

# Phase3

Karla Otoude and Chloé Marquis

**Link to our github project**

[Our Github Project](#)

## I. Research question

**How are changes in health and social protection expenditures associated with avoidable mortality across OECD countries over time?**

Population health differs significantly across countries. One key factor explaining these differences is the strength of social and health protection systems. In particular, public investment in healthcare and social protection is expected to reduce mortality from causes that are considered preventable or treatable through effective public policies.

This report seeks to understand the relationship between social and health protection spending and avoidable mortality, used as an indicator of the effectiveness of a country's health system. Avoidable mortality captures premature deaths (under age 75) that could be prevented through healthcare, making it a relevant proxy for population health.

There are two goals in our analysis. First, using panel data for 41 OECD countries between 2000 and 2023, this report aims to describe and analyse the global evolution of avoidable mortality across OECD countries and assess if high levels of social and health spending are correlated with lower avoidable mortality rates. Second, it focuses on the Covid-19 crisis, to understand if countries with a stronger health protection pre-Covid experienced lower avoidable mortality rates from 2019 to 2021.

## Method

To conduct our study we will be using a **linear mixed model (LMM)**. Lagged variables will be used as investments in both the social and health systems likely impact the population's general health with a delay. Such a model will allow us to properly study the OECD countries taking into account their differences.

For explanatory variables, in order to estimate the strength and quality of a country's health protection scheme, we will use the following variables:

- Health expenditure
- Social expenditure aggregates
- Number of hospitals beds

We also have more general variables such as :

- GDP per capita
- Unemployment rate
- Human development index
- Life expectancy

They will be used as control variables, since richer countries tend to have better living conditions which may lead to a generally healthier population, without the need for a strong health protection.

Finally, the explained variable will be :

- Avoidable mortality

## II. Description of the datasets

This report relies on 8 datasets. Here is a short description of each of them. More details on the datasets and their descriptive statistics can be found in Phase 2.

### **Health expenditure - OECD (2025)**

This dataset details the annual health expenditure and financing as a percentage of GDP for 52 countries. The data is available for a period ranging from 2000 to 2024. It was computed using financial flows related to the consumption of healthcare goods and services from different health providers such as Hospitals or Residential long-term care facilities.

### **Avoidable mortality - OECD (2025)**

This dataset describes the number of “avoidable deaths” per 100,000 inhabitants annually for the period 2000-2024 for 46 countries. It contains both “preventable mortality” and “treatable (or amenable) mortality”. The first refers to deaths that can be avoided through effective public health and primary prevention interventions. The second refers to timely and effective health care interventions, including prevention and treatment. Both indicators refer to premature mortality (under age 75).

### **GDP per Capita - World Bank (2024)**

The GDP per Capita dataset presents the GDP per capita (constant 2015 US\$) from 1960 to 2024 for 70 countries. This indicator is expressed in constant prices, meaning the series has been adjusted to account for price changes over time. The reference year for this adjustment is 2015. This indicator is expressed in United States dollars.

### **Hospital beds - OECD (2025)**

This dataset provides data on the number of total hospitals beds by function of healthcare and by type of care (ie. somatic or psychiatric care) for the period 2000-2024 for 48 different countries. Total hospital beds are the sum of the following categories: Curative care (acute care) beds in hospitals, Rehabilitative care beds in hospitals, Long-term care beds in hospitals, All other beds in hospitals not elsewhere classified.

### **Life expectancy - World Bank (2023)**

This dataset provides the life expectancy in years for women and men for a period ranging from 1960 to 2023, for 69 different countries or economies.

### **Social Expenditure Aggregate - OECD (2025)**

This dataset includes statistics on public and private social expenditure at program level (e.g. Old age, Survivors, Incapacity-related benefits, Health, etc) as a percentage of GDP. It covers 38 OECD countries for the period 1980-2021/23 and estimates for aggregates for 2022-24.

### **Unemployment Rate - IMF (2025)**

This IMF dataset describes the percentage of the labor force that is unemployed and actively seeking employment. It is available from 1980 to 2025 and features 120 geographic regions.

### **HDI (Human Development Index) - UNDP (2025)**

This dataset describes the human development index, per country (195), annually (1990-2023). It is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living. The HDI is between 0 (low human development) and 1 (very high human development).

Table 1: Brief summary of the datasets used in this report

Dataset	Observations	Variables
Health Expenditure	1246	46
Avoidable Mortality	997	44
GDP per Capita	266	70
Number of Hospital Beds	1034	38
Life Expectancy	266	70
Social Expenditure	1046	34
Unemployment	122	32
HDI	4484	10

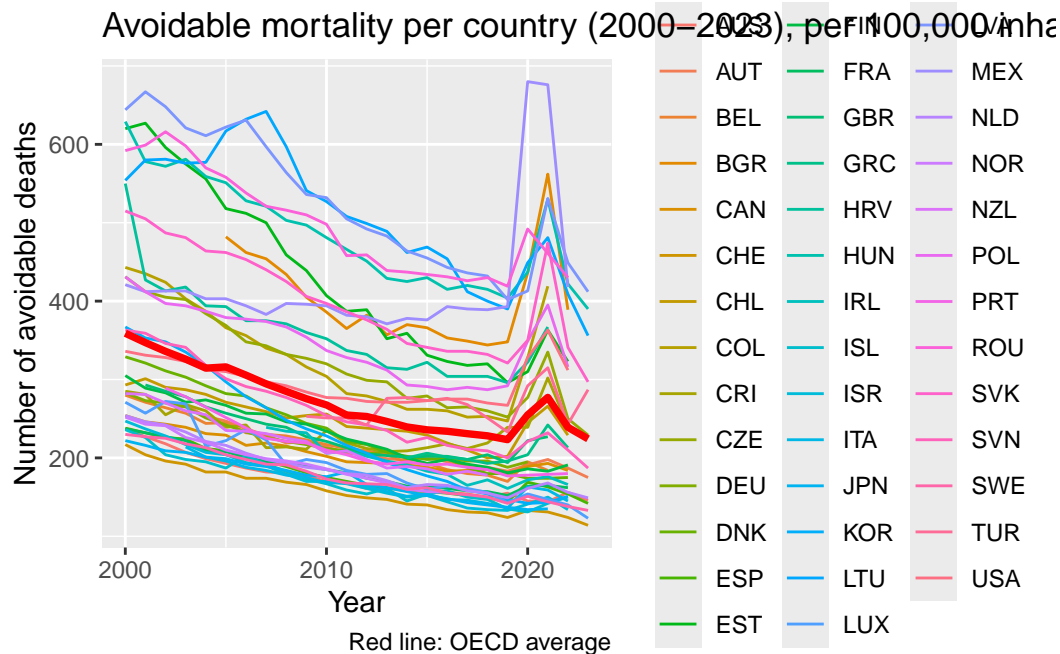
### Description of the merged and cleaned dataset

Number of rows	Number of columns	Percentage of NA	Number of countries studied	Time period
888	10	1.05%	41	2000-2023

## 3. Data Analysis

We first start by a visual exploration of our main variable : avoidable mortality (*OBS\_VALUE\_mortality*) and its link to explanatory variables.

## Evolution of avoidable mortality per country

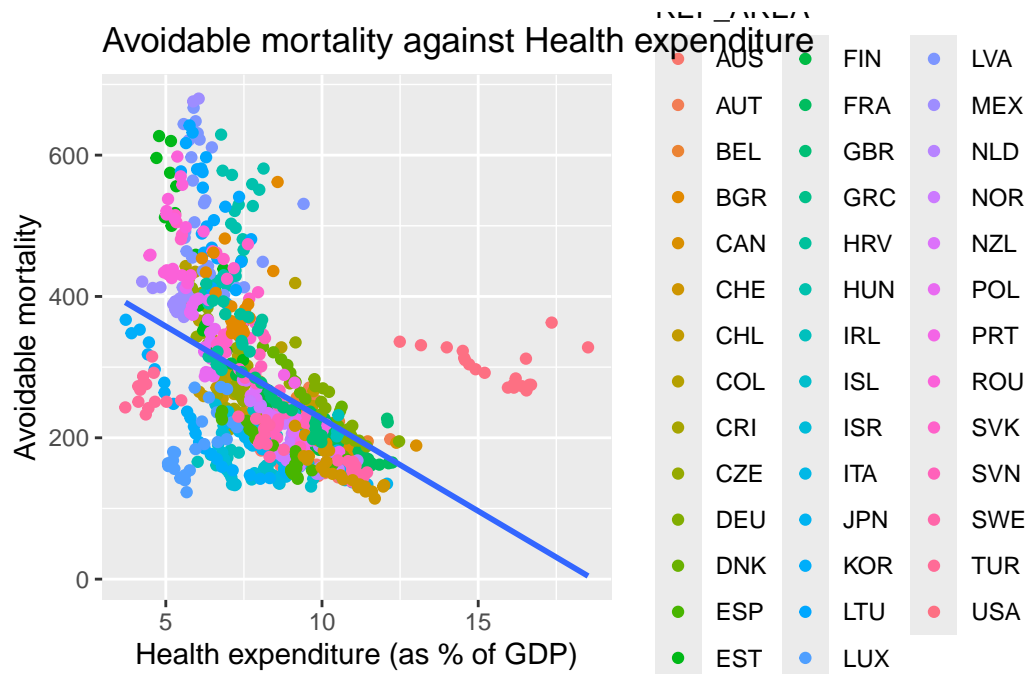


Most countries share the same downward sloping trend in avoidable mortality. A distinct increase appears in 2019, with a peak in 2020 and a decrease to previous trend values from 2021 in all countries. The country with the lowest avoidable death rate is Switzerland (CHE) and the highest is Latvia at the beginning of the 2000s, then surpassed by Lithuania (LTU) around 2007. Additionally, there is a strong increase in avoidable deaths from around 400 to almost 650 per 100,000 inhabitants in Mexico in 2019.

It is important to note that this report primarily studies developed countries with the most available data from 2000 to 2023. We chose to exclude the following countries : South Africa (ZAF), Argentina (ARG), Peru (PER), Thailand (THA) and Brazil (BRA) as the data was not sufficient to proceed with proper data analysis. Thus, our conclusions will be biased due to the fact that we study more developed countries with, perhaps, more similar institutional structures and resistance to global health crises such as Covid-19.

## Visualising *avoidable\_mortality* against independent variables

The goal of our analysis is to determine if the evolution of health and social expenditures are correlated with the evolution of avoidable mortality. We can start the data analysis by simply plotting some independent variables against *avoidable\_mortality* and observe the general trend.



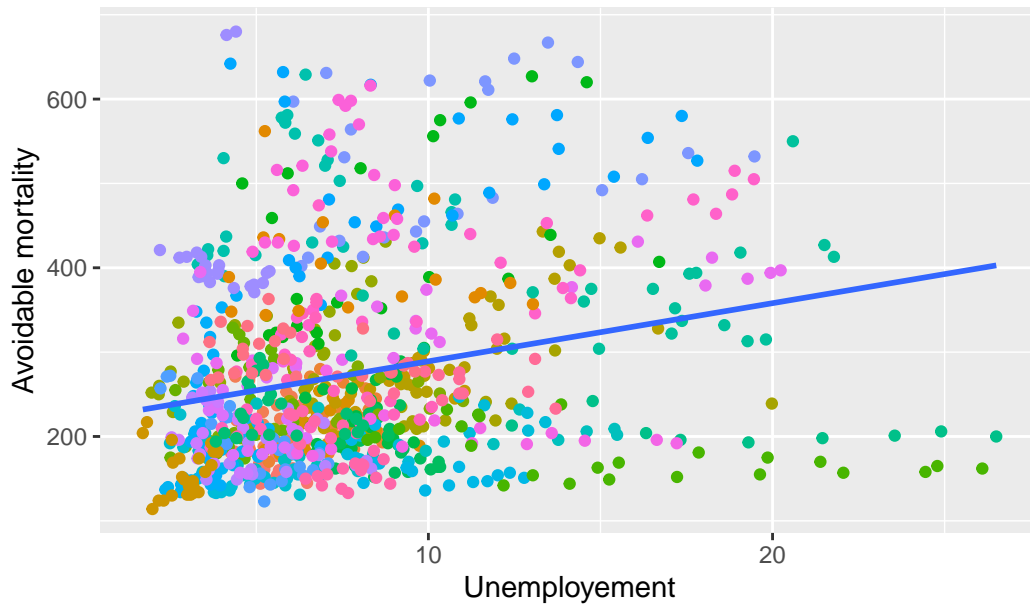
As expected, as the level of health expenditure increases, avoidable mortality tends to decrease. Australia (AUS) detaches itself from other countries by having a particularly high level of health expenditure, but still a moderately high level of avoidable mortality. Most countries are situated on the left side of the graph (with low health expenditure) showcasing countries with high and low rates of avoidable mortality for the same level of health expenditure.

Avoidable mortality against Social Expenditure



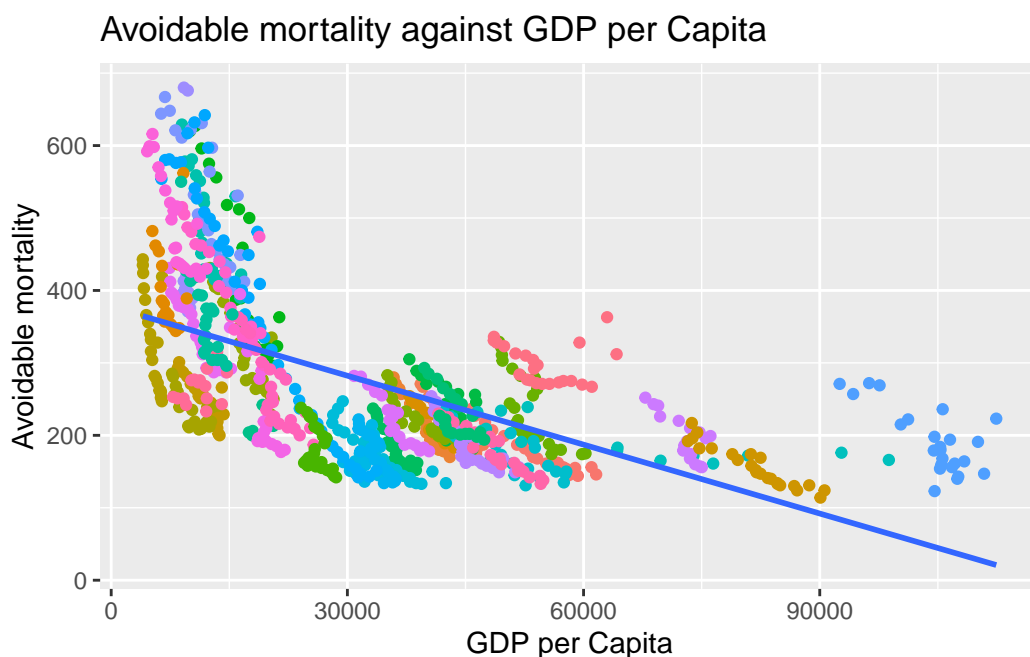
As social expenditure increases, the avoidable mortality rate decreases. Most countries' social expenditure is within 10% and 25% of their GDP. However, for these countries similar in social expenditure levels, avoidable mortality goes from 100 to almost 700 per 100,000 deaths, suggesting an important variability. This observation is consistent with the trend observed using the health expenditure variable. For similar levels of health or social expenditure, avoidable mortality seems to vary significantly between countries.

Avoidable mortality against Unemployment



Here, the trend indicates that as unemployment increases, avoidable mortality does too. Most countries have low levels of unemployment and low rates of avoidable mortality. In the bottom right, Spain (ESP) and Greece (GRC) display very high rates of unemployment while keeping their avoidable mortality rates among the lowest of all countries. On the contrary, countries such as Mexico (MEX) in purple showcase high levels of avoidable mortality with very low rates of unemployment.





This graph suggests that as GDP per capita increases, the rate of avoidable mortality decreases. The avoidable mortality rate doesn't seem to go below 100 deaths per 100,000 inhabitants. Thus, as GDP per capita increases, avoidable mortality remains between 100 and 300 deaths per 100,000 inhabitants suggesting marginal change in mortality as a country's population becomes richer.

#### 4. Econometric Analysis : Linear Mixed Model (LMM)

##### Correlation matrix for control variables

	OBS_VALUE_mortality	OBS_VALUE_gdp	OBS_VALUE_hdi	OBS_VALUE_life_exp	OBS_VALUE_unemp
OBS_VALUE_mortality	1.0000000	-0.6047638	-0.6593684	-0.9559693	0.2393598
OBS_VALUE_gdp	-0.6047638	1.0000000	0.7179984	0.6348101	-0.4273404
OBS_VALUE_hdi	-0.6593684	0.7179984	1.0000000	0.7670182	-0.3755789
OBS_VALUE_life_exp	-0.9559693	0.6348101	0.7670182	1.0000000	-0.2437086
OBS_VALUE_unemp	0.2393598	-0.4273404	-0.3755789	-0.2437086	1.0000000

The correlation matrix reveals strong linear relationships among several control variables.

- Avoidable mortality is highly negatively correlated with life expectancy (-0.96), reflecting the fact that both variables capture very similar aspects of general population health.

- Life expectancy is also strongly correlated with HDI (0.76) and GDP per capita (0.65), meaning there is an overlap between these variables.

In order to avoid multicollinearity issues in our regression, we choose to exclude HDI and life expectancy from our control variables. We retain GDP per capita and unemployment, which can adjust for differences in economic development between countries and remain sufficiently distinct from avoidable mortality.

## Creating lag operators/ lagged variables

Before further modelling, we create lagged version of our variables. We believe that past investments in health sector and social protection have an impact on mortality. Using one-year lags will allow us to study short term effects such as changes in government budgets from one year to the next. 3-year and 5-year moving averages will also be used to capture long-term investment impacts. Indeed, we assume increased hospital capacities or training staff takes a few years before benefiting the health system.

We will create lagged versions of 4 variables : mortality, health expenditure, social expenditure and hospital beds.

Consequently we have :

- `mortality_lag1` : 1-year lag of avoidable mortality
- `healthExp_lag1` : 1-year lag of health expenditure (short-run)
- `healthExp_ma3` : 3-year average lag of health expenditure (medium-run)
- `healthExp_ma5` : 5-year average lag of health expenditure (long-run)
- `hospBeds_lag1` : 1-year lag of hospital beds
- `hospBeds_ma3` : 3-year average lag of hospital beds
- `hospBeds_ma5` : 5-year average lag of hospital beds
- `socialExp_lag1` : 1-year lag of social expenditure
- `socialExp_ma3` : 3-year average lag of social expenditure
- `socialExp_ma5` : 5-year average lag of social expenditure

Additionally, to account for the Covid-19 shock, we introduce a pandemic dummy variable equal to 1 for the years 2019 to 2021 and 0 otherwise. We also add 2 interaction terms with Health expenditure and Social expenditure to assess the impact of these factors on avoidable mortality during this period.

MODIF

## Model testing

### Testing for the short, medium and long-run

We do a first test to see how our variables interact with each other, especially lags

The model we first want to test is :

$$\begin{aligned} Mortality_{i,t} = & \gamma_{0,i} + \beta_1 mortality_{i,t-1} + \beta_2 health.expenditure_{i,t-1} + \beta_3 health.expenditure_{i,t-[1,5]} \\ & + \beta_4 hospital.beds_{i,t-1} + \beta_5 hospital.beds_{i,t-[1,5]} + \beta_6 social.expenditure_{i,t-1} \\ & + \beta_7 social.expenditure_{i,t-[1,5]} + \gamma_8 GDP.per.capita_{i,t} + \gamma_9 Unemployment_{i,t} \end{aligned}$$

With

- i = index for country i
- t = index for period t (here our periods are years)

We also assume that mortality (at time t-1), social expenditure and hospital beds, our explanatory variables, are fixed effects. GDP per capita and Unemployment, our control variables, are random effects (meaning that we take into account in this model the fact that they will be different for each country).

- Fixed effects have their coefficient represented by the symbol beta
- Random effects have their coefficient represented by the symbol gamma

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula:

```
OBS_VALUE_mortality ~ mortality_lag1 + healthExp_lag1 + healthExp_ma5 +  
  hospBeds_lag1 + hospBeds_ma5 + socialExp_lag1 + socialExp_ma5 +  
  OBS_VALUE_gdp + OBS_VALUE_unemp + (1 | REF_AREA)
```

Data: data\_lagged

Control: lmerControl(autoscale = TRUE)

AIC	BIC	logLik	-2*log(L)	df.resid
5680.7	5733.6	-2828.4	5656.7	593

Scaled residuals:

Min	1Q	Median	3Q	Max
-8.9536	-0.2717	-0.0653	0.1424	11.1295

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	0.0	0.00
Residual		673.2	25.95

Number of obs: 605, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	3.096e+01	8.534e+00	3.628
mortality_lag1	9.255e-01	1.257e-02	73.615
healthExp_lag1	-5.647e+00	3.564e+00	-1.585
healthExp_ma5	6.053e+00	3.703e+00	1.634
hospBeds_lag1	-2.182e-04	1.163e-04	-1.876
hospBeds_ma5	2.139e-04	1.151e-04	1.858
socialExp_lag1	1.263e+00	1.188e+00	1.063
socialExp_ma5	-1.826e+00	1.205e+00	-1.516
OBS_VALUE_gdp	-1.965e-04	6.719e-05	-2.925
OBS_VALUE_unemp	-1.366e-02	3.306e-01	-0.041

Correlation of Fixed Effects:

	(Intr)	mrtl_1	hltE_1	hltE_5	hspB_1	hspB_5	sclE_1	sclE_5	OBS_VALUE_g
mortality_lag1	-0.780								
healthExp_lag1	-0.019	0.013							
healthExp_ma5	-0.054	0.014	-0.984						
hospBeds_lag1	-0.414	0.238	-0.058	0.072					
hospBeds_ma5	0.412	-0.234	0.057	-0.073	-1.000				
socialExp_lag1	0.049	-0.044	-0.696	0.677	0.008	-0.010			
socialExp_ma5	-0.121	0.088	0.682	-0.677	0.066	-0.064	-0.980		
OBS_VALUE_g	-0.428	0.495	0.088	-0.129	0.085	-0.078	-0.059	0.057	
OBS_VALUE_n	-0.342	0.105	0.142	-0.139	0.087	-0.081	-0.191	0.154	0.393

optimizer (nloptwrap) convergence code: 0 (OK)  
boundary (singular) fit: see help('isSingular')

After this first test, we notice multiple problems :

- The correlation between mortality\_lag1 and the intercept is too high, nullifying the effect of the random intercept which ends up having a variance and a standard deviation of 0. We will remove mortality\_lag1 from the next regressions to avoid such a result.
- Using both the first lag and the long-run moving average of variables in the same model is not possible. The correlation between the two makes it hard to distinguish which has an effect on mortality and the long-run moving average of explanatory variables end up having incoherent coefficients.

Therefore we will compare short-run, medium-run and long-run in 3 different models

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula:

OBS\_VALUE\_mortality ~ healthExp\_lag1 + hospBeds\_lag1 + socialExp\_lag1 +  
OBS\_VALUE\_gdp + OBS\_VALUE\_unemp + (1 | REF\_AREA)

Data: data\_lagged

Control: lmerControl(autoscale = TRUE)

AIC	BIC	logLik	-2*log(L)	df.resid
8065.7	8102.9	-4024.8	8049.7	770

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.8907	-0.4959	-0.1371	0.3434	7.2939

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	8662	93.07
Residual		1422	37.71

Number of obs: 778, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.592e+02	2.345e+01	23.850
healthExp_lag1	-1.109e+01	2.422e+00	-4.579
hospBeds_lag1	-4.659e-05	3.150e-05	-1.479
socialExp_lag1	-3.609e+00	1.027e+00	-3.514
OBS_VALUE_gdp	-4.174e-03	3.980e-04	-10.488
OBS_VALUE_unemp	1.511e+00	6.273e-01	2.409

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: OBS\_VALUE\_mortality ~ healthExp\_ma3 + hospBeds\_ma3 + socialExp\_ma3 +  
OBS\_VALUE\_gdp + OBS\_VALUE\_unemp + (1 | REF\_AREA)

Data: data\_lagged

Control: lmerControl(autoscale = TRUE)

AIC	BIC	logLik	-2*log(L)	df.resid
7086.9	7123.2	-3535.4	7070.9	683

Scaled residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-2.6635 -0.4656 -0.1229 0.2703 7.5496

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	7513	86.68
	Residual	1239	35.21

Number of obs: 691, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.660e+02	2.431e+01	23.288
healthExp_ma3	-9.485e+00	2.686e+00	-3.532
hospBeds_ma3	-4.599e-05	3.129e-05	-1.470
socialExp_ma3	-6.106e+00	1.183e+00	-5.161
OBS_VALUE_gdp	-3.404e-03	4.066e-04	-8.372
OBS_VALUE_unemp	1.362e+00	6.392e-01	2.130

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: OBS\_VALUE\_mortality ~ healthExp\_ma5 + hospBeds\_ma5 + socialExp\_ma5 +  
OBS\_VALUE\_gdp + OBS\_VALUE\_unemp + (1 | REF\_AREA)

Data: data\_lagged

Control: lmerControl(autoscale = TRUE)

AIC	BIC	logLik	-2*log(L)	df.resid
6185.0	6220.2	-3084.5	6169.0	597

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.9289	-0.4044	-0.0966	0.2031	7.6467

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	6564	81.02
	Residual	1168	34.17

Number of obs: 605, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.505e+02	2.611e+01	21.082
healthExp_ma5	-8.276e+00	3.001e+00	-2.758
hospBeds_ma5	-4.099e-05	3.171e-05	-1.293
socialExp_ma5	-6.323e+00	1.333e+00	-4.745

OBS_VALUE_gdp	-3.057e-03	4.342e-04	-7.041
OBS_VALUE_unemp	5.655e-01	6.569e-01	0.861