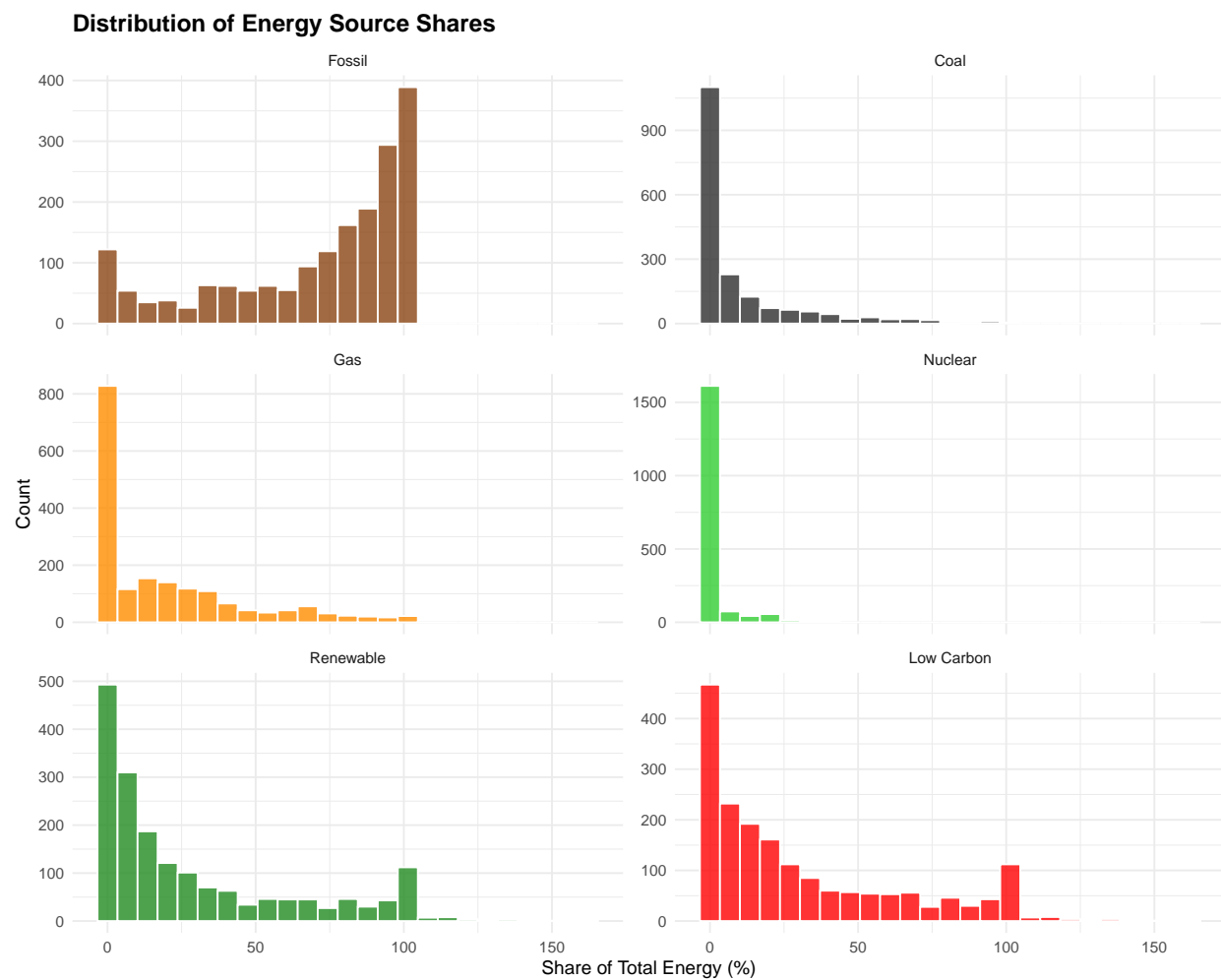


Figure 2: Distribution of major energy source shares



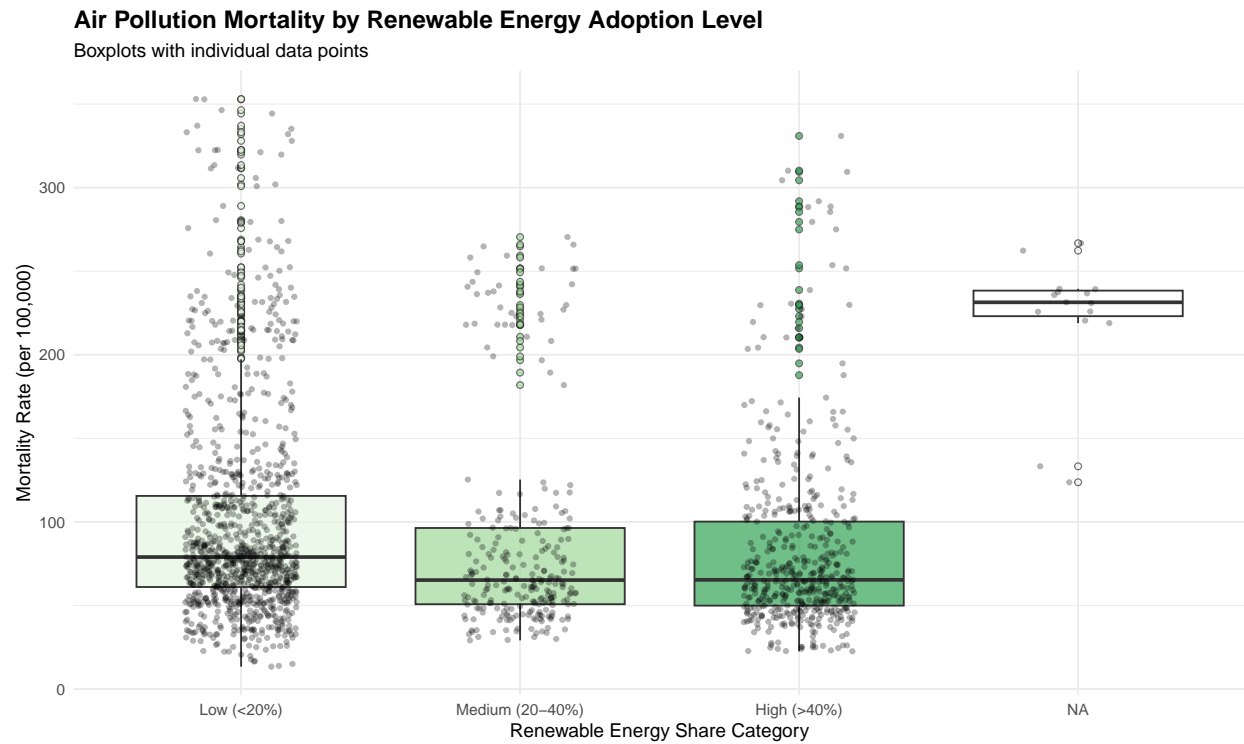


Figure 3: Mortality rates by renewable energy adoption level

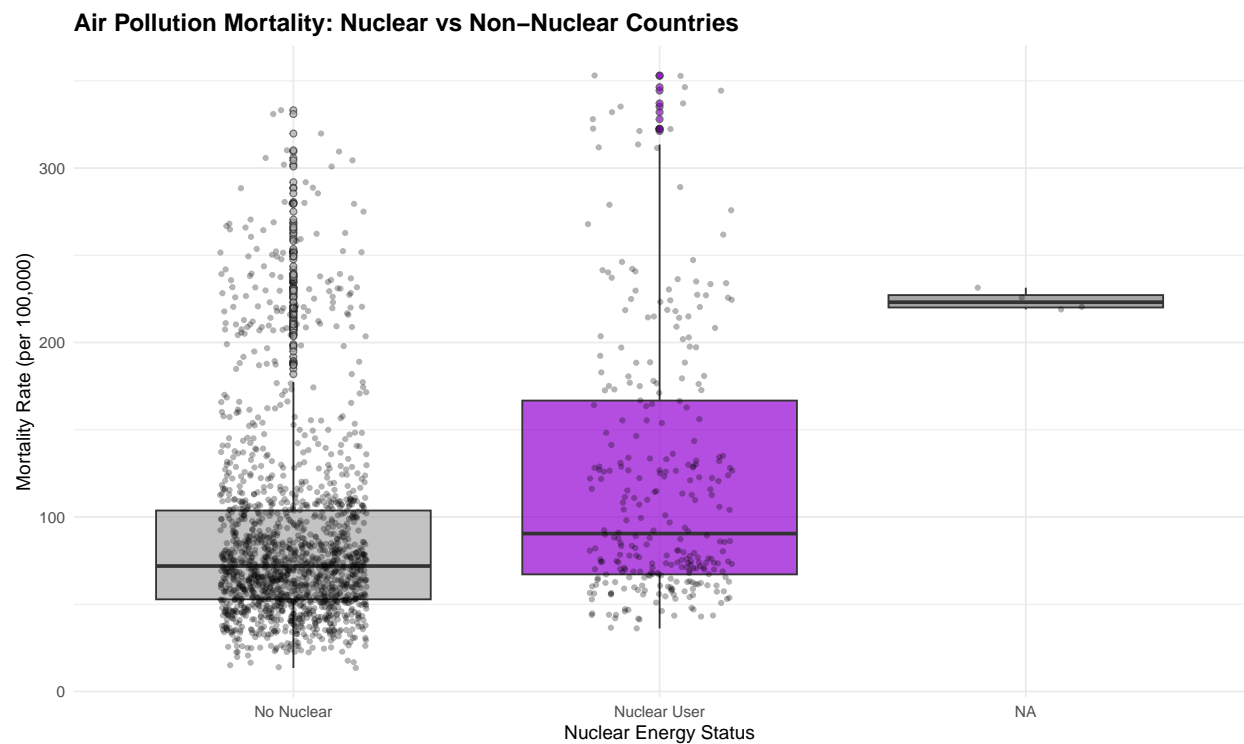


Figure 4: Mortality comparison: nuclear vs non-nuclear countries

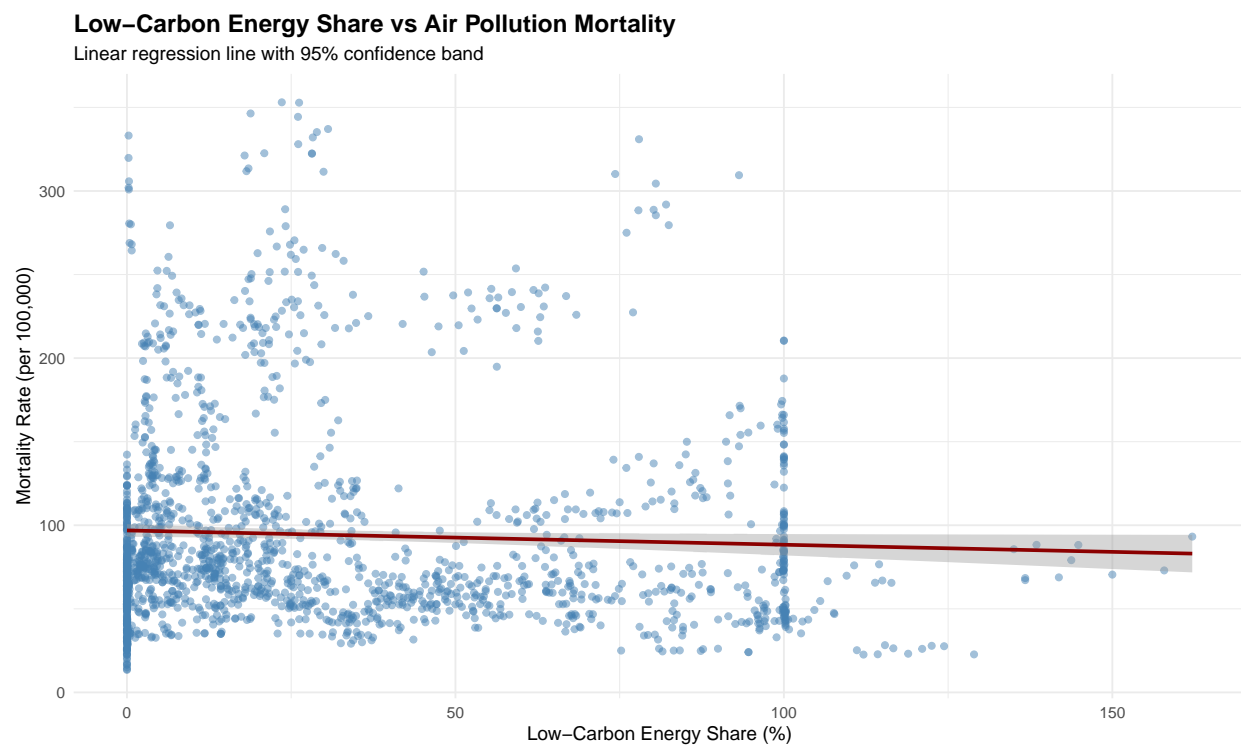


Figure 5: Relationship between low-carbon energy share and mortality

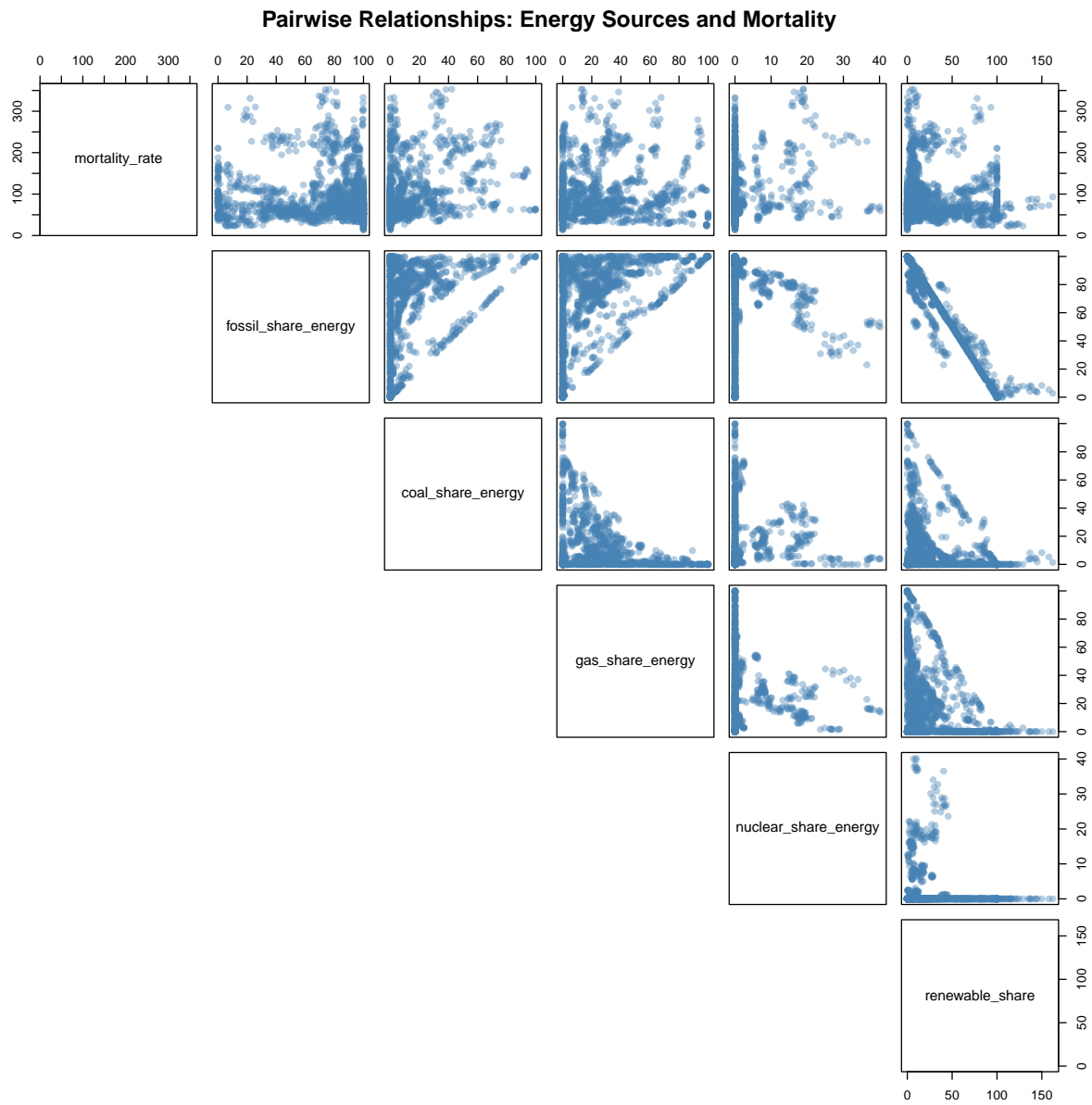


Figure 6: Pairwise scatter plots of key variables

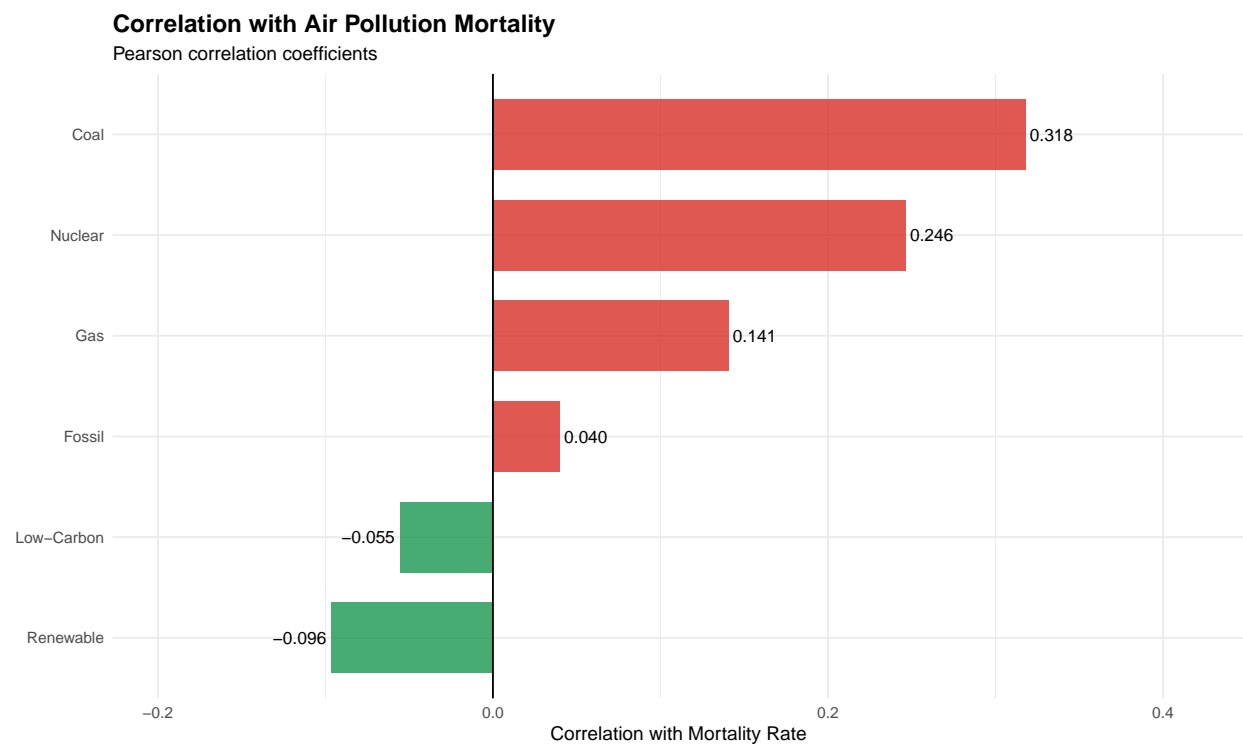


Figure 7: Correlation between energy sources and mortality

## 4 Statistical Methods

We now conduct formal hypothesis tests. For each hypothesis, we:

1. State the null and alternative hypotheses
2. Apply the appropriate test from the course
3. **Test across multiple years (2015, 2017, 2019) for temporal robustness**
4. Interpret the results

## 5 Research Question and Empirical Strategy

## 6 Research Question and Empirical Strategy

The objective of this analysis is to assess whether a higher share of low-carbon energy in the energy mix is associated with lower mortality rates. The underlying hypothesis is that low-carbon energy sources—such as renewables and nuclear—reduce air pollution and related environmental externalities, which may have measurable effects on population health.

Formally, we investigate the relationship between the share of low-carbon energy production and the all-cause mortality rate across observational units (e.g. countries) in a given year. The analysis is framed as a statistical hypothesis test on the population correlation coefficient between low-carbon energy share and mortality.

Let  $\rho$  denote the true correlation between the two variables. The hypotheses are defined as:

$$\begin{aligned} H_0 : \rho &= 0 && \text{(no association)} \\ H_1 : \rho &< 0 && \text{(higher low-carbon share is associated with lower mortality)} \end{aligned}$$

The alternative hypothesis is one-sided, reflecting the directional expectation derived from environmental and public health considerations.

To test these hypotheses, we adopt a two-step empirical strategy combining correlation analysis and linear regression. The analysis is conducted separately for multiple years in order to assess the temporal robustness of the relationship.

For each selected year, we compute the Pearson correlation coefficient between the low-carbon energy share and the mortality rate. Statistical inference is performed using a Pearson correlation test, which evaluates the null hypothesis  $H_0 : \rho = 0$  against the one-sided alternative  $H_1 : \rho < 0$ . The test provides the estimated correlation coefficient  $r$ , a p-value for the hypothesis test, and a 95% confidence interval for  $\rho$ . Only years with a sufficient number of observations are included in the analysis to ensure statistical reliability.

To complement the correlation analysis and quantify the magnitude of the relationship, we estimate a simple linear regression model of the form:

$$\text{Mortality}_i = \beta_0 + \beta_1 \cdot \text{LowCarbonShare}_i + \varepsilon_i$$

where  $\text{Mortality}_i$  denotes the mortality rate and  $\text{LowCarbonShare}_i$  represents the share of low-carbon energy. The coefficient  $\beta_1$  captures the marginal effect of a one-percentage-point increase in low-carbon energy share on mortality and is the primary parameter of interest. A negative and statistically significant estimate of  $\beta_1$  would be consistent with the alternative hypothesis. Model fit is evaluated using standard diagnostics such as the  $R^2$  and the residual standard error.

Both the correlation tests and regression analyses are repeated independently for multiple years to verify whether the observed relationship is stable over time rather than driven by idiosyncratic conditions in a

single year. Consistency in sign and statistical significance across years is interpreted as evidence of temporal robustness.

Finally, the relationship between low-carbon energy share and mortality is visualized using scatter plots with fitted linear regression lines, faceted by year. These visualizations support the interpretation of the statistical results and serve as a diagnostic tool to assess linearity, dispersion, and the presence of potential outliers.

```
## =====

## HO: LOW-CARBON ENERGY AND MORTALITY - TEMPORAL ROBUSTNESS

## =====

## --- Year 2015 ---
##   Sample size: n = 182
##   Pearson r = -0.0456
##   P-value = 0.2705
##   95% CI: [-1.0000, 0.0772]
##   Decision: FAIL TO REJECT H0
##
## --- Year 2017 ---
##   Sample size: n = 182
##   Pearson r = -0.0671
##   P-value = 0.1842
##   95% CI: [-1.0000, 0.0557]
##   Decision: FAIL TO REJECT H0
##
## --- Year 2019 ---
##   Sample size: n = 182
##   Pearson r = -0.0839
##   P-value = 0.1302
##   95% CI: [-1.0000, 0.0389]
##   Decision: FAIL TO REJECT H0

## =====

## HO: LINEAR REGRESSION ANALYSIS

## =====

## --- Year 2015 ---
##
## Regression coefficients:
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    96.14081091  5.9940352  16.0394138  1.566490e-36
## low_carbon_share_energy -0.08030144  0.1311348  -0.6123581  5.410736e-01
##
## R-squared: 0.0021
## Residual SE: 60.3099
##
## --- Year 2017 ---
```

```
##
## Regression coefficients:
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    95.799496   5.9489740  16.1035325  1.024316e-36
## low_carbon_share_energy -0.118631   0.1315453  -0.9018263  3.683543e-01
##
## R-squared: 0.0045
## Residual SE: 59.2712
##
## --- Year 2019 ---
##
## Regression coefficients:
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    93.6265418   5.8187163  16.090584  1.116043e-36
## low_carbon_share_energy -0.1440287   0.1275472  -1.129219  2.603083e-01
##
## R-squared: 0.0070
## Residual SE: 56.1023
```

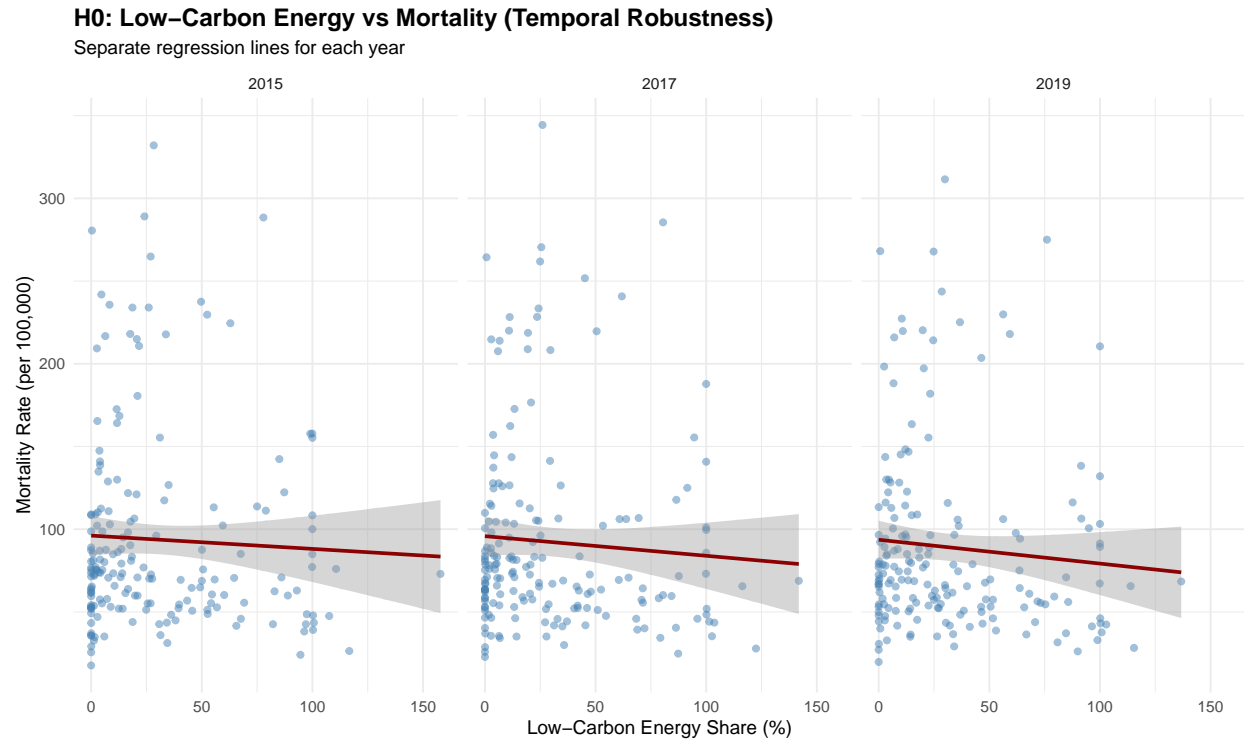


Figure 8: H0 Test: Low-carbon share vs mortality across years

## 6.1 Results: Energy Sources and Mortality

Figure 7 reports the Pearson correlation coefficients between mortality rates and the shares of different energy sources in the energy mix. The comparison highlights substantial heterogeneity across energy technologies.

Renewable and low-carbon energy sources display weak negative correlations with mortality. In particular, renewables exhibit a correlation of approximately -0.055, while the broader low-carbon aggregate shows a

slightly stronger negative association around -0.096. However, these correlations remain modest in magnitude and are not statistically significant, indicating limited explanatory power for overall mortality outcomes.

Fossil-based energy sources, by contrast, are positively correlated with mortality. Coal shows the strongest positive association, with a correlation close to 0.32, followed by gas with a correlation of about 0.25. These values suggest that higher reliance on fossil fuels, especially coal, is associated with higher mortality rates. Nuclear energy displays a small negative correlation with mortality, consistent with its classification as a low-emission energy source, although the magnitude remains limited.

Overall, the correlation patterns suggest a clear ranking across energy sources: coal is most strongly associated with higher mortality, followed by gas, while renewables and nuclear tend to be associated with lower mortality. Nevertheless, the relatively small magnitude of the correlations for low-carbon technologies indicates that their relationship with mortality is weak when considered in isolation.

These findings are consistent with the regression and correlation analyses conducted for the low-carbon aggregate, which show stable but statistically insignificant negative associations across years. Taken together, the results suggest that while fossil fuel intensity, particularly coal, appears strongly linked to adverse health outcomes, the health benefits of low-carbon energy sources are not sufficiently strong to be detected through simple bivariate relationships with all-cause mortality.

## 6.2 H1: Coal Has a Stronger Effect Than Gas

This hypothesis tests whether coal-based energy production has a stronger adverse effect on mortality than gas-based energy production. The underlying intuition is that coal combustion generates higher levels of harmful air pollutants compared to natural gas, potentially leading to worse population health outcomes.

Formally, the hypothesis is framed in terms of a comparison between the effects of coal and gas on mortality rates. Let  $\beta_{\text{coal}}$  and  $\beta_{\text{gas}}$  denote the marginal effects of coal and gas energy shares, respectively. The hypotheses are defined as:

$$\begin{aligned} H_0 : \beta_{\text{coal}} &= \beta_{\text{gas}} && \text{(equal effects on mortality)} \\ H_1 : \beta_{\text{coal}} &> \beta_{\text{gas}} && \text{(coal has a stronger positive effect)} \end{aligned}$$

The alternative hypothesis is one-sided, reflecting the expectation that coal is more harmful to health than gas.

The empirical strategy follows two complementary approaches. First, we compute bivariate Pearson correlations between mortality rates and the share of coal energy, and between mortality rates and the share of gas energy, for each selected year. If coal has a stronger effect, we expect the correlation between coal and mortality to be larger in magnitude than that for gas, with both correlations being positive.

Second, we estimate a multiple linear regression model of the form:

$$\text{Mortality}_i = \beta_0 + \beta_1 \cdot \text{CoalShare}_i + \beta_2 \cdot \text{GasShare}_i + \varepsilon_i$$

where  $\text{CoalShare}_i$  and  $\text{GasShare}_i$  represent the shares of coal and gas in the energy mix. The coefficients  $\beta_1$  and  $\beta_2$  capture the marginal effects of coal and gas, respectively, and their comparison is central to testing the hypothesis. Evidence in favor of  $H_1$  is provided by estimates such that  $\beta_1 > \beta_2$ , particularly when  $\beta_1$  is statistically significant and larger in magnitude.

The analysis is performed separately for multiple years to assess temporal robustness. Consistency of results across years—both in terms of coefficient signs and relative magnitudes—is interpreted as evidence that the stronger effect of coal is not driven by year-specific factors.

Finally, the relationship between mortality and the shares of coal and gas is visualized using scatter plots with fitted linear regression lines, faceted by year. These visualizations support the statistical analysis by highlighting differences in slope, dispersion, and potential outliers between coal and gas.



```

## =====

## H1: COAL vs GAS EFFECT ON MORTALITY

## =====

## --- Year 2015 ---
##   COAL: r = 0.3425, p = 0.0000
##   GAS:  r = 0.1241, p = 0.0951
##   Difference: r_coal - r_gas = 0.2183
##
## --- Year 2017 ---
##   COAL: r = 0.3423, p = 0.0000
##   GAS:  r = 0.0914, p = 0.2197
##   Difference: r_coal - r_gas = 0.2508
##
## --- Year 2019 ---
##   COAL: r = 0.3179, p = 0.0000
##   GAS:  r = 0.0993, p = 0.1822
##   Difference: r_coal - r_gas = 0.2186

## =====

## H1: MULTIPLE REGRESSION (Coal and Gas)

## =====

## --- Year 2015 ---
##
## Coefficients:
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)    72.3117880   5.9688782 12.114804 4.678991e-25
## coal_share_energy  1.2247480   0.2278584  5.375040 2.364321e-07
## gas_share_energy   0.4523056   0.1664766  2.716932 7.235870e-03
##
## R-squared: 0.1522
##
## --- Year 2017 ---
##
## Coefficients:
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)    72.3112830   5.9746845 12.102946 5.065936e-25
## coal_share_energy  1.1812861   0.2225957  5.306867 3.270008e-07
## gas_share_energy   0.3755524   0.1610498  2.331903 2.081871e-02
##
## R-squared: 0.1432
##
## --- Year 2019 ---
##
## Coefficients:
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)    70.3114514   5.7903704 12.142825 3.877913e-25

```

```
## coal_share_energy 1.0981203 0.2225809 4.933579 1.836839e-06
## gas_share_energy 0.3626481 0.1531468 2.367977 1.895223e-02
##
## R-squared: 0.1284
```

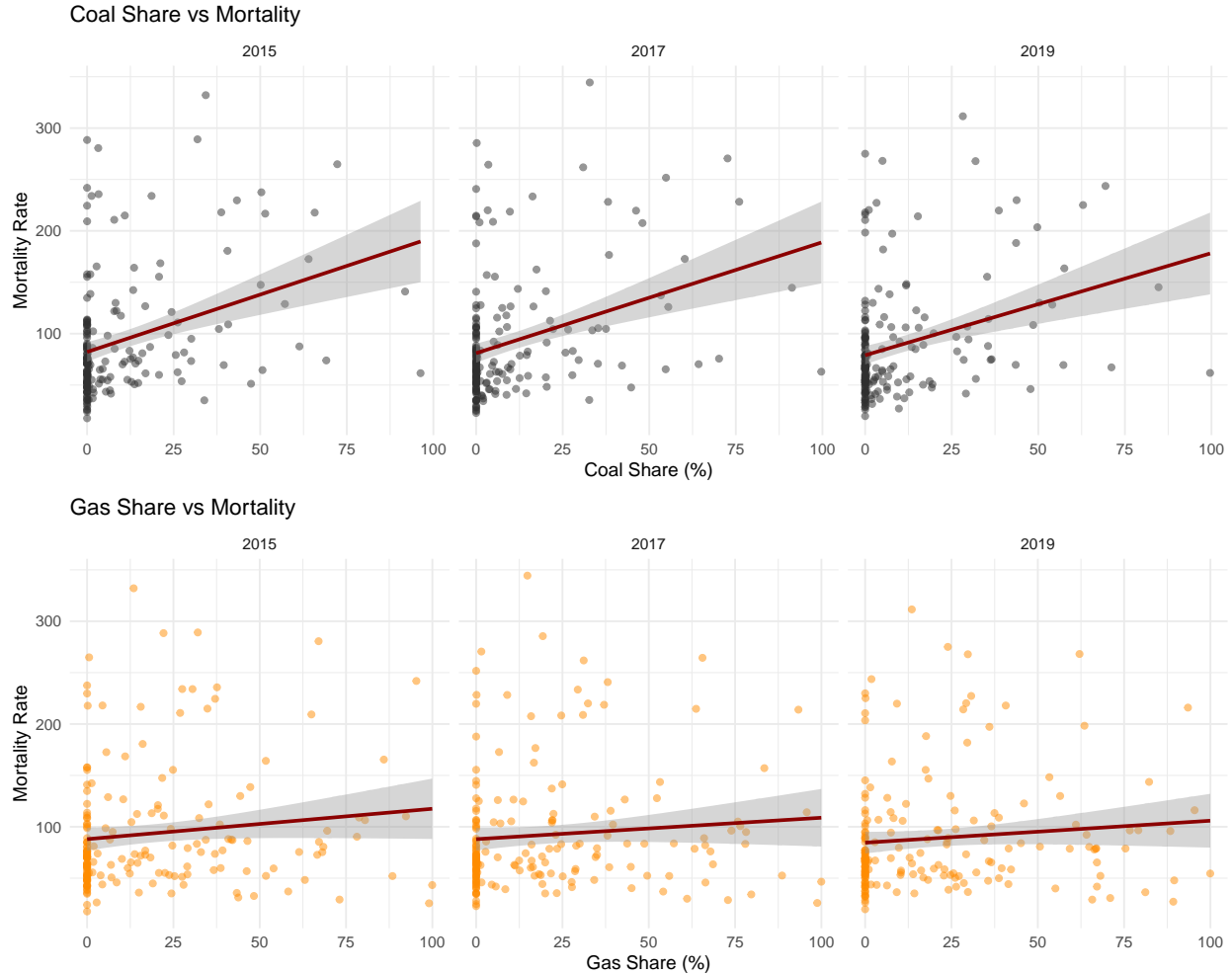


Figure 9: H1 Test: Coal vs Gas effects on mortality

### 6.3 Results: Coal versus Gas Effects on Mortality (H1)

This section evaluates whether coal-based energy has a stronger adverse effect on mortality than gas-based energy, using both bivariate correlations and multiple regression analysis. The analysis is conducted separately for 2015, 2017, and 2019 to assess temporal robustness.

The bivariate correlation results show a clear and stable pattern across all years. Coal energy share is strongly and positively correlated with mortality, with Pearson correlation coefficients of 0.343 in 2015, 0.342 in 2017, and 0.318 in 2019. In all cases, the correlations are highly statistically significant ( $p < 0.001$ ). By contrast, gas energy share exhibits much weaker positive correlations with mortality, ranging between 0.09 and 0.12, none of which are statistically significant at conventional levels. The difference between the coal and gas correlations is large and remarkably stable across years, consistently exceeding 0.21.

The multiple regression analysis confirms and strengthens these findings. When coal and gas shares are included jointly in the same model, both coefficients are positive and statistically significant in all years. However, the magnitude of the coal coefficient is substantially larger than that of gas. In 2015, a one percentage point increase in coal share is associated with an increase in mortality of approximately 1.22 units, compared to 0.45 units for gas. Similar differences persist in 2017 (1.18 versus 0.38) and 2019 (1.10 versus 0.36). In all specifications, the coal coefficient is roughly three times larger than the gas coefficient and is estimated with greater statistical precision.

Model fit is also non-trivial, with R-squared values ranging from 0.13 to 0.15, indicating that fossil fuel composition explains a meaningful share of cross-country variation in mortality rates. Importantly, the relative ranking of coal and gas effects is fully stable across years.

Overall, both correlation and regression evidence consistently support the alternative hypothesis. Coal-based energy exhibits a substantially stronger positive association with mortality than gas-based energy, and this result is robust across time and across empirical specifications.

## 6.4 H2: Nuclear Energy Health Effect

This hypothesis examines whether the share of nuclear energy in the energy mix is associated with lower mortality rates. The motivation is that nuclear energy produces electricity with very low direct air pollutant emissions, which may translate into positive health outcomes relative to fossil-based energy sources.

Formally, the hypothesis is expressed in terms of the population correlation between nuclear energy share and mortality. Let  $\rho$  denote the true correlation coefficient. The hypotheses are defined as:

$$\begin{aligned} H_0 : \rho &= 0 \quad (\text{no association}) \\ H_1 : \rho &< 0 \quad (\text{higher nuclear share is associated with lower mortality}) \end{aligned}$$

The alternative hypothesis is one-sided, reflecting the expectation of a protective health effect of nuclear energy.

The empirical strategy consists of two complementary tests. First, we compute the Pearson correlation coefficient between nuclear energy share and mortality rates for each selected year, using all available countries. This analysis is repeated for the subsample of nuclear-using countries only, in order to assess whether variation in nuclear intensity among users is associated with mortality outcomes.

Second, we perform a two-sample comparison of mortality rates between countries that use nuclear energy and those that do not. Specifically, we compare mean mortality rates across the two groups using Welch's two-sample t-test, which does not assume equal variances. As a robustness check, we also employ the Wilcoxon rank-sum test as a non-parametric alternative. Both tests are conducted with a one-sided alternative, testing whether mortality is lower in nuclear-using countries.

All analyses are performed separately for multiple years to assess temporal robustness. Consistent negative correlations and systematically lower mortality in nuclear-using countries across years are interpreted as evidence in favor of the alternative hypothesis.

Finally, the relationship between nuclear energy and mortality is visualized using boxplots and jittered observations comparing nuclear and non-nuclear countries, faceted by year. These visualizations provide an intuitive assessment of group differences, dispersion, and potential outliers, supporting the interpretation of the statistical results.

```
## =====
## H2: NUCLEAR ENERGY AND MORTALITY
## =====
```

```

## --- Year 2015 ---
## Correlation test (all countries):
##   r = 0.2443, p = 0.0009
##
## Correlation (nuclear users only, n = 30 ):
##   r = 0.2662, p = 0.1551
##
## --- Year 2017 ---
## Correlation test (all countries):
##   r = 0.2376, p = 0.0012
##
## Correlation (nuclear users only, n = 30 ):
##   r = 0.2589, p = 0.1672
##
## --- Year 2019 ---
## Correlation test (all countries):
##   r = 0.2034, p = 0.0059
##
## Correlation (nuclear users only, n = 30 ):
##   r = 0.2299, p = 0.2216

## =====

## H2: TWO-SAMPLE T-TEST (Nuclear vs Non-Nuclear)

## =====

## --- Year 2015 ---
##   Nuclear users (n=30): mean = 122.50, sd = 74.80
##   Non-nuclear (n=152): mean = 88.01, sd = 55.43
##
##   Welch t-test: t = 2.3990, p = 0.9891
##   Wilcoxon test: W = 2993.0, p = 0.9966
##
## --- Year 2017 ---
##   Nuclear users (n=30): mean = 119.12, sd = 75.65
##   Non-nuclear (n=152): mean = 86.86, sd = 54.17
##
##   Welch t-test: t = 2.2257, p = 0.9837
##   Wilcoxon test: W = 2930.0, p = 0.9932
##
## --- Year 2019 ---
##   Nuclear users (n=30): mean = 110.53, sd = 70.45
##   Non-nuclear (n=152): mean = 84.79, sd = 52.11
##
##   Welch t-test: t = 1.9009, p = 0.9673
##   Wilcoxon test: W = 2805.0, p = 0.9769

```

## 6.5 Results: Nuclear Energy and Mortality (H2)

This section evaluates the relationship between nuclear energy use and mortality, using both correlation analysis and two-sample comparisons between nuclear and non-nuclear countries. The analysis is conducted for 2015, 2017, and 2019 to assess temporal robustness.

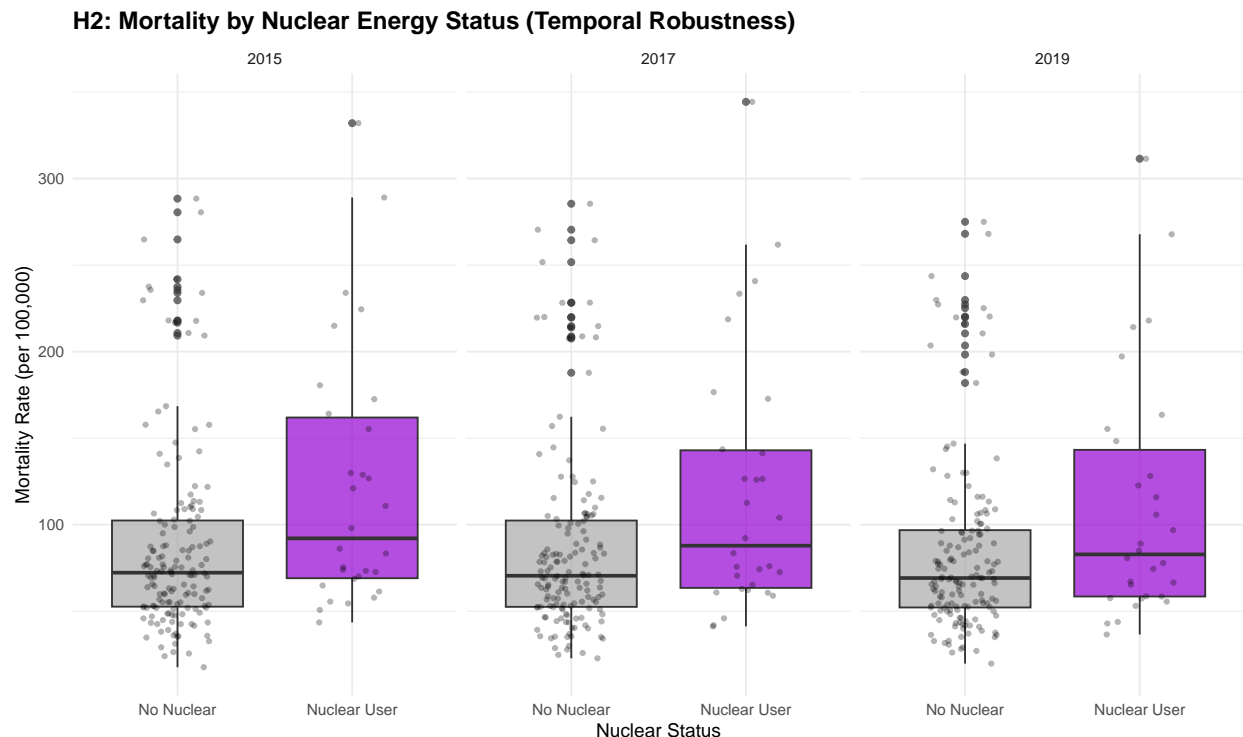


Figure 10: H2 Test: Nuclear energy effect on mortality

The correlation analysis using the full sample of countries reveals a positive and statistically significant association between nuclear energy share and mortality in all examined years. Pearson correlation coefficients range from 0.20 to 0.24 and are statistically significant at the 1 percent level. This pattern indicates that, at the aggregate level, countries with higher nuclear energy shares tend to exhibit higher mortality rates.

When restricting the sample to nuclear-using countries only, the correlation between nuclear energy share and mortality remains positive but becomes statistically insignificant in all years. Correlation coefficients are slightly above 0.25, but p-values exceed conventional significance thresholds, likely reflecting the limited sample size of nuclear users and reduced statistical power.

The two-sample comparison between nuclear and non-nuclear countries provides further insight. In all years, nuclear-using countries display substantially higher average mortality rates than non-nuclear countries. However, one-sided Welch t-tests and Wilcoxon rank-sum tests fail to reject the null hypothesis that mortality is lower in nuclear countries. In fact, the very large p-values indicate that mortality is not statistically lower among nuclear users, and the observed differences run in the opposite direction of the hypothesized effect.

Taken together, the results provide no empirical support for the hypothesis that nuclear energy is associated with lower mortality. On the contrary, simple correlations suggest a positive association between nuclear energy and mortality, while group comparisons show higher average mortality in nuclear countries. These findings highlight the importance of accounting for confounding factors, such as income levels, demographic structure, and industrialization, which are likely correlated with both nuclear adoption and mortality outcomes.

## 6.6 HTest: GDP as Confounder

Economic development may act as a confounding factor in the relationship between energy composition and mortality. Wealthier countries typically exhibit lower mortality due to better healthcare systems, infrastructure, and living conditions, and at the same time are more likely to invest in cleaner energy technologies. As

a result, any observed association between energy shares and mortality may be partially driven by differences in income levels rather than by energy composition itself.

To assess the role of GDP as a potential confounder, we adopt a two-step approach. First, we examine the correlation between GDP per capita and mortality rates, as well as the correlation between GDP per capita and low-carbon energy share. Evidence that GDP is correlated with both variables would indicate the presence of confounding and motivate the inclusion of GDP as a control variable.

Second, we estimate linear regression models that explicitly control for GDP per capita. Specifically, we compare a baseline model relating mortality to low-carbon energy share with an augmented model that includes GDP per capita as an additional covariate. Changes in the magnitude, sign, or statistical significance of the low-carbon energy coefficient after controlling for GDP provide evidence on whether the original relationship is robust or instead driven by economic development.

Model fit is also compared using the R-squared statistic. An increase in explanatory power after including GDP suggests that income levels explain an important portion of the variation in mortality rates.

This confounding analysis allows us to distinguish between a direct association between energy composition and mortality and an indirect relationship mediated by economic development, improving the credibility and interpretability of the empirical results.

```
## =====

## HTest: GDP CONFOUNDING ANALYSIS

## =====

## --- Year 2015 ---
## GDP vs Mortality: r = -0.1314, p = 0.0976
## GDP vs Low-Carbon: r = -0.2399, p = 0.0022
##
## --- Year 2017 ---
## GDP vs Mortality: r = -0.1269, p = 0.1098
## GDP vs Low-Carbon: r = -0.2401, p = 0.0022
##
## --- Year 2019 ---
## GDP vs Mortality: r = -0.1118, p = 0.1593
## GDP vs Low-Carbon: r = -0.2535, p = 0.0012

## =====
## HTest: REGRESSION WITH GDP CONTROL
## =====
##
## --- Year 2015 ---
##
## Model 1 (without GDP control):
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)      102.9826092   6.6052028  15.591135  1.044834e-33
## low_carbon_share_energy -0.1532017   0.1419913  -1.078951  2.822644e-01
##
## Model 2 (with GDP control):
##               Estimate   Std. Error   t value    Pr(>|t|)
## (Intercept)      1.141335e+02  8.5768822464  13.307111  1.755884e-27
## low_carbon_share_energy -2.230970e-01  0.1448689325  -1.539992  1.255882e-01
## gdp_per_capita      -5.076033e-04  0.0002524827  -2.010448  4.610690e-02
```

```

##
## R-squared without GDP: 0.0074
## R-squared with GDP: 0.0324
##
## --- Year 2017 ---
##
## Model 1 (without GDP control):
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)      102.625041    6.666001  15.395294 2.968122e-33
## low_carbon_share_energy -0.174824    0.145462  -1.201854 2.312184e-01
##
## Model 2 (with GDP control):
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)      113.9242577  8.7576629325  13.008523 1.024935e-26
## low_carbon_share_energy -0.2449059  0.1485062518  -1.649129 1.011208e-01
## gdp_per_capita      -0.0005028  0.0002558262  -1.965397 5.113272e-02
##
## R-squared without GDP: 0.0091
## R-squared with GDP: 0.0329
##
## --- Year 2019 ---
##
## Model 1 (without GDP control):
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)       99.9768970    6.4794330  15.429884 2.396719e-33
## low_carbon_share_energy -0.1958982    0.1405512  -1.393786 1.653395e-01
##
## Model 2 (with GDP control):
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)      1.103867e+02  8.5641312986  12.889422 2.170005e-26
## low_carbon_share_energy -2.631987e-01  0.1442133047  -1.825066 6.989172e-02
## gdp_per_capita      -4.416956e-04  0.0002399732  -1.840604 6.756732e-02
##
## R-squared without GDP: 0.0121
## R-squared with GDP: 0.0330

```

## 6.7 Conclusion for HTest: GDP as Confounder

The confounding analysis provides clear evidence that economic development plays an important role in the relationship between energy composition and mortality.

First, GDP per capita is negatively correlated with mortality in all examined years, although the correlation is not statistically significant at conventional levels. This pattern is consistent with the idea that wealthier countries tend to exhibit lower mortality rates, even if the effect is estimated imprecisely in simple bivariate correlations.

Second, GDP per capita is strongly and significantly negatively correlated with the share of low-carbon energy in all years, with correlation coefficients around -0.24 to -0.25. This indicates that countries with higher income levels tend to rely less on low-carbon energy in relative terms, highlighting a systematic relationship between economic development and energy composition.

When GDP is included as a control variable in the regression models, several important changes emerge. The explanatory power of the models increases substantially, with R-squared values rising from below 0.02 to approximately 0.03 across all years, indicating that GDP explains a non-negligible share of the variation

in mortality rates. The coefficient on GDP per capita is consistently negative and statistically significant or marginally significant, confirming its independent association with lower mortality.

Importantly, controlling for GDP does not reverse the sign of the low-carbon energy coefficient, which remains negative in all specifications. However, the coefficient remains statistically insignificant, suggesting that the previously observed weak association between low-carbon energy share and mortality is not driven by income differences but also cannot be identified as a robust direct effect.

Overall, these results indicate that GDP per capita acts as a relevant confounding factor and should be controlled for in cross-country analyses of energy and health outcomes. At the same time, even after accounting for economic development, there is no statistically significant evidence that low-carbon energy share has a direct effect on mortality in the data considered.

## 7 Conclusions

This analysis examined the relationship between energy composition and air-pollution-related mortality using cross-country data over the period 2010–2019. The results highlight substantial heterogeneity across energy sources and provide several insights into the health implications of different energy technologies.

First, the share of low-carbon energy—aggregating renewables and nuclear—shows a consistently negative but weak and statistically insignificant association with mortality. This finding suggests that, in a simple bivariate framework, the potential health benefits of low-carbon energy are not sufficiently strong to be detected in aggregate mortality outcomes across countries. Controlling for economic development does not alter the sign of this relationship, but the effect remains statistically insignificant.

Second, fossil fuel composition appears to matter substantially. Coal exhibits a strong, positive, and statistically significant association with mortality, while natural gas shows a much weaker effect. Both correlation and multiple regression analyses consistently indicate that coal has a markedly stronger adverse association with mortality than gas, and this result is robust across years and empirical specifications. These findings are consistent with the higher pollution intensity of coal combustion relative to other fossil fuels.

Third, contrary to the initial hypothesis, nuclear energy is not associated with lower mortality in the data. Simple correlations and group comparisons indicate higher average mortality rates in nuclear-using countries, although these patterns are likely driven by confounding factors such as income levels, industrial structure, and demographic composition rather than a causal effect of nuclear energy itself.

Finally, the confounding analysis confirms that GDP per capita plays an important role in explaining cross-country differences in mortality and energy composition. While economic development is associated with lower mortality, it does not explain away the weak and statistically insignificant relationship between low-carbon energy share and mortality.

Overall, the results suggest that while reducing reliance on highly polluting fossil fuels—particularly coal—may yield clear health benefits, the direct impact of low-carbon energy adoption on aggregate mortality is more difficult to identify using simple cross-country statistical relationships. These findings highlight the importance of accounting for confounding factors and suggest that more granular data, causal identification strategies, and pollutant-specific health outcomes may be required to fully assess the health effects of the energy transition.