



The predictability of COVID-19 mortality rates based on *ex-ante* economic, health and social indicators

EMESE KOVÁCS*  and PÉTER MIHÁLYI 

Institute of Economics, Corvinus University of Budapest, Budapest, Fővám tér 8., H-1093, Hungary

Received: July 15, 2021 • Revised manuscript received: August 3, 2021 • Accepted: August 15, 2021

© 2021 The Author(s)



ABSTRACT

The paper analyses the differences of COVID-19 mortality rates (MR) in 24 European countries. We explain MRs on the available, reliable *ex-ante* economic, health and social indicators pertaining to the year 2019 – i.e., before the outbreak of the pandemic. Using simple regression equations, we received statistically significant results for 11 such variables out of 28 attempts. Our best model with two *ex-ante* independent variables explains 0.76 of the variability of our *ex-post* dependent variable, the logarithm of Cumulative COVID Deaths. The estimated coefficient for the variable *Density of Nurses* shows that having one more nurse per 1,000 of population decreases cumulative COVID deaths by almost 15%. Similarly, one more unit *Consumption of Non-Prescribed Medicine* decreases cumulative deaths by 5%. It seems that until now those European countries were successful in minimising the fatalities where the population had a high level of health literacy, people pursue healthier lifestyle and the healthcare systems worked with a relatively large nursing force already prior to the COVID pandemic.

KEYWORDS

COVID-19, pandemic, mortality rate, *ex-post* and *ex-ante* variables

JEL CLASSIFICATION INDICES

I12, I15, I18

* Corresponding author. E-mail: emeseblacksmith@gmail.com

1. INTRODUCTION

We know three things for sure about the pandemic unfolding before our eyes since early 2020, and there is one *known unknown* factor. What we can be certain of is that:

- (i) the coronavirus syndrome is an infectious disease caused by a newly discovered type of the coronavirus family. Most people infected with the virus experience mild to moderate respiratory illness and recover without requiring any special treatment. But in some cases, the disease can cause severe medical complications and lead to death.
- (ii) medical science is now much better prepared than ever before to prevent and treat a pandemic like COVID, and as a result
- (iii) this disease causes significantly fewer deaths, both in absolute numbers and relative to the population than similar health cataclysms of the past centuries.

What we do not understand, what we do not know, is:

- (iv) why does the spread of the disease through human-to-human transmission, which potentially threatens the entire human race evenly, differs so considerably among the countries.

The present paper focuses on this last question from a novel, econometric approach. We shall refer to this as “the black box approach”. Our country sample reflects the limitations of data availability on both sides of the regression analysis (two dependent and 28 independent variables), as well as our assumptions about the cross-country comparability concerning commonly used definitions and reliable reporting. We selected 16 West European countries and 8 European transition economies (24 in total).

Looking back from a 21st century perspective, we can state without any doubt that the history of humankind was a history of communicable diseases. On the one hand, low-intensity infections have always been present in all countries (e.g., seasonal influenza, malaria), but on the other hand, epidemics have flared up dramatically from time to time and in some places, causing the deaths of thousands and millions. In this regard, we refer here to just two important historical examples. The bubonic plague of 1,347–1,351 killed 30–40% of the population of Western Europe, and in England the figure is thought to have been as high as 70%, while Eastern Europe largely escaped the pandemic. The death toll is estimated at between 50 and 200 million.¹

As widely known, the last major pandemic to affect humanity as a whole was the 1918–1920 influenza (known at that time as the Spanish flu), when the H1N1 influenza virus was estimated to have infected 1/3 of the total human population in three years with a death toll of 50–100 million, or 2–4% of the world’s population (Taubenberger – Morens 2006). The available estimates for individual countries are, not surprisingly, sporadic and not comparable directly. In the United States, the Spanish flu killed 1% of those infected (case fatality rate, CFR = 1%)²,

¹Plague is transmitted between rodents by rodent fleas and can be transmitted to people when infected rodent fleas bite them. As with many zoonotic diseases, plague is a very severe disease in people, with case fatality rates (CFR) of 50–60% if left untreated. Today, bubonic plague is entirely curable with antibiotics, but preventive vaccination has not yet been developed. There are about 2000 cases of the disease worldwide every year, with Madagascar being the most affected country (WHO 2000).

²Case fatality rate (CFR) = Proportion of people who die from a specified disease among all individuals diagnosed with the disease over a certain period of time.



with a total of 675,000 deaths. In Ghana, 5% of the population died in the first two months after the outbreak, in Samoa in the Polynesian archipelago 20% of the population died, etc.

So far, the present coronavirus pandemic has caused 4–10 million deaths in 18 months,³ a 0.05% loss of the world's population. This is two orders of magnitude less than the proportion the world experienced during the Spanish flu or the apocalyptic attacks of the bubonic plague bacteria (Table 1, column 5).

The etiology of the plague was not understood until the general acceptance of the germ theory in the 19th century. From this, scientists established that the cause of the striking disparities in loss of life between the two parts of Europe was straightforward: the germs were carried from Asia to Europe through sea-trade, thus the landlocked part of Europe was fortunate to be spared from the deadly communicable *Bacteria yersinia pestis*. The actual cause of the Spanish flu – the H1N1 virus that caused the disease – was not identified until 1933; the proven safe vaccine against influenza was not available until 1945. By contrast, the responsible pathogen for COVID, the severe acute respiratory syndrome virus (SARS-CoV-2) was isolated in China literally within days of the first clinically diagnosed cases and it took less than 12 months to develop different types of the first generation of effective vaccines in the US, the UK, China and Russia.

According to the Worldometer database, which is updated daily, in July 2021, cumulative mortality per 100 people (i.e., COVID-related deaths as a percentage of the total population) was highest in the world in Peru and Hungary: 0.6 and 0.3, respectively. At that time, these two countries topped the "list of shame".⁴ The United States came 21st with 0.2 and Sweden 36th with 0.1. India, with 1.4 billion inhabitants, came 107th with 0.03 per 100 inhabitants. In China, which is roughly the same size as India, ranked 200th on the list, there were only 3 (!) deaths per 1 million inhabitants – this figure does not appear to be reliable.

Unfortunately, there is a false perception of the pandemic in the public opinion worldwide. The media has been bombarding the people from the very beginning with absolute death figures, such as "4 million total deaths in the world" or "130 thousand deaths in the UK", without emphasising that the population size of different countries varies enormously. E.g., the Hungarian public was overwhelmed with the high absolute number of COVID-deaths in Italy and the UK, although proportionally the number of fatalities was bigger in Hungary than in the other two countries. Not surprisingly, this false perception distorts the thinking horizon of policy experts as well, because they must take into consideration what people think – even if these thoughts are ill-founded.

It is a matter of ethical and/or political perspective whether the disparities in MRs among these above-mentioned countries and their deviation from the world average (0.05%) is considered *dramatically significant* or *relatively insignificant*. There are legitimate and sensible arguments for both positions. One can say that every death matters. But one can also defend a position by saying that on per capita basis, small and changing country differences should not be allowed to influence the strategies of the public health system. Furthermore, there is also a

³The lower figure is the registered number, the higher figure is an expert estimate. On 21 May 2021, the WHO estimated that the proper number would be two to three times higher than the Worldometer's calculation based on officially reported national data.

⁴It is worth noting that among the top 20 countries of this shame-list, 10 countries can be categorised as post-communist economies (Bosnia-Herzegovina, Bulgaria, Croatia, Czechia, Hungary, Montenegro, North Macedonia, Slovakia, Slovenia, and Poland). It is, however, beyond the scope of the present paper to dig deeper into the causes of this striking phenomenon.



Table 1. Major world epidemics

Year/Period (1)	Name of the disease and the pathogen (2)	Estimated total number of fatalities in millions (3)	Percentage of the affected population (Mortality rate) (4)	Percentage of the Earth's population (Mortality rate) (5)	Case fatality rate (CFR) (6)
B.C. 430-425	Bubonic plague - Bacteria yersinia pestis	<0.1	In Athens: 25	n. a.	50-60
C. E. 541		30	In Constantinople: 20-40	10	
C. E. 1346-1353 (9 major waves)		75-200	30-60 of affected countries in Asia, Africa, and Western Europe, 70 of England	20-50	
1918-1920	Influenza (Spanish flu) - H1N1 virus	17-100		1.0-5.4	>2.5
1957-1958	Asian influenza - H2N2 virus	2		0.07	0.2
2020-2021	COVID-19 - SARS-2 COV virus	4.2-8.5		0.04-0.1	3.5
Current infections, million/year					
Latest year before 2020	Tuberculosis	1.5		0.02	
	Hepatitis B and C	1.3		0.02	
	AIDS - HIV virus	0.5-1.0		0.007-0.013	
	Seasonal flu - Influenza A, B, C and D virus	0.6-0.7		0.008-0.01	
	Malaria	0.4		0.005	

Source: Authors' compilation.

logical explanation what the leader of the Swedish anti-COVID public health policy said at the very outset, namely that within 1-2 years, the infection rates would be roughly the same in all countries of the world and herd immunity would be eventually reached at about the same time everywhere.⁵ This may be the result soon, but so far, the statistics have not borne out this assumption. In the eyes of the public, the total rounded number of COVID deaths in Sweden (15

⁵See the interview with Anders Tegnell in *Nature*, published on 21 April 2020. <https://www.nature.com/articles/d41586-020-01098-x>. For a more broadly supported similar view in the scientific community see the so-called *Great Barrington Declaration* (2020).



thousand), Hungary (30 thousand), and in Israel (6,500) signifies massive disparities, given the trivial fact that these three countries are roughly equal in population size.

2. MEASURING THE DISEASE BURDEN – DEPENDENT VARIABLES

The present paper puts the COVID-related official death numbers into focus by emphasising that more than 90% of the registered worldwide coronavirus cases end with recovery.⁶ What really matters from a public health perspective is the loss of life, not the illness itself. There are several methods to measure the death burden. One way is to look at the CFR as we did in column 6 of Table 1 above for the comparison of different epidemics in history and as Pažitný et al. (2021) show in the present Special Issue of this journal, using CFR as the sole dependent variable of their econometric model. We built our database from two other country-by-country metrics as *ex-post dependent variables*:

- 1: Excess death as reported by Eurostat and
- 2: Cumulated death, as reported by Worldometer.

1. The excess death indicator was created by Eurostat in April 2020 with the help of national statistical institutes. The number of deaths from all causes is compared with the expected number of deaths extrapolated from data of a certain period in the past. The indicator is expressed as the percentage rate of additional deaths in a month, compared to a base period (2016–2019). The higher the value, the more additional deaths have occurred compared to the base, on the contrary, a negative value means that fewer deaths happened in a month under reviewed compared to the base period (see Appendix).
2. Cumulated death is reported by Worldometer. It shows the total number of detected COVID-related death cases between the beginning of January 2020 and July 16, 2021 (our cut-off date) per 1 million population.⁷

To make the two variables directly comparable, the excess death data needed to be aggregated across time – i.e., across 16 months, because of the *tremendous volatility* of the monthly time series in many countries of the sample.⁸ As a quick illustration of our findings, let us take the case of Poland, where the excess death in November 2020 was 97% higher than in the base period – the largest figure in the table. Or look at Slovenia, where in November 2020, excess mortality was 91% above the 2016–2019 base period, while three months later, in February 2021 the number of total deaths was 1 per cent below the base period average.

In column 2 of Table 2, the above-mentioned aggregation was done already. The data are presented in alphabetic order of the countries to ease the readers' navigation. It is undeniable

⁶For the world as a whole, 173 million out of 190 million people – to be precise.

⁷Worldometer (formerly Worldometers), is a reference website that provides counters and real-time statistics for diverse topics. It is owned and operated by a data company Dadax. In 2020, the website attained popularity due to hosting statistics relating to the COVID pandemic. The underlying cumulated death data used in the present paper are almost identical with another frequently used database, *Our World in Data* compiled from the daily data set provided by Johns Hopkins University. <https://www.worldometers.info/coronavirus/>.

⁸This is visible from the standard deviation data presented in the last column of the Appendix.



that the differences among the countries are significant. The worst results belong to Poland and the Czech Republic, in which until May 2021, 23% more people died from all causes than in the base period of 2016–2019. The best figures were reported from Norway, in which not *more* but 2% *less* people died during the COVID-crisis compared to the base period. Column 3 shows the total number of COVID-caused death cases per 1 million population. In our sample, the highest

Table 2. Dependent variables

Countries [1]	Arithmetic average of monthly excess death expressed in percentage [2]	Total COVID-19 death per 1 million of population [3]	Ranking col. [2] [4]	Ranking col. [3] [5]
Austria	9	1,184	14	17
Belgium	11	2,165	10	4
Czechia	23	2,827	2	2
Denmark	1	437	23	21
Estonia	8	957	17	20
Finland	2	176	21	22
France	10	1,703	11	9
Germany	5	1,093	19	18
Greece	9	1,234	15	16
Hungary	13	3,115	7	1
Iceland	2	87	22	24
Italy	14	2,118	6	6
Latvia	5	1,363	20	13
Lithuania	10	1,642	13	11
Luxembourg	8	1,287	16	14
Netherlands	10	1,035	11	19
Norway	-2	146	24	23
Poland	23	1,989	1	7
Portugal	12	1,691	8	10
Slovakia	21	2,293	3	3
Slovenia	16	2,128	4	5
Spain	15	1,734	5	8
Sweden	6	1,437	18	12
Switzerland	12	1,250	9	15
Spearman correlation				0.84

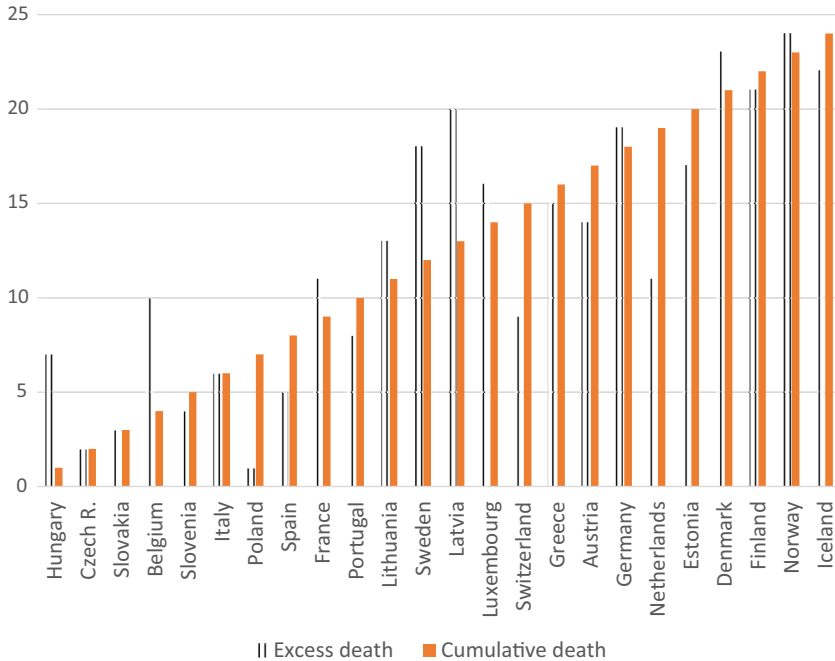


Fig. 1. The comparison of two ex-post dependent variables (Rankings: 1-24)

number was reported from Hungary (a land-locked, middle-size country in Central Europe) and the lowest one from Iceland (essentially comprising from a remote, huge main island at the juncture of the North Atlantic and Arctic Oceans).

In columns 4 and 5 the absolute figures were converted into ranking, which allow us to compare the differences between the main messages of the two underlying data series. The Spearman correlation analysis confirms the readers' first visual inspection of Fig. 1: The two series are strongly correlating ($\rho = 0.84$). Being assured by the similarity of the two dependent variable time series, in the remaining part of the paper only the total COVID death number (Table 2, column 2) will be used.

3. GETTING AROUND THE BLACK BOX

The starting point of our research hypothesis was that COVID deaths per country are ultimately causally related to two groups of variables. First, we know the economic, health and social conditions in each country before the pandemic. We referred to all these data as *ex-ante independent* variables. Without exception, we used explanatory data for the year 2019 or before.

On the other hand, the measures implemented in each country to contain the transmission of the virus and to treat and save the lives of people who became ill obviously matter a lot, most likely more than the *ex-ante* variables. Without much intellectual effort, even average newspaper readers can think of hundreds of *ex-post* variables, where the *ex-post* restriction is meant to indicate that only what happened during the pandemic (i.e., after 1 January 2020) matters in the

explanation of the dependent variable(s). Schematically, our modelling alternatives look like this below:

<p>IDEAL MODEL: Ex-ante independent variables + Ex-post independent variables → Ex-post dependent variable(s) MEANINGFUL LINE OF THOUGHT, BUT NOT PERTINENT IN THE PRESENT RESEARCH: Ex-ante independent variables → Ex-post independent variable(s) MODEL APPLIED IN THE PRESENT PAPER: Ex-ante independent variables → Ex-post dependent variable(s)</p>

However, in our attempt to collect *ex-post* independent variables, we ended up with a confound of abundance. In retrospect, this was inevitable.⁹ All countries of the world, including, of course, the countries selected in our sample, have applied a vast range of non-pharmaceutical interventions, but not at the same time and not with the same stringency. The most obvious example is the curfew, which was applied to some extent by most countries, but in some cases starting at 8 p.m., in others only at 10 p.m. or midnight. Some countries that hermetically sealed their borders for long, but others did so only selectively and from time to time. Even the recommended space of social distancing (1.5 m or 2.0 m) or the choice of face masks (cloth mask, surgical mask or N95 mask) varied from country to country and from time to time. In many countries, such as Germany and the United Kingdom, the regulations varied in many dimensions from one federal state to another and from one county to another, so that it is not possible to find representative variables for a single country.¹⁰

The time elapsed since the beginning of the pandemic has been far from homogenic in at least three dimensions.¹¹ Firstly, the chances of infection, hospitalization and ultimately death were fundamentally different before and after the start of mass *vaccination* in each European country. Secondly, the hard fact is that there are currently no effective drugs targeting SARS-CoV-2 directly, specifically and effectively. As a second-best alternative, the research community and hundreds of clinicians worldwide are falling back on the existing repertoire of approved drugs to probe into their anti-coronavirus properties. Such practices, however, also vary within the countries. Thirdly, the SARS-Cov-2 virus, which causes the disease, cannot be considered as a single biological entity. As was quickly recognised, the virus, first identified in China, mutated and is still mutating at a dizzying rate. A good example is what happened on the Diamond Princess, a ship carrying mostly Japanese tourists in February 2020. One person carried the infection on board, where 400 of the 3,711 passengers and crew showed symptomatic indications (9.7%), and 9 of them died (2.4%), while at least 24 new mutations developed (Yeah – Contreras 2021).

⁹Baum et al. (2021) have come to the same conclusion, when they tried to use the Global Health Security Index and the Epidemic Preparedness Index both of which had been computed a few months *before* the COVID outbreak to measure the preparedness of 195 countries of the world for epidemics or pandemics. As it turned out, while these studies correctly showed that the world as a whole was not prepared to combat “a fast-spreading respiratory disease”, both studies failed to predict national COVID morbidity and mortality figures. In their *ex-post* study, the authors identified 10 factors which seemed to have contributed to this failure. Some of these factors will be shown in our study as well.

¹⁰Within Japan, a large regional disparity in COVID mortality was observed. The ratio of mortality rates in the most and least affected territorial units was 83 to 1 during the first wave of the pandemic (Osaki et al. 2021).

¹¹WHO declared the outbreak of pandemic on 11 March 2020.



At time of this paper's cut-off date (16 July 2021), the Delta variant in the US went from making up just 2% of the cases as recently as mid-May 2021 to 80% by mid-July of the same year. The UK experienced a similar, earlier rise, and it now accounts for over 99% of the analysed cases there. In Brazil, the Gamma variant, which was first identified there, accounted for 96% of the analysed cases. In Chile, in addition to Gamma, the Lambda variant made up a significant share of analysed cases, etc. By our cut-off date, the Alpha variant – the original Wuhan virus – has been practically eradicated in most countries of the Earth.¹²

In Fig. 2 below, these complicating matters are illustrated in more detail. The arrows show the “impossible strategy”: to explain the variation of the dependent variables with the help of the ex-post independent variables (thick arrows), as well as “the getting around the black box” strategy, where the possible logical links leading from the ex-ante independent variables to the dependent variables are shown by the thinner arrows.

Taking into consideration of the above complications, we did not even try to identify with econometric tools *directly* the causality links between ex-post independent variables and our two dependent variables at a country-by-country level due to the vastly inhomogeneous content of the variables, such as the curfew (already mentioned) or the closing of schools, restaurants, shops, etc. When we decided to “get around the black box”, we have committed ourselves to *indirect* estimations. Let us consider two examples. It makes sense to suppose that the geographical characteristics of the countries (i.e., being landlocked or accessible through maritime transport means) do matter for the spread of the coronavirus. In the same way, it makes sense to assume that the development of the healthcare sector (e.g., number of doctors) also matters. But there is a price to pay. We were convinced from the very beginning that our *simple regressions* will be capable in this indirect way, in the best case, explain a small part of the variation of the COVID-death disparities among 24 countries. In other words, even if we find statistically significant explanatory variables with ordinary least squares (OLS) regressions calculated always by one independent variable a time, it is unlikely that our regression equations will yield high R^2 values. But as John Maynard Keynes once allegedly quipped, “it is better to be roughly right than precisely wrong”.

4. TESTING 28 INDEPENDENT VARIABLES ONE-BY-ONE

Altogether 28 independent variables were tested for 24 European countries. All data were taken from Eurostat, reflecting the state of affairs in 2019 (i.e., before the COVID pandemic) or the latest year before that. More than half of them (17 variables to be precise) did not prove to be significant from the perspective of our preferred “raw” dependent variable (see column 3 in Table 2), which was converted to its logarithmic value as is often the case in linear regression analysis.¹³

Taking the logarithm was justified for multiple reasons. The distribution of Cumulative Deaths is not normal (see the histogram in Fig. 3 below). Getting closer to normal distribution was achieved by the logarithmic transformation, which increased the reliability of our estimates.

¹²Our World in Data: Biweekly digest, 16 July 2021.

¹³In this regard, we followed the technique applied in Osaki et al. (2011).



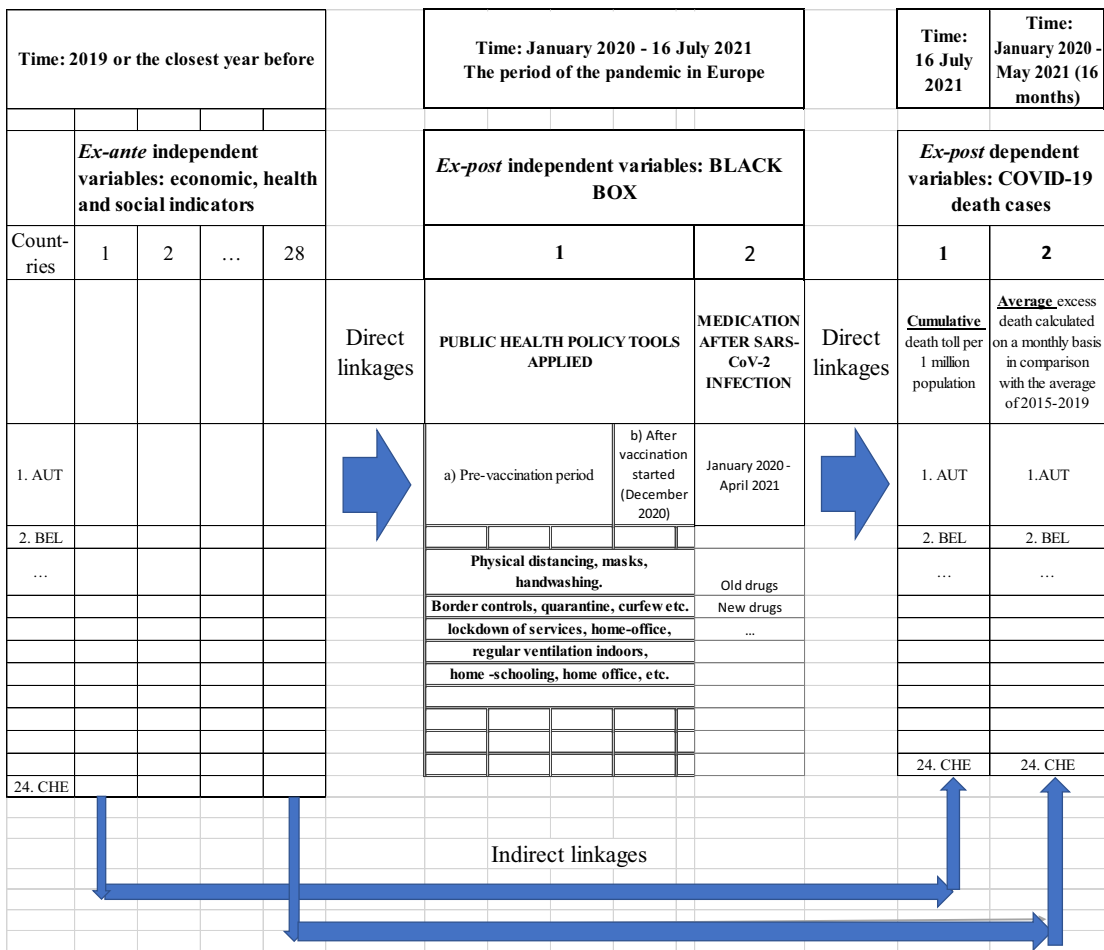


Fig. 2.

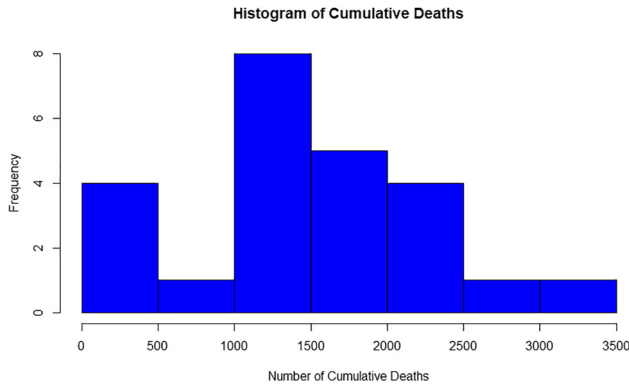


Fig. 3.

By taking the logarithm, the numbers shrunk in absolute value, closer to the magnitudes of our independent variables. In addition, this procedure can eliminate heteroscedasticity. As a result, the interpretation of the estimated coefficients of the independent variables changed to per cent.

The general form of our ordinary least squares (OLS) model is as follows:

$$\log(Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_q X_q \quad q \geq 0$$

where Y is the dependent, β_0 is the intercept, β_1 is the estimated coefficient of variable X_1 .

The list of the tested variables and the first results of the variable-by-variable regression analyses are presented in Table 3.

In selecting independent variables, special attention was paid to pick indicators associated with chronic illnesses, because clinical experience in the European countries showed that comorbidities such as hypertension increase the likelihood of fatal COVID outcomes. It is also known that the severity of COVID is higher among smokers, obese men and women; living in nursing homes or staying in hospital for a long time for any reason increases the chances of COVID transmission. This linkage – at least in Europe – appears to be so strong that the *Density of Hospital Beds* (Variable 6) does not appear to have a negative association with COVID deaths, as we originally thought. The abundance of hospital beds seems to be the consequence of the aging population and the high share of people with chronic diseases.

The most significant variable was *Consumption of Non-Prescribed Medicine*¹⁴(28), the R^2 of which was 0.58 (meaning that the simple regression equation explains 58% of the dependent variable's variation). The independent variable with the second-highest explanatory power was *Regular Exercising* (24) with 0.51. Furthermore, two other healthy lifestyle-related variables had R^2 's above 0.2: *Consumption of Prescribed Medicine*¹⁵ (27) (0.23) and *Alcohol Consumption* (23) (0.27). The high R^2 values of these four variables reflect the increased health consciousness of a considerable proportion of the European population – people with higher education and higher incomes. So, a health-conscientious life pre-COVID was not in vain. At the same time, variables

¹⁴Annual self-reported use expressed in per cent of population.

¹⁵Annual self-reported use expressed in per cent of population.



Table 3. The pair-wise relationships between ex-ante independent variables and the logarithm of COVID Cumulative Deaths in 24 European countries

		Data availability	Direction of suspected causality	Significance at 5% level	R^2	Adjusted R^2	Estimated coefficient
I.	Economic development level						
1	GDP per head	24 countries	Negative	xx	0.24	0.2	-1.815e-05
2	Housing density	24 countries	Positive				
3	Health expenditures per capita	24 countries	Negative				
4	Density of physicians	24 countries	Negative				
5	Density of nurses	24 countries	Negative	xx	0.37	0.34	-0.14
6	Density of hospital beds	24 countries	Negative	xx	0.20	0.17	0.25*
7	Density of long-term beds #	21 countries	Positive				
8	Density of old-age homes #	23 countries	Positive				
9	Eurohealth Consumer Index	24 countries	Negative	xx	0.23	0.19	-0.004
II.	Geography and history						
10	Population density	24 countries	Positive				
11	Landlocked or not	24 countries	Positive (0,1)				
12	Post-socialist country	24 countries	Positive (0,1)	xx	0.17	0.14	0.8
13	Trust in government	24 countries	Negative	xx	0.27	0.24	-0.03
III.	Health of the population						
14	Male life expectancy at 65	24 countries	Negative				
15	Female life expectancy at 65	24 countries	Negative				
16	Median age of the population #	24 countries	Positive	xx	0.27	0.24	0.2

(continued)

Table 3. Continued

		Data availability	Direction of suspected causality	Significance at 5% level	R^2	Adjusted R^2	Estimated coefficient	
	17	Self-percieved health	24 countries	Negative				
	18	Long-standing illness #	24 countries	Positive				
	19	High blood pressure #	24 countries	Positive				
	20	Suicide	24 countries	Positive				
IV.		Healthy life-style	24 countries					
	21	Obesity #	24 countries	Positive				
	22	Smoking #	24 countries	Positive				
	23	Alcohol consumption #	24 countries	Positive	xx	0.27	0.24	0.27
	24	Regular exercising	21 countries	Negative	xx	0.51	0.48	-0.07
	25	Fruit consumption	23 countries	Negative				
	26	Vegetable consumption	23 countries	Negative				
	27	Consumption of prescribed medicine #	19 countries	Positive	xx	0.23	0.18	0.04
	28	Consumption of non-prescribed medicine	19 countries	Negative	xx	0.58	0.55	-0.05

Notes: *: Opposite to expectation, #: Strong association with chronic diseases.

Source: Authors' collection of data predominantly from Eurostat publications.



Obesity (21) and *Smoking* (22) proved to be unimportant in explaining the differences among the country variation of COVID's death burden. Simply said, it seems that among the highly developed sample countries, there is no divergence: less and less people smoke, and more and more people are obese everywhere.¹⁶ This logic, however, does not apply to the *Median Age* variable ($R^2 = 0.27$), which is a demographic reality and not a matter of conscientious health behaviour. What we found here is in line with the dominant view in the literature: as age increases, so does the probability of COVID mortality.

It was perhaps even more surprising that the variables, such as *Density of Physicians* (4) and *Health Expenditure per Capita* (3), turned out to be statistically insignificant. If these results are at least "roughly right" – to be in line with Keynes – spending more money on health and healthcare were not the deciding factor when the so-far unknown virus arrived at the borders of our sample countries.¹⁷ The economic development level reflected in the customary *GDP per Head* (1) figures were significant, but its estimated coefficients were negligibly small.

One of the probable lessons of the present pandemic is that the availability of trained nurses in a country's healthcare system is beneficial both in "normal" times and during a pandemic. It improves the comfort of hospitalized patients and alleviates the burden on the shoulders of the overburdened clinicians. The high R^2 value (0.37) of the *Density of Nurses* (5) variable thus can be explained.

The two most important independent variables were also plotted against our dependent variable on two separate scatter graphs to see whether the countries of the study form distinct groups in the upper left and lower right quadrants as separated by the median values of the two selected variables. And this seems to be the case. In Fig. 4 – for example – the quadrant division clearly confirms our linear OLS-regression results above. Health-conscientious countries (here represented by *Non-Prescribed Medicine Consumption* (28)) performed better in terms of COVID deaths. Out of 19 countries, 7 unequivocally fell into the upper left quadrant (these are the poorly performing countries) while another 7 falls into the lower right quadrant (the countries with relatively low COVID death figures). Four Nordic countries – Iceland, Denmark, Norway and Finland – have squarely outperformed the rest of the sample. It sounds meaningful to conclude that the relatively high level of the use of non-prescribed medicines is a good indicator of health literacy, healthy lifestyle (but also higher incomes). A very similar message can be read from Fig. 5, where the relationship between COVID-related deaths and the *Regular Exercising* (24) variable is shown.

5. TOWARDS BUILDING A COMBINED MODEL WITH MULTIPLE REGRESSIONS

The best model had an adjusted- R^2 of 0.76 with two *ex-ante* independent variables, both of which are significant at 5%: *Non-Prescribed Medicine Use* and *Number of Nurses*. Two other models were also found to hold considerable explanatory powers but only at the 10%

¹⁶It is noteworthy that according to our preliminary (unpublished) analysis of a worldwide sample of 109 countries obesity and smoking are two out of the three most significant variables.

¹⁷In fact, the Baum et al. (2021) study already arrived at the same conclusion.



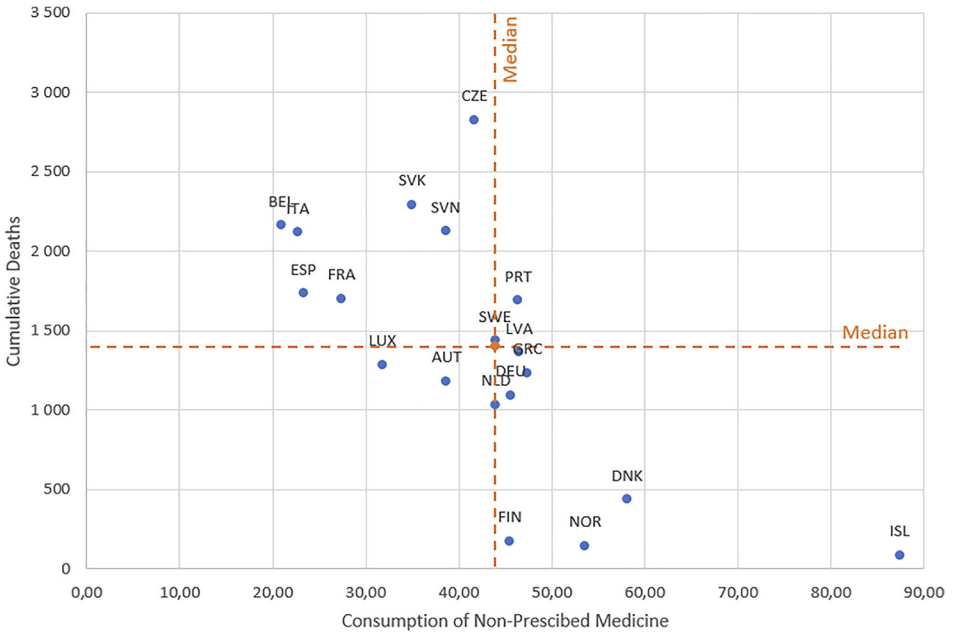


Fig. 4.

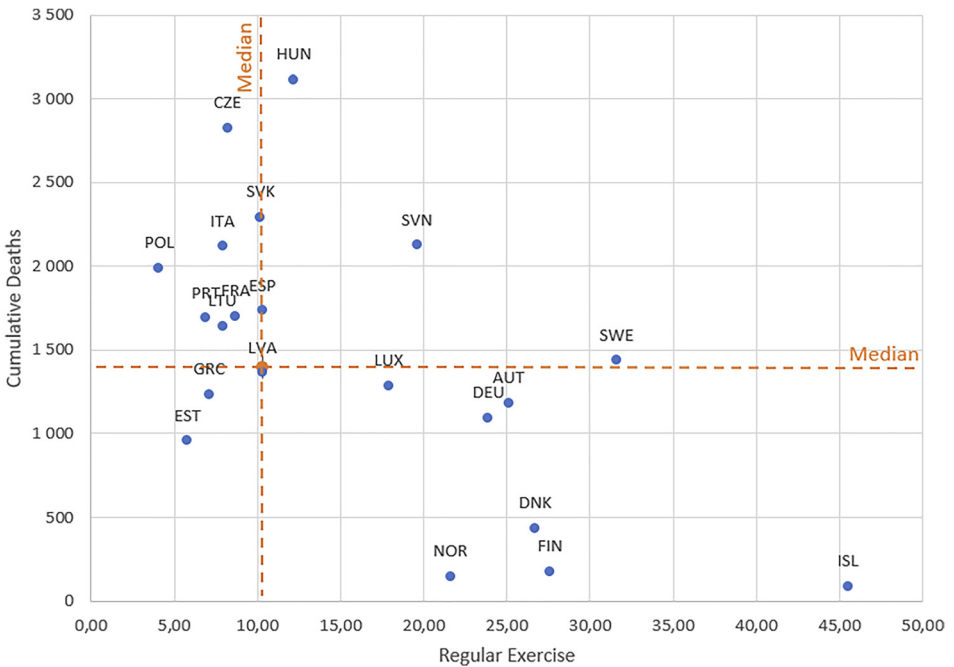


Fig. 5.



significance level. Adding *Alcohol Consumption* to the above-mentioned variables increased the adjusted- R^2 to 0.81 and doing the same with *Prescribed Medicine Use* gave almost the same results.¹⁸ The output tables of the three best models are presented below.

The best model

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.06314 -0.31947  0.07824  0.31664  0.82917

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.693027   0.411804   23.54 7.66e-14 ***
Fin$Number_of_nurses -0.119074  0.034821   -3.42 0.003512 **
Fin$NonPrescribed_eu -0.038712  0.008879   -4.36 0.000486 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5254 on 16 degrees of freedom
(5 observations deleted due to missingness)
Multiple R-squared:  0.7559, Adjusted R-squared:  0.7254
F-statistic: 24.78 on 2 and 16 DF, p-value: 1.259e-05
```

The best model explains 0.76 of the variability of *Log Cumulative COVID Deaths*. Both independent variables are significant at the standard level. The estimated coefficient for the variable *Density of Nurses* (5) shows that having one more nurse decreases cumulative COVID deaths by 13%. Similarly, one more unit *Consumption of Non-Prescribed Medicine* (28) decreases cumulative deaths by 4%.

The second-best model

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.79066 -0.35899 -0.06734  0.45780  0.66620

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.55330   0.69830   12.249 7.19e-09 ***
Fin$Number_of_nurses -0.13745  0.03603   -3.814 0.0019 **
Fin$NonPrescribed_eu -0.02784  0.00991   -2.810 0.0139 *
Fin$Prescribed_eu    0.02342  0.01156    2.026 0.0623 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4922 on 14 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared:  0.8116, Adjusted R-squared:  0.7713
F-statistic: 20.11 on 3 and 14 DF, p-value: 2.416e-05
```

The second-best model of our study has the explanatory power of 0.81. Two out of three variables – namely variables (5) and (28) – are significant at 5%, while the third variable, *Consumption of Prescribed Medicine* (27) is significant only at 10%. Similarly to our best model, the estimated coefficients for variables (5) and (28) are negative and are of almost the same magnitude as in the best model. At the same time, variable (27) has the opposite effect, meaning

¹⁸The Akaike Information Criterion of the three best models above in order of mention were as follows: 34.20, 31.04 and 31.91.



that consuming one more unit (per cent) of prescribed medicine leads to 2% more COVID deaths. This is consistent with the following interpretation: a higher level of prescribed medicine consumption is – to a large extent – the reflection of a higher prevalence of chronic diseases, which in turn is one of the circumstances increasing the probability of death in case of an eventual COVID infection.

The third-best model

```

Residuals:
    Min       1Q   Median       3Q      Max
-0.9805 -0.1953 -0.1145  0.2943  0.8487

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.248912   0.832263   9.911 5.62e-08 ***
Fin$Number_of_nurses -0.112737   0.032285  -3.492 0.00328 **
Fin$NonPrescribed_eu -0.034227   0.008508  -4.023 0.00111 **
Fin$Alcohol       0.127022   0.065135   1.950 0.07010 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4847 on 15 degrees of freedom
(5 observations deleted due to missingness)
Multiple R-squared:  0.8053,    Adjusted R-squared:  0.7664
F-statistic: 20.68 on 3 and 15 DF,  p-value: 1.38e-05

```

Finally, the third-best model has an R^2 of 0.81 and had three independent variables just like the second-best model. Variables (5) and (28) are significant at 5% and have similar estimated coefficients to the two models presented before. The third independent variable is *Alcohol Consumption* (23), which is only significant at 10% and its estimated coefficient suggests that consuming one more litre of pure alcohol increases cumulative COVID deaths by 14%.

5. CONCLUSION AND SUMMARY

This paper explains the country differences of COVID mortality rates in 24 European countries with econometric tools from the beginning of the pandemic until our cut-off date for information (16 July 2021). Since the quantification of the various country-specific anti-COVID health policy actions is impossible, we explain MRs with available, reliable, ex-ante economic, health and social indicators pertaining to the year 2019 – i.e., before the outbreak of the pandemic.

We built three models. Our best model with two ex-ante independent variables explains 0.76 of the variability of our ex-post dependent variable, the logarithm of Cumulative COVID Deaths. The estimated coefficient for the variable *Density of Nurses* shows that having one more nurse per 1,000 of population decreases cumulative COVID deaths by almost 15%. Similarly, one more unit *Consumption of Non-Prescribed Medicine* decreases cumulative deaths by 5%. All in all, it seems that until now those European countries were successful in minimising the fatalities where the population had a high level of health literacy, pursue healthier lifestyle and the healthcare systems worked with a relatively large nursing force already prior to COVID.

Our econometric estimates could be improved by the inclusion of a larger number of countries, but the trade-off appears to be high for us. If we went beyond the broadly defined



borders of the European Union and tried to include 20–40 less developed countries, our main dependent variable used in this paper, the cumulative number of COVID-related death figures would become less and less reliable. As we already mentioned above, outside of Europe, the real number of COVID death could be 2–3 times higher than the officially reported data. Thus, mixing reliable and unreliable data would bound to lead (paraphrasing the words of Keynes) to “precisely wrong” estimates.

REFERENCES

- Baum, F. – Freeman, T. – Musolino, C. – Abramovitz, M. – De Ceukelaire, W. – Flavel, J. – Friel, S. – Giugliani, C. – Howden-Chapman, P. – Huong, N. T. – London, L. (2021): Explaining COVID-19 Performance: What Factors might Predict National Responses? *British Medical Journal*, 372(91). <https://doi.org/10.1136/bmj.n91>.
- Osaki, Y. – Otsuki, H. – Imamoto, A. – Kinjo, A. – Fujii, M. – Kuwabara, Y. – Kondo, Y. (2021): Why are COVID-19 Mortality Rates by Country or Region so Different? An Ecologic Study of Factors Associated with Mortality from Novel Coronavirus Infections by Country. *Yonago Acta Medica*, 64(1): 80–91. <https://doi:10.33160/yam.2021.02.011>.
- Pažitný, P. – Kandilaki, D. – Mužik, R. – Benková, B. (2021): COVID-19 Case Fatality Rate in the Context of Healthcare System Organization and EHCI Performance: Focus on the Visegrad (V4) Countries. *Acta Oeconomica*, 71(S): 35–52. <https://doi.org/10.1556/032.2021.00028>.
- Taubenberger, J. K. – Morens, D. M. (2006): 1918 Influenza: The Mother of all Pandemics. *Emerging Infectious Diseases*, 12(1): 15–22. <https://doi.org/10.3201/eid1201.050979>.
- The Great Barrington Declaration (2020), <https://gbdeclaration.org/>.
- WHO Department of Communicable Disease Surveillance and Response (2000): *WHO Report on Global Surveillance of Pandemic-Prone Infectious Diseases*. Chapter 3. [WHO/CDS/CSR/ISR/2000.1](http://www.who.int/cds/publications/2000/1).
- Yeh, Ting-Yu – Contreras, P. G. (2021): Viral Transmission and Evolution Dynamics of SARS-CoV-2 in Shipboard Quarantine. *Bull World Health Organ*, 99: 486–495. <http://dx.doi.org/10.2471/BLT.20.255752>.



APPENDIX

The monthly changes of excess mortality during the COVID pandemic in selected European countries, %

TIME	ISO_3	2020-01	2020-02	2020-03	2020-04	2020-05	2020-06	2020-07	2020-08	2020-09	2020-10	2020-11	2020-12	2021-01	2021-02	2021-03	2021-04	2021-05	
Austria	AUT	-2	-1	5	11	2	2	3	5	9	12	48	38	9	-3	0	12	9	9
Belgium	BEL	-4	-5	12	73	8	-1	-7	18	2	23	59	19	2	-9	-11	7	5	16
Czechia	CZE	-2	0	-1	3	-1	4	4	7	11	53	76	45	54	40	61	31	6	23
Denmark	DNK	-2	-7	-4	6	-1	1	2	2	4	2	6	11	11	-7	-12	-3	6	5
Estonia	EST	-8	-5	-1	7	3	3	5	6	8	2	7	14	12	9	30	26	13	7
Finland	FIN	-10	0	1	8	6	6	2	4	7	2	6	4	-2	-2	-4	-1	4	4
France	FRA	-5	-2	16	36	2	3	-1	6	8	16	31	16	10	9	7	17	8	8
Germany	DEU	-2	-4	-3	9	2	3	0	7	6	5	13	31	23	-2	-10	6	7	7
Greece	GRC	1	7	8	1	4	-2	6	7	10	4	28	20	-8	-3	16	25	26	9
Hungary	HUN	-10	-6	-2	1	-3	0	-2	-1	4	16	59	46	6	0	50	48	5	18
Iceland	ISL	14	4	-2	4	8	-21	-5	-7	1	11	6	5	4	2	:	:	:	6
Italy	ITA	-10	-1	50	42	4	0	3	5	6	16	52	27	7	3	16	19	5	14
Latvia	LVA	-8	-9	-13	-3	-1	4	-1	4	-1	4	11	29	29	14	-2	4	14	9
Lithuania	LIT	-12	-11	-1	2	-2	8	4	7	7	10	36	70	29	2	6	10	12	12
Luxembourg	LUX	-8	-11	6	19	7	0	0	12	7	6	45	33	9	-4	10	8	1	9
Netherlands	NLD	-4	-4	15	54	4	1	-3	9	6	18	21	23	19	4	-3	10	8	10
Norway	NOR	-4	-2	-1	3	-3	-2	-2	2	5	2	1	-2	-6	-11	-9	-4	-3	3
Poland	POL	-6	-2	-1	4	4	5	6	11	11	45	97	50	26	11	38	65	26	22
Portugal	PRT	-4	-4	6	16	10	4	26	7	14	15	26	21	60	24	-5	-6	-1	12
Slovakia	SVK	-4	-2	2	-1	0	0	2	5	5	21	39	58	74	68	54	28	2	23
Slovenia	SVN	-6	0	-1	6	2	-10	4	3	8	26	91	79	27	-1	3	10	12	19
Spain	ESP	-5	-3	54	81	5	-3	6	12	15	22	24	7	18	10	0	4	4	15
Sweden	SWE	-3	-5	2	38	24	11	-1	-1	-2	-3	11	25	17	-5	-8	-5	0	11
Switzerland	CHE	-4	-1	12	28	-1	2	2	5	5	15	66	60	23	-6	-7	1	-3	16

Source: Eurostat, accessed on July 16, 2021.

Open Access. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes - if any - are indicated. (SID_1)

