

# Phase3

Karla Otuode and Chloé Marquis

## Link to our github project

[Our Github Project](#)

### I. Research question

**How does the level of social protection impact the general population health ?**

We want to identify if a stronger social and health protection (quantified by different variables such as health expenditure, tax revenues, etc) has an impact on the avoidable mortality rate. In order to do so we will analyse data collected in OECD countries from 2000 to 2023.

There are two goals in our analysis : The first is to understand the global evolution of the data and the reasons behind its trends. The second is to study the Covid-19 crisis. We believe it revealed how necessary having a proper social protection is to limit the number of deaths during a pandemic.

### Method

To conduct our study we will be using a **linear mixed model (LMM)** with lags. Such a model will allow us to properly study the OECD countries taking into account their differences. It is also a fitting model since we are working with time-based data.

Considering our research question, we will use multiple datasets and variables.

For explanatory variables, in order to estimate the strength and quality of a country's health protection scheme, we will have : - 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 It will be used to estimate the population's general health.

## **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 hospital 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 internationally comparable statistics on public and mandatory and voluntary private social expenditure at programme level (e.g. Old age, Survivors, Incapacity-related benefits, Health, etc) It covers 38 OECD countries and some accession 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

Dataset	Observations	Variables
Social Expenditure	1046	34
Unemployment	122	32
HDI	4484	10

### Description of the merged and cleaned dataset

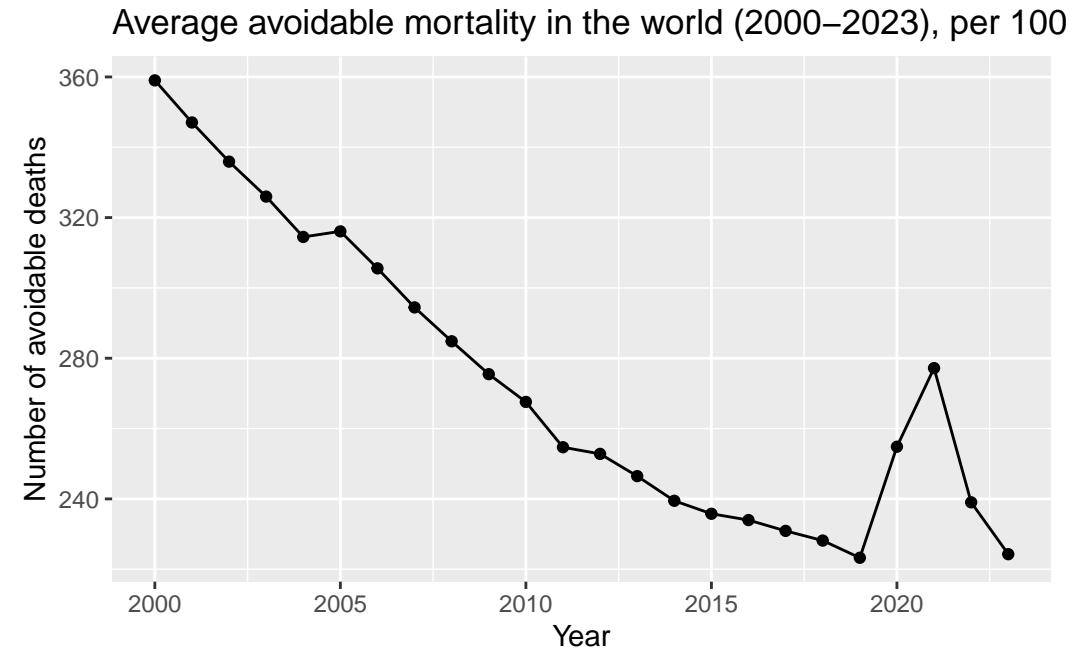
Number of rows	Number of columns	Percentage of NA	Number of countries studied	Time period
888	10	1.05%	41	2000-2023

## 3. Data Analysis

*Graphical representation of our main variable : avoidable mortality (OBS\_VALUE\_mortality)*

### Global avoidable mortality

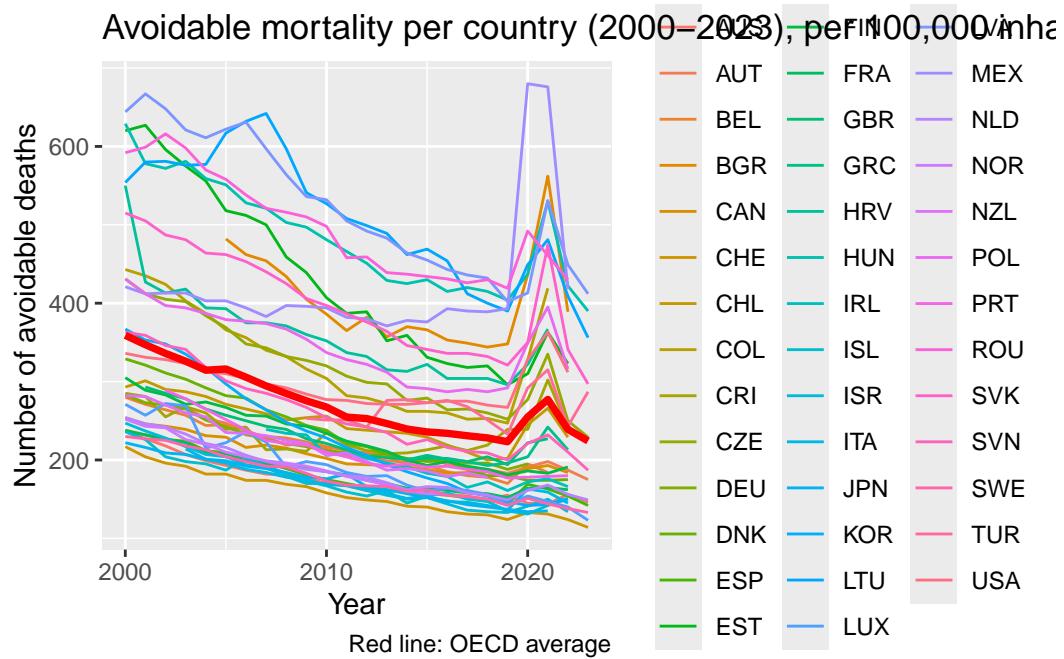
First we can take a look at the number of avoidable deaths globally



This graph reveals a downward trend in avoidable mortality from 2000 to 2019 cut by a sudden increase in 2020 and 2021 during Covid-19. However it seems that, afterward, the variable will go back to its previous trend.

### Avoidable mortality per country

We also check the `avoidable_mortality` variable per country.

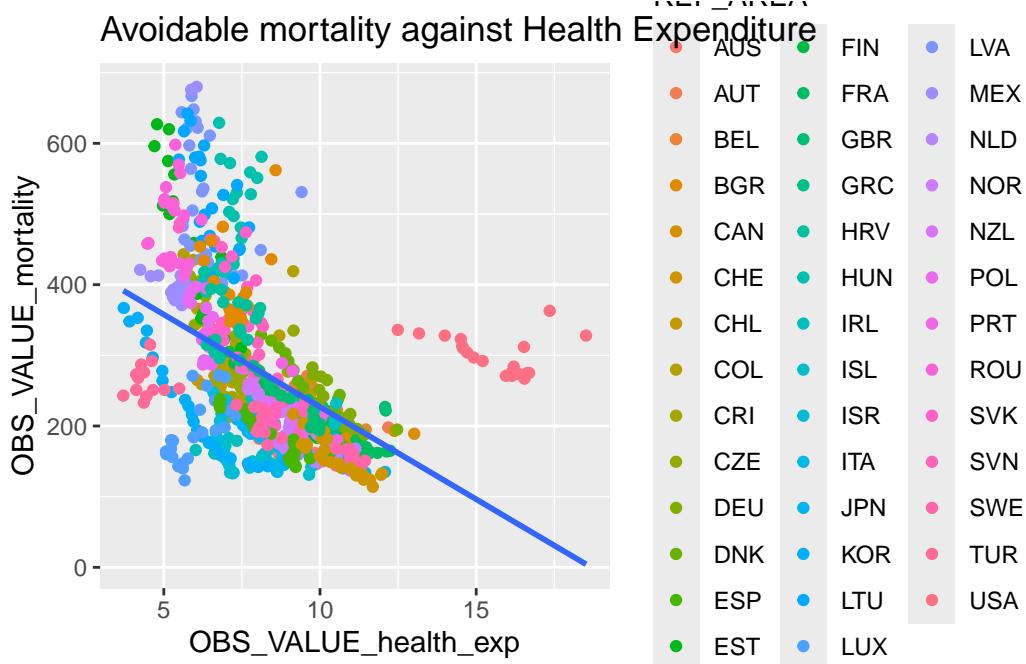


Most countries share the same downward sloping trend in avoidable mortality. A distinct increase appears in 2019, with a peak in 2020 and a decrease to previous trend values from 2021 in all countries. The country with the lowest avoidable death rate is Switzerland (CHE) and the highest is Latvia at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

It is important to note that we 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 is 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.

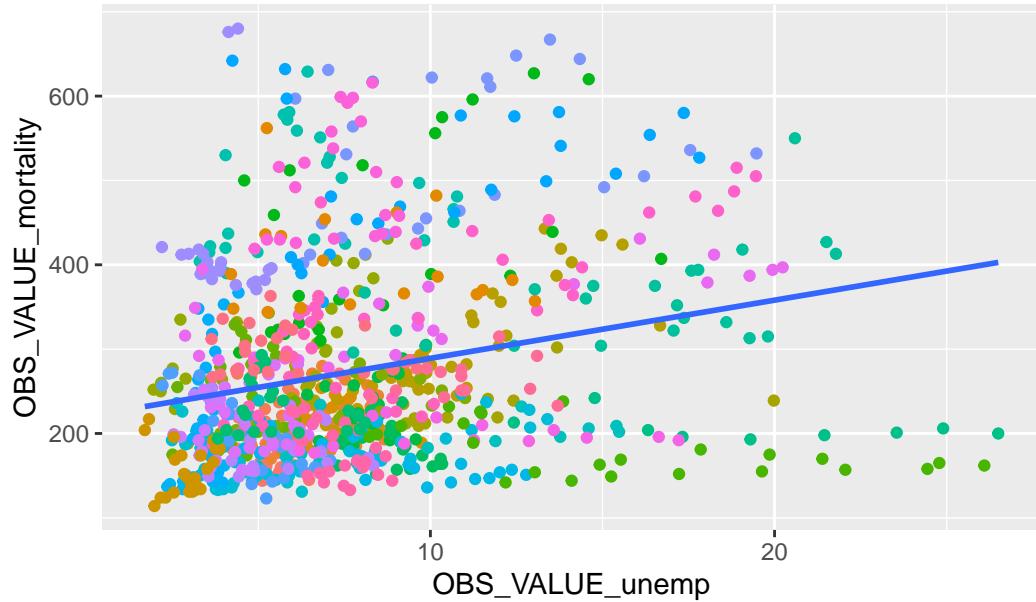
## Visualising *avoidable\_mortality* against independent variables

The goal of our analysis is to determine if certain variables explain the level of avoidable mortality. We can start the data analysis by simply plotting some independent variables against *avoidable\_mortality* and observe the general trend.



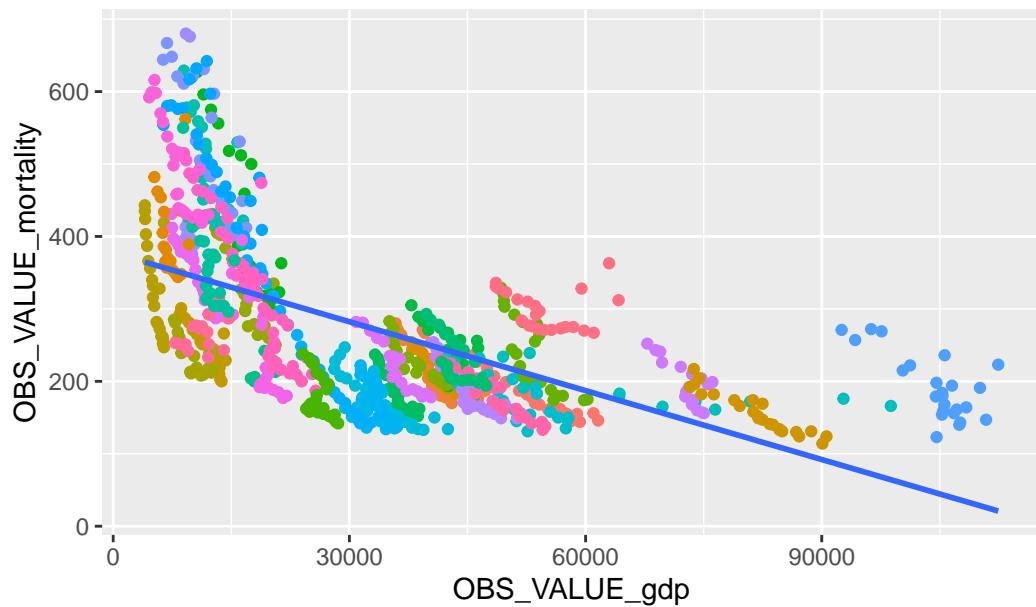
As expected, as the level of health expenditure increases, avoidable mortality tends to decrease. Australia (AUS) detaches itself from other countries by having a particularly high level of health expenditure, but still a moderately high level of avoidable mortality. Most countries are situated on the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

## Avoidable mortality against Unemployment



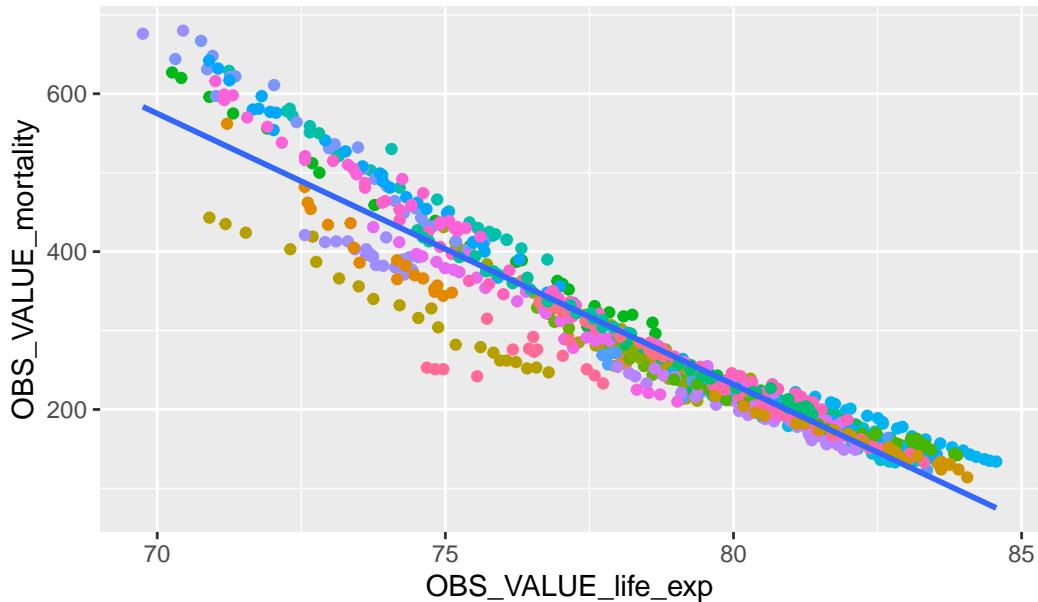
Here, the trend indicates that as unemployment increases, avoidable mortality does too. Most countries have low levels of unemployment and low rates of avoidable mortality. In the bottom right, Spain (ESP) and Greece (GRC) display very high rates of unemployment while keeping their avoidable mortality 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.

### Avoidable mortality against GDP per Capita



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.

## Avoidable mortality against Life Expectancy



The correlation between avoidable mortality and life expectancy seems strong in this graph. Indeed, as life expectancy increases, avoidable mortality decreases linearly for all countries. Two countries (Türkiye (TUR) in pink and Colombia (COL) in brown) find themselves slightly below the trend line, with lower mortality than the trend would suggest for their level of life expectancy.

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

### **Creating lag operators/ lagged variables**

Before making any regression, we created lagged version of our variables. We believe that past investments in health sector and social protection have an impact on mortality and we want this to appear in our model.

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

Consequently we have :

- `mortality_lag1` : first lag of the mortality variable (at time t-1)
- `mortality_lag2` : second lag of the mortality variable (at time t-2)
- `healthExp_lag1` : first lag of the health expenditure variable (at time t-1)
- `healthExp_lag2` : second lag of the health expenditure variable (at time t-2)

- `hospBeds_lag1` : first lag of the hospital Beds variable (at time t-1)
- `hospBeds_lag2` : second lag of the hospital Beds variable (at time t-2)
- `socialExp_lag1` : first lag of the social expenditure variable (at time t-1)
- `socialExp_lag2` : second lag of the social expenditure variable (at time t-2)

## Model testing

### First test

The model we first want to test is :

$$\begin{aligned}
 Mortality_{i,t} = & \gamma_0{}_i + \beta_1 mortality_{i,t-1} + \beta_2 mortality_{i,t-2} + \beta_3 health.expenditure_{i,t-1} \\
 & + \beta_4 health.expenditure_{i,t-2} + \beta_5 Hospital.Beds_{i,t-1} + \beta_6 Hospital.Beds_{i,t-2} \\
 & + \gamma_7 GDP.per.capita_{i,t} + \gamma_8 Unemployment_{i,t} + \gamma_9 HDI_{i,t} \\
 & + \gamma_{10} Life.expectancy_{i,t}
 \end{aligned}$$

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, Unemployment, Life expectancy and HDI, 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 + mortality_lag2 + healthExp_lag1 +
  healthExp_lag2 + hospBeds_lag1 + hospBeds_lag2 + OBS_VALUE_gdp +
  OBS_VALUE_unemp + OBS_VALUE_hdi + OBS_VALUE_life_exp + (OBS_VALUE_gdp +
  OBS_VALUE_unemp + OBS_VALUE_hdi + OBS_VALUE_life_exp | REF_AREA)
Data: data_lagged

      AIC        BIC    logLik -2*log(L)  df.resid
6304.6     6429.2   -3125.3     6250.6       719
  
```

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.0405	-0.4830	-0.0190	0.4847	11.8870

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
REF_AREA (Intercept)		9.020e+01	9.497517	
OBS_VALUE_gdp		3.569e-05	0.005974	-0.99
OBS_VALUE_unemp		1.507e+01	3.882621	0.38 -0.38
OBS_VALUE_hdi		1.965e+02	14.018316	0.23 -0.24 0.01
OBS_VALUE_life_exp		5.730e+00	2.393691	0.67 -0.68 -0.03 0.83
Residual		1.301e+02	11.407150	

Number of obs: 746, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.750e+03	6.920e+01	39.734
mortality_lag1	1.864e-01	2.688e-02	6.936
mortality_lag2	-1.441e-01	2.678e-02	-5.382
healthExp_lag1	-3.610e-01	1.322e+00	-0.273
healthExp_lag2	3.708e-01	1.340e+00	0.277
hospBeds_lag1	-2.252e-04	1.266e-04	-1.779
hospBeds_lag2	1.830e-04	1.264e-04	1.447
OBS_VALUE_gdp	-1.396e-03	1.104e-03	-1.265
OBS_VALUE_unemp	8.273e-01	7.596e-01	1.089
OBS_VALUE_hdi	7.160e+02	7.383e+01	9.698
OBS_VALUE_life_exp	-3.939e+01	1.093e+00	-36.049

Correlation of Fixed Effects:

(Intr)	mrtl_1	mrtl_2	hltE_1	hltE_2	hspB_1	hspB_2	OBS_VALUE_g
mrtlty_lg1	-0.431						
mrtlty_lg2	-0.181	-0.664					
hlthExp_lg1	-0.098	-0.210	0.282				
hlthExp_lg2	0.229	0.103	-0.228	-0.719			
hospBds_lg1	0.028	-0.035	0.023	0.036	0.028		
hospBds_lg2	-0.051	0.028	-0.030	-0.003	-0.008	-0.972	
OBS_VALUE_g	0.105	-0.026	0.064	-0.050	-0.039	-0.005	-0.016
OBS_VALUE_n	0.050	0.006	0.027	-0.116	0.010	0.011	-0.033 -0.072
OBS_VALUE_h	-0.203	-0.108	0.171	-0.102	-0.056	-0.048	0.049 -0.254
OBS_VALUE__	-0.639	0.420	-0.038	0.132	-0.150	0.003	0.003 -0.193
			OBS_VALUE_n	OBS_VALUE_h			
			mrtlty_lg1				

```

mortlty_lg2
hlthExp_lg1
hlthExp_lg2
hospBds_lg1
hospBds_lg2
OBS_VALUE_g
OBS_VALUE_n
OBS_VALUE_h -0.046
OBS_VALUE__ -0.079      -0.502
fit warnings:
Some predictor variables are on very different scales: consider rescaling.
You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

After the first test, respecting our first model, we can see two things :

- Firstly, having a second lag for each of the explanatory variables is not as useful to our model as we could have thought. Their coefficient are not coherent compared to variables lagged only once (`healthExp_lag1` compared to `healthExp_lag2` for example). This can be explained by the fact that there is not really a direct effect of the second lag on the explained variable and that they only affect the explained variable through their relation with the first lag. Moreover the coefficients between explanatory variables lagged twice and lagged once are too high, having both is unnecessary. Therefore we can remove second lags from our regressions.
- Secondly, when assuming this many variables have random effects (here, gdp, hdi, unemployment and life expectancy), we notice there is a lot of correlation between the variables with random effects. Especially, there is a lot of correlation between each of these variables and the intercept which is random too (to take in account structural differences between countries). We can conclude it is probably not relevant to have these variables as random effects.

## Second test

After this first test, what we are going to try is to largely reduce the number of random effects, and only use one lag. The model we will use for our regression is now :

$$\begin{aligned}
Mortality_{i,t} = & \gamma_{0,i} + \beta_1 mortality_{i,t-1} + \beta_2 health.expenditure_{i,t-1} \\
& + \beta_3 Hospital.Beds_{i,t-1} + \beta_4 GDP.per.capita_{i,t} + \beta_5 Unemployment_{i,t} \\
& + \beta_6 HDI_{i,t} + \beta_7 Life.expectancy_{i,t}
\end{aligned}$$

```

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula:
OBS_VALUE_mortality ~ mortality_lag1 + healthExp_lag1 + hospBeds_lag1 +
  OBS_VALUE_gdp + OBS_VALUE_unemp + OBS_VALUE_hdi + OBS_VALUE_life_exp +
  (1 | REF_AREA)
Data: data_lagged

      AIC      BIC   logLik -2*log(L)  df.resid
6657.8    6704.5   -3318.9     6637.8       780

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1635 -0.4504 -0.0189  0.4252 11.4544

Random effects:
Groups   Name        Variance Std.Dev.
REF_AREA (Intercept) 453.0    21.28
Residual           215.4    14.68
Number of obs: 790, groups: REF_AREA, 41

Fixed effects:
            Estimate Std. Error t value
(Intercept) 2.266e+03 7.018e+01 32.291
mortality_lag1 2.449e-01 2.189e-02 11.188
healthExp_lag1 9.080e-01 7.087e-01 1.281
hospBeds_lag1 6.237e-06 9.063e-06 0.688
OBS_VALUE_gdp 1.689e-04 1.422e-04 1.188
OBS_VALUE_unemp 1.351e+00 2.335e-01 5.788
OBS_VALUE_hdi 4.119e+02 5.048e+01 8.159
OBS_VALUE_life_exp -3.108e+01 1.013e+00 -30.691

Correlation of Fixed Effects:
              (Intr) mrtly_1 hltE_1 hspB_1 OBS_VALUE_g OBS_VALUE_n OBS_VALUE_h
mortlty_lg1 -0.858
hlthExp_lg1  0.174 -0.004
hospBds_lg1  0.013 -0.003 -0.043
OBS_VALUE_g  0.166  0.000  0.011 -0.010
OBS_VALUE_n  0.043 -0.107 -0.179  0.002  0.258
OBS_VALUE_h  0.017 -0.031 -0.293  0.021 -0.304      0.237
OBS_VALUE__ -0.845  0.699 -0.058 -0.038 -0.042      -0.187      -0.536
fit warnings:
Some predictor variables are on very different scales: consider rescaling.
You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.

```

The results seem better and more coherent than in the first model. However we do notice that the AIC - Akaike information criterion, which is a measure for how well fitted a model is - is higher than it was in the previous model (6657.8 versus 6325.4). The AIC checks both if explanatory variables explain the observed variable correctly and for whether or not the number of parameters in the model is too high. Generally the lower the AIC is, the better the model is. The AIC is not absolute, for example here we do see that our second model is more coherent but it encourages us to make more tests.

### Third test

For our third test model we add non lagged explanatory variables (variables at lag = 0). We will also add social expenditures which was not in the previous regression to see if it is relevant in our model (OBS\_VALUE\_social\_expenditure and socialExp\_lag1)

Our model becomes :

$$\begin{aligned} Mortality_{i,t} = & \gamma_0,i + \beta_1 mortality_{i,t-1} + \beta_2 social.expenditure_{i,t} + \beta_3 social.expenditure_{i,t-1} \\ & + \beta_4 health.expenditure_{i,t} + \beta_5 health.expenditure_{i,t-1} + \beta_6 Hospital.Beds_{i,t} \\ & + \beta_7 Hospital.Beds_{i,t-1} + \beta_8 GDP.per.capita_{i,t} + \beta_9 Unemployment_{i,t} \\ & + \beta_{10} HDI_{i,t} + \beta_{11} Life.expectancy_{i,t} \end{aligned}$$

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: OBS_VALUE_mortality ~ mortality_lag1 + OBS_VALUE_social_expenditure +
           socialExp_lag1 + OBS_VALUE_health_exp + healthExp_lag1 +
           OBS_VALUE_hosp_beds + hospBeds_lag1 + OBS_VALUE_gdp + OBS_VALUE_unemp +
           OBS_VALUE_hdi + OBS_VALUE_life_exp + (1 | REF_AREA)
Data: data_lagged

      AIC        BIC      logLik -2*log(L)  df.resid
6495.0     6560.0    -3233.5     6467.0       758
```

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.0433	-0.4683	-0.0107	0.4330	11.7165

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	438.2	20.93
	Residual	209.4	14.47

Number of obs: 772, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.140e+03	7.276e+01	29.407
mortality_lag1	2.948e-01	2.335e-02	12.627
OBS_VALUE_social_expenditure	-7.100e-02	5.858e-01	-0.121
socialExp_lag1	1.248e-01	5.755e-01	0.217
OBS_VALUE_health_exp	6.857e+00	1.903e+00	3.603
healthExp_lag1	-4.987e+00	1.880e+00	-2.652
OBS_VALUE_hosp_beds	1.278e-04	1.499e-04	0.853
hospBeds_lag1	-1.226e-04	1.486e-04	-0.825
OBS_VALUE_gdp	2.072e-04	1.426e-04	1.453
OBS_VALUE_unemp	1.231e+00	2.517e-01	4.890
OBS_VALUE_hdi	3.860e+02	5.472e+01	7.053
OBS_VALUE_life_exp	-2.948e+01	1.057e+00	-27.890

Correlation of Fixed Effects:

	(Intr)	mrtl_1	OBS_VALUE_s_	sclE_1	OBS_VALUE_hl_	hltE_1
mortlty_lg1	-0.868					
OBS_VALUE_s_	0.003	0.049				
soclExp_lg1	-0.004	0.008	-0.750			
OBS_VALUE_hl_	-0.156	0.164	-0.649	0.542		
hlthExp_lg1	0.195	-0.161	0.536	-0.645	-0.878	
OBS_VALUE_hs_	-0.019	0.035	0.053	0.031	-0.013	-0.017
hospBds_lg1	0.020	-0.036	-0.051	-0.029	0.010	0.017
OBS_VALUE_g	0.145	0.030	0.085	0.020	0.013	-0.047
OBS_VALUE_n	0.057	-0.143	-0.120	-0.102	0.016	0.006
OBS_VALUE_h	0.031	-0.075	-0.071	-0.106	-0.081	0.035
OBS_VALUE_l_	-0.839	0.717	0.019	0.051	0.160	-0.181
	OBS_VALUE_hs_	hspB_1	OBS_VALUE_g	OBS_VALUE_n	OBS_VALUE_h	
mortlty_lg1						
OBS_VALUE_s_						
soclExp_lg1						
OBS_VALUE_hl_						
hlthExp_lg1						
OBS_VALUE_hs_						
hospBds_lg1	-0.998					
OBS_VALUE_g	0.038	-0.038				
OBS_VALUE_n	0.005		-0.006	0.186		
OBS_VALUE_h	-0.066		0.066	-0.325	0.340	
OBS_VALUE_l_	0.042		-0.044	-0.010	-0.244	-0.557
fit warnings:						

Some predictor variables are on very different scales: consider rescaling.  
 You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.

The AIC is lower in this model than it was in the second test (6495.0 vs 6657.8). However, contrary to what we could believe theoretically, social expenditures at time t and time t-1 do not have a significant impact on mortality at time t (we see that their t-value are close to 0 and not 1 in absolute value)

#### **Fourth test**

Fourth test with an additional random effect : GDP per capita (in order to try taking into account countries' richness level)

$$\begin{aligned} Mortality_{i,t} = & \gamma_0,i + \beta_1 mortality_{i,t-1} + \beta_2 social.expenditure_{i,t} + \beta_3 social.expenditure_{i,t-1} \\ & + \beta_4 health.expenditure_{i,t} + \beta_5 health.expenditure_{i,t-1} + \beta_6 Hospital.Beds_{i,t} \\ & + \beta_7 Hospital.Beds_{i,t-1} + \gamma_8 GDP.per.capita_{i,t} + \beta_9 Unemployment_{i,t} \\ & + \beta_{10} HDI_{i,t} + \beta_{11} Life.expectancy_{i,t} \end{aligned}$$

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: OBS_VALUE_mortality ~ mortality_lag1 + OBS_VALUE_social_expenditure +
  socialExp_lag1 + OBS_VALUE_health_exp + healthExp_lag1 +
  OBS_VALUE_hosp_beds + hospBeds_lag1 + OBS_VALUE_gdp + OBS_VALUE_unemp +
  OBS_VALUE_hdi + OBS_VALUE_life_exp + (1 + OBS_VALUE_gdp |      REF_AREA)
Data: data_lagged
```

AIC	BIC	logLik	-2*log(L)	df.resid
6391.7	6466.1	-3179.9	6359.7	756

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.8550	-0.4896	0.0076	0.4040	11.8947

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
REF_AREA	(Intercept)	1.679e+03	40.974794	
	OBS_VALUE_gdp	4.113e-06	0.002028	-0.88
Residual		1.632e+02	12.776543	

Number of obs: 772, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.337e+03	6.846e+01	34.139
mortality_lag1	1.875e-01	2.259e-02	8.300
OBS_VALUE_social_expenditure	1.436e-01	5.318e-01	0.270
socialExp_lag1	1.012e-01	5.272e-01	0.192
OBS_VALUE_health_exp	4.985e+00	1.721e+00	2.897
healthExp_lag1	-2.873e+00	1.740e+00	-1.651
OBS_VALUE_hosp_beds	1.292e-04	1.344e-04	0.962
hospBeds_lag1	-1.288e-04	1.328e-04	-0.970
OBS_VALUE_gdp	-9.085e-04	4.469e-04	-2.033
OBS_VALUE_unemp	6.838e-01	2.692e-01	2.540
OBS_VALUE_hdi	5.759e+02	5.797e+01	9.935
OBS_VALUE_life_exp	-3.350e+01	1.009e+00	-33.188

Correlation of Fixed Effects:

	(Intr)	mrtl_1	OBS_VALUE_s_	sclE_1	OBS_VALUE_hl_	hltE_1
mortlty_lg1	-0.817					
OBS_VALUE_s_	0.014	0.041				
soclExp_lg1	-0.029	0.005	-0.678			
OBS_VALUE_hl_	-0.172	0.183	-0.648	0.514		
hlthExp_lg1	0.204	-0.174	0.485	-0.636	-0.824	
OBS_VALUE_hs_	-0.010	0.039	0.057	0.036	-0.004	-0.006
hospBds_lg1	0.005	-0.040	-0.054	-0.031	0.004	0.011
OBS_VALUE_g	0.065	0.088	0.041	-0.058	-0.009	-0.048
OBS_VALUE_n	0.081	-0.010	-0.129	-0.144	0.036	-0.028
OBS_VALUE_h	-0.010	-0.098	-0.094	-0.105	-0.065	0.038
OBS_VALUE_l_	-0.804	0.670	0.026	0.082	0.164	-0.189
	OBS_VALUE_hs_	hspB_1	OBS_VALUE_g	OBS_VALUE_n	OBS_VALUE_h	
mortlty_lg1						
OBS_VALUE_s_						
soclExp_lg1						
OBS_VALUE_hl_						
hlthExp_lg1						
OBS_VALUE_hs_						
hospBds_lg1	-0.996					
OBS_VALUE_g	0.016	-0.033				
OBS_VALUE_n	0.004	-0.018	0.311			
OBS_VALUE_h	-0.079	0.078	-0.288	0.184		
OBS_VALUE_l_	0.041	-0.037	0.015	-0.218	-0.565	
fit warnings:						

Some predictor variables are on very different scales: consider rescaling.

You may also use `(g)lmerControl(autoscale = TRUE)` to improve numerical stability.

```

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

```

## Fifth test

Fifth test using HDI as the additional random effect (and no longer GDP per capita) :

$$\begin{aligned}
Mortality_{i,t} = & \gamma_{0,i} + \beta_1 mortality_{i,t-1} + \beta_2 social.expenditure_{i,t} + \beta_3 social.expenditure_{i,t-1} \\
& + \beta_4 health.expenditure_{i,t} + \beta_5 health.expenditure_{i,t-1} + \beta_6 Hospital.Beds_{i,t} \\
& + \beta_7 Hospital.Beds_{i,t-1} + \beta_8 GDP.per.capita_{i,t} + \beta_9 Unemployment_{i,t} \\
& + \gamma_{10} HDI_{i,t} + \beta_{11} Life.expectancy_{i,t}
\end{aligned}$$

```

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: OBS_VALUE_mortality ~ mortality_lag1 + OBS_VALUE_social_expenditure +
           socialExp_lag1 + OBS_VALUE_health_exp + healthExp_lag1 +
           OBS_VALUE_hosp_beds + hospBeds_lag1 + OBS_VALUE_gdp + OBS_VALUE_unemp +
           OBS_VALUE_hdi + OBS_VALUE_life_exp + (1 + OBS_VALUE_hdi |      REF_AREA)
Data: data_lagged

```

AIC	BIC	logLik	-2*log(L)	df.resid
6209.1	6283.5	-3088.5	6177.1	756

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5379	-0.4702	-0.0087	0.3922	12.8280

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
REF_AREA	(Intercept)	153841.5	392.23	
	OBS_VALUE_hdi	188467.1	434.13	-1.00

Residual 121.1 11.01

Number of obs: 772, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.585e+03	8.870e+01	29.140
mortality_lag1	9.396e-02	2.092e-02	4.490
OBS_VALUE_social_expenditure	6.424e-01	4.737e-01	1.356
socialExp_lag1	6.618e-02	4.764e-01	0.139

OBS_VALUE_health_exp	2.431e+00	1.518e+00	1.601
healthExp_lag1	-2.430e+00	1.566e+00	-1.552
OBS_VALUE_hosp_beds	8.580e-05	1.172e-04	0.732
hospBeds_lag1	-7.008e-05	1.159e-04	-0.605
OBS_VALUE_gdp	2.010e-05	1.676e-04	0.120
OBS_VALUE_unemp	5.812e-01	2.372e-01	2.450
OBS_VALUE_hdi	6.253e+02	9.299e+01	6.725
OBS_VALUE_life_exp	-3.705e+01	9.379e-01	-39.506

Correlation of Fixed Effects:

	(Intr)	mrtl_1	OBS_VALUE_s_	sclE_1	OBS_VALUE_hl_	hltE_1
mortlty_lg1	-0.584					
OBS_VALUE_s_	0.014	0.025				
soclExp_lg1	-0.019	0.002	-0.571			
OBS_VALUE_hl_	-0.113	0.199	-0.649	0.453		
hlthExp_lg1	0.158	-0.177	0.423	-0.646	-0.757	
OBS_VALUE_hs_	-0.012	0.034	0.060	0.048	-0.009	-0.023
hospBds_lg1	0.008	-0.035	-0.054	-0.041	0.005	0.019
OBS_VALUE_g	0.132	0.062	0.093	-0.030	0.013	-0.069
OBS_VALUE_n	-0.037	0.021	-0.176	-0.147	0.057	-0.019
OBS_VALUE_h	-0.578	0.019	-0.069	-0.069	-0.053	-0.034
OBS_VALUE_l_	-0.512	0.590	0.035	0.082	0.166	-0.145
		OBS_VALUE_hs_	hspB_1	OBS_VALUE_g	OBS_VALUE_n	OBS_VALUE_h
mortlty_lg1						
OBS_VALUE_s_						
soclExp_lg1						
OBS_VALUE_hl_						
hlthExp_lg1						
OBS_VALUE_hs_						
hospBds_lg1	-0.995					
OBS_VALUE_g	0.032	-0.032				
OBS_VALUE_n	0.004	-0.006	0.228			
OBS_VALUE_h	-0.047	0.051	-0.219	0.256		
OBS_VALUE_l_	0.051	-0.053	0.006	-0.250	-0.396	

fit warnings:

Some predictor variables are on very different scales: consider rescaling.  
 You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.

## Sixth

Sixth test using both GDP and HDI as additional random effects :

$$\begin{aligned}
Mortality_{i,t} = & \gamma_0 + \beta_1 mortality_{i,t-1} + \beta_2 social.expenditure_{i,t} + \beta_3 social.expenditure_{i,t-1} \\
& + \beta_4 health.expenditure_{i,t} + \beta_5 health.expenditure_{i,t-1} + \beta_6 Hospital.Beds_{i,t} \\
& + \beta_7 Hospital.Beds_{i,t-1} + \gamma_8 GDP.per.capita_{i,t} + \beta_9 Unemployment_{i,t} \\
& + \gamma_{10} HDI_{i,t} + \beta_{11} Life.expectancy_{i,t}
\end{aligned}$$

Linear mixed model fit by maximum likelihood ['lmerMod']  
 Formula: OBS\_VALUE\_mortality ~ mortality\_lag1 + OBS\_VALUE\_social\_expenditure +  
           socialExp\_lag1 + OBS\_VALUE\_health\_exp + healthExp\_lag1 +  
           OBS\_VALUE\_hosp\_beds + hospBeds\_lag1 + OBS\_VALUE\_gdp + OBS\_VALUE\_unemp +  
           OBS\_VALUE\_hdi + OBS\_VALUE\_life\_exp + (1 + OBS\_VALUE\_hdi +  
           OBS\_VALUE\_gdp | REF\_AREA)  
 Data: data\_lagged

AIC	BIC	logLik	-2*log(L)	df.resid
6407.7	6496.1	-3184.9	6369.7	753

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.8545	-0.4861	-0.0056	0.4184	11.7795

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
REF_AREA	(Intercept)	1.012e+03	31.807070	
	OBS_VALUE_hdi	1.225e+03	35.000608	0.72
	OBS_VALUE_gdp	9.937e-06	0.003152	-0.74 -0.95
	Residual	1.575e+02	12.548771	

Number of obs: 772, groups: REF\_AREA, 41

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.360e+03	6.785e+01	34.788
mortality_lag1	1.779e-01	2.246e-02	7.921
OBS_VALUE_social_expenditure	1.660e-01	5.281e-01	0.314
socialExp_lag1	1.358e-01	5.221e-01	0.260
OBS_VALUE_health_exp	4.869e+00	1.702e+00	2.861
healthExp_lag1	-2.827e+00	1.723e+00	-1.641
OBS_VALUE_hosp_beds	1.102e-04	1.329e-04	0.829
hospBeds_lag1	-1.149e-04	1.312e-04	-0.876
OBS_VALUE_gdp	-1.156e-03	6.216e-04	-1.859

```

OBS_VALUE_unemp          6.019e-01  2.803e-01  2.147
OBS_VALUE_hdi            6.167e+02  6.133e+01  10.055
OBS_VALUE_life_exp       -3.417e+01  1.013e+00 -33.716

Correlation of Fixed Effects:
              (Intr) mrtl_1 OBS_VALUE_s_ sclE_1 OBS_VALUE_hl_ hltE_1
mortlty_lg1      -0.808
OBS_VALUE_s_     0.020  0.041
soclExp_lg1     -0.030  0.001 -0.656
OBS_VALUE_hl_   -0.174  0.182 -0.647      0.505
hlthExp_lg1     0.201 -0.175  0.472     -0.632 -0.815
OBS_VALUE_hs_  -0.012  0.040  0.056      0.036 -0.003      -0.005
hospBds_lg1    0.004 -0.040 -0.053     -0.029  0.005      0.013
OBS_VALUE_g     0.064  0.083  0.046     -0.056 -0.010      -0.040
OBS_VALUE_n     0.086  0.015 -0.122     -0.156  0.038      -0.035
OBS_VALUE_h     -0.019 -0.100 -0.104     -0.098 -0.064      0.035
OBS_VALUE_l_   -0.791  0.657  0.026      0.082  0.166     -0.182
                  OBS_VALUE_hs_ hospB_1 OBS_VALUE_g OBS_VALUE_n OBS_VALUE_h
mortlty_lg1
OBS_VALUE_s_
soclExp_lg1
OBS_VALUE_hl_
hlthExp_lg1
OBS_VALUE_hs_
hospBds_lg1   -0.995
OBS_VALUE_g    0.016      -0.031
OBS_VALUE_n    0.005      -0.024  0.290
OBS_VALUE_h    -0.083      0.079 -0.344      0.126
OBS_VALUE_l_   0.046      -0.040  0.031     -0.202     -0.571
fit warnings:
Some predictor variables are on very different scales: consider rescaling.
You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

Tests four to six helped us notice that :

- GDP and HDI were not varying enough to be random effects. Since we already have a moving intercept and there is not much of a difference between having a moving intercept from one country to another and having a random effect for GDP and HDI per country) Particularly having a random effect for HDI is not great since it has a perfect correlation (-1) with the intercept, meaning that having HDI as a random effect would be useless. We

see the same thing with GDP but at a lower level since the correlation is slightly less high.

- Health expenditure at present time and t-1 are relevant in the regression
- Social expenditure not so much (they still have a t value closer to 0 than 1 in absolute value). However it is interesting to keep this variable in our model since it is different from what we were expecting
- The Hospital beds variable is barely relevant (0.8)
- Control variables are really relevant (especially life expectancy)

We also notice that the fifth test (which is not that great since it has correlation problems) has the lowest AIC. We were able to find a lower AIC but it ended up not being a coherent model. Consequently we will have to compromise by choosing a more coherent model with a slightly higher AIC.

### **Chosen model**

The best model we were able to find ended up being our third test.

So our chosen model is :

$$\begin{aligned} Mortality_{i,t} = & \gamma_0,i + \beta_1 mortality_{i,t-1} + \beta_2 social.expenditure_{i,t} + \beta_3 social.expenditure_{i,t-1} \\ & + \beta_4 health.expenditure_{i,t} + \beta_5 health.expenditure_{i,t-1} + \beta_6 Hospital.Beds_{i,t} \\ & + \beta_7 Hospital.Beds_{i,t-1} + \beta_8 GDP.per.capita_{i,t} + \beta_9 Unemployment_{i,t} \\ & + \beta_{10} HDI_{i,t} + \beta_{11} Life.expectancy_{i,t} \end{aligned}$$

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: OBS_VALUE_mortality ~ mortality_lag1 + OBS_VALUE_social_expenditure +
  socialExp_lag1 + OBS_VALUE_health_exp + healthExp_lag1 +
  OBS_VALUE_hosp_beds + hospBeds_lag1 + OBS_VALUE_gdp + OBS_VALUE_unemp +
  OBS_VALUE_hdi + OBS_VALUE_life_exp + (1 | REF_AREA)
Data: data_lagged
```

AIC	BIC	logLik	-2*log(L)	df.resid
6495.0	6560.0	-3233.5	6467.0	758

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.0433	-0.4683	-0.0107	0.4330	11.7165

Random effects:

Groups	Name	Variance	Std.Dev.
REF_AREA	(Intercept)	438.2	20.93
	Residual	209.4	14.47
Number of obs: 772, groups: REF_AREA, 41			

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.140e+03	7.276e+01	29.407
mortality_lag1	2.948e-01	2.335e-02	12.627
OBS_VALUE_social_expenditure	-7.100e-02	5.858e-01	-0.121
socialExp_lag1	1.248e-01	5.755e-01	0.217
OBS_VALUE_health_exp	6.857e+00	1.903e+00	3.603
healthExp_lag1	-4.987e+00	1.880e+00	-2.652
OBS_VALUE_hosp_beds	1.278e-04	1.499e-04	0.853
hospBeds_lag1	-1.226e-04	1.486e-04	-0.825
OBS_VALUE_gdp	2.072e-04	1.426e-04	1.453
OBS_VALUE_unemp	1.231e+00	2.517e-01	4.890
OBS_VALUE_hdi	3.860e+02	5.472e+01	7.053
OBS_VALUE_life_exp	-2.948e+01	1.057e+00	-27.890

Correlation of Fixed Effects:

	(Intr)	mrtl_1	OBS_VALUE_s_	sclE_1	OBS_VALUE_hl_	hltE_1
mortlty_lg1	-0.868					
OBS_VALUE_s_	0.003	0.049				
soclExp_lg1	-0.004	0.008	-0.750			
OBS_VALUE_hl_	-0.156	0.164	-0.649	0.542		
hlthExp_lg1	0.195	-0.161	0.536	-0.645	-0.878	
OBS_VALUE_hs_	-0.019	0.035	0.053	0.031	-0.013	-0.017
hospBds_lg1	0.020	-0.036	-0.051	-0.029	0.010	0.017
OBS_VALUE_g	0.145	0.030	0.085	0.020	0.013	-0.047
OBS_VALUE_n	0.057	-0.143	-0.120	-0.102	0.016	0.006
OBS_VALUE_h	0.031	-0.075	-0.071	-0.106	-0.081	0.035
OBS_VALUE_l_	-0.839	0.717	0.019	0.051	0.160	-0.181
	OBS_VALUE_hs_	hspB_1	OBS_VALUE_g	OBS_VALUE_n	OBS_VALUE_h	
mortlty_lg1						
OBS_VALUE_s_						
soclExp_lg1						
OBS_VALUE_hl_						
hlthExp_lg1						
OBS_VALUE_hs_						
hospBds_lg1	-0.998					
OBS_VALUE_g	0.038		-0.038			
OBS_VALUE_n	0.005		-0.006	0.186		

```
OBS_VALUE_h -0.066      0.066 -0.325      0.340
OBS_VALUE_l_  0.042      -0.044 -0.010      -0.244      -0.557
fit warnings:
Some predictor variables are on very different scales: consider rescaling.
You may also use (g)lmerControl(autoscale = TRUE) to improve numerical stability.
```