Open Access

Epidemiology & Public Health

Article Type: Research Article

Volume 3 Issue 1 - 2025

Modelling of Covid-19 cases, deaths and excess mortality in 19 European countries

Dearden John C, PhD*; Rowe Philip H, PhD

School of Pharmacy & Biomolecular Sciences, Liverpool John Moores University, UK.

Abstract

Introduction: This study seeks to model Covid-19 cases and deaths and excess all-cause mortality in 19 European countries over the period 2020-2023, using a number of socio-demographic factors as independent variables.

Methods: We used QSAR (Quantitative Structure-Activity Relationship) multiple linear regression modelling to examine which of over 20 socio-demographic factors best modelled the data. The Minitab 'Best Subsets' routine was used to select the best descriptor sets for each model.

Results: We were unable to obtain good QSAR correlations for cases, probably because of significant under-reporting of Covid-19 infections. Despite some under-reporting of Covid-19 deaths, they were well modelled with two factors, namely a pollution index and levels of Cardiovascular problems (CHD, CVD, stroke), both of which increased Covid-19 deaths.

Excess all-cause mortality is now regarded as a more accurate indication of deaths, and we obtained very good QSAR models for seven publicly available excess all-cause mortality data in the 19 European countries. Modelling selected the same two factors in each case, namely level of Covid-19 vaccination and latitude N (north), both with negative signs.

Discussion and conclusion: The models show that Covid-19 deaths appeared to be controlled largely by levels of pollution and of cardiovascular diseases. Excess all-cause mortality during the pandemic was modelled best by vaccination levels and latitude north. The rôle of latitude N is not clear, but we suggest that it could be a proxy for exercise levels and temperature, both of which correlate reasonably well with latitudes of the 19 European countries.

***Corresponding author:** Dearden John C

School of Pharmacy & Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.

Tel: +44-(0)1928722292; Email: j.c.dearden@ljmu.ac.uk

Received: Feb 04, 2025

Accepted: Mar 21, 2025

Published: Mar 28, 2025

Epidemiology & Public Health - www.jpublichealth.org

Citation: Dearden JC, Rowe PH. Modelling of Covid-19 cases, deaths and excess mortality in 19 European countries. Epidemiol Public Health. 2025; 3(1): 1066.

Keywords: Covid-19; Cases; Deaths; Excess all-cause mortality; Modelling; QSAR.

Introduction

The viral infection Severe Acute Respiratory Syndrome 2 (SARS-Cov2), generally referred to as Covid-19, is a severe viral infection that started in Wuhan, China, in late 2019, and spread rapidly throughout the world. It was declared a pandemic by the World Health Organisation (WHO) on 12 March 2020 [1]. The western world in particular was totally unprepared for this catastrophe, as the last time such a huge event had occurred was a severe influenza outbreak, the so-called Spanish flu, in 1918, which killed at least 50 million people worldwide [2].

In the months following the initial outbreak of Covid-19, a number of possible factors that could affect infection and death rates were proposed in scientific reports and in the media; these included the effect of vitamin D on reducing infection rates [3] and a range of socio-demographic factors such as obesity and levels of smoking.

In an attempt to elucidate which if any of these factors were contributing to the pandemic, we used the QSAR (Quantitative Structure-Activity Relationship) approach [4] to correlate cumulative Covid-19 cases and deaths per million inhabitants in 20

European countries up to 9th May 2020 with 20 potential factors [5]. We found that a key factor in infection rates was, as Ilie et al. [3] had reported, levels of vitamin D in populations. Other controlling factors for infection rates were levels of stroke deaths, levels of smoking and levels of respiratory deaths. Covid-19 death rates were found to correlate best with population densities, proportions of elderly (>65) people and levels of inactivity.

It must be pointed out that such modelling requires the input data to be as accurate as possible. However, it is known that the numbers of both cases and deaths in the pandemic almost certainly contain considerable error. For example, Starnini et al. [6], focussing on Italy and Spain, found "several biases of case-based surveillance data and temporal and spatial limitations in the data", and called for "an improvement in the process of COVID-19 data collection, management, storage, and release". Kobak [7] stated that "for many countries the reported numbers of cases and deaths can be gross underestimations". In some instances, the opposite is the case; for example, in the United Kingdom if a person died in hospital with Covid-19, the death was reported as a Covid death even if the person died of something else. Nevertheless, we believe that QSAR modelling of Covid-19 cases and deaths should yield indications of which socio-demographic factors are important in controlling the development of the pandemic.

The present work extends that previously reported [5] by using Covid-19 cumulative cases and deaths per million inhabitants on two additional dates: 5th May 2023 (the date on which the WHO declared the pandemic to be over) and 7th November 2021 (midway between the other two dates), for 19 European countries. Turkey was removed from our original list of 20 countries because it is partly in Europe and partly in Asia, and it was not always clear whether Turkish data related only to its European area or to the whole of Turkey.

We also modelled all-cause excess mortality in the selected 19 European countries during periods in late 2020 to late 2022. This statistic is defined as the increase in all-cause mortality relative to the expected mortality [8]. It is regarded as "a more comprehensive measure of the total impact of the pandemic than the confirmed Covid-19 death count alone. It captures not only the confirmed deaths, but also COVID-19 deaths that were not correctly diagnosed and reported as well as deaths from other causes that are attributable to the overall crisis conditions" [9].

Methods

Cumulative Covid-19 cases and deaths per million inhabitants were taken from Wikipedia [10]. Excess mortalities were taken from seven sources: Office for National Statistics (ONS) [11]; Our World in Data (OWID) [12]; Eurostat (Euro) [13]; Organisation for Economic Co-operation and Development (OECD) [14]; The Collaborators (Coll) [15]; The Economist (Econ) [16]; World Health Organization (WHO) [17]. Data for the potential socio-demographic factors (descriptors) for the 19 European countries were taken from Dearden and Rowe [5]. Two of them require explanation: HDI is a Human Development Index, a composite index of life expectancy, education and per capita income [18]; UHC (Universal Health Coverage) is a measure of access to the full range of quality health services without financial hardship [19]. Lengths (days) of the first and second lockdowns (LD1 and LD2) were also included as descriptors [20], as were three datasets giving Covid-19 vaccination levels; the percentage of people fully vaccinated in July 2022 (VaccJ22) [21] and in December 2022 (VaccD22) [22], and the number of vaccinations per 100 people by late 2022-early 2023 (Vacc22/23) [23]. Statistical analysis of the data was performed with Minitab statistical software, version 20.0 [24], using the best sub-sets routine to find the factors (descriptors) that best modelled the Covid-19 data. For 19 objects (in this case countries) no more than three or four descriptors should be used in a model, to minimise the risk of chance correlations [25]. It should be noted that almost all types of data contain error, particularly *in vivo* data, so a good statistical model should have a coefficient of determination (r^2) in the region of 0.7-0.8, and a predictive r^2 (termed q^2) of ≥0.5. The probability (p) of each of the descriptors in a correlation being there by chance should be ≤0.05 (≤1 in 20).

European countries were not fully vaccinated against Covid-19 until the spring of 2022 [26]. Therefore, in order to assess the effect of vaccination on the progress of the pandemic, cumulative cases/million and deaths/million for 5th July 2022 were subtracted from those for 5th May 2023.

Results

Table 1 shows the cumulative cases and deaths per million inhabitants for the 19 European countries.

Table 1: Cumulative Covid cases and deaths per million inhabitants for 19 European countries.

Country	Cases/million			Deaths/million		
	09.05.20	07.11.21	05.05.23	09.05.20	07.11.21	05.05.23
Belgium	4509	122845	412347	741	2263	2946
Czechia	777	175279	442009	25	2975	4075
Denmark	1755	68617	581120	50	466	1461
Estonia	1299	155327	465721	42	1218	2260
Finland	1038	29772	267231	47	235	1728
France	2054	108270	603309	386	1788	2527
Germany	2040	57296	460404	89	1208	2085
Hungary	325	92084	227072	40	3219	5035
Iceland	4919	35502	562838	27	91	700
Ireland	4548	93252	341598	285	1127	1778
Italy	3605	81192	436574	501	2239	3213
Netherlands	2410	125452	490974	307	1058	1311
Norway	1501	40084	273349	41	182	1011
Portugal	2653	106981	542684	108	1769	2589
Slovakia	267	95902	343786	5	2439	3891
Spain	4732	106145	291112	558	1876	2542
Sweden	2444	112192	257662	307	1432	2298
Switzerland	3511	102139	505049	177	1249	1607
U.K.	2181	138884	365016	470	2517	3351

Correlations between cases and deaths are: for 09.05.20, r^2 = (+) 0.425; for 07.11.21, r^2 = (+) 0.425; for 05.05.23, r^2 = (-) 0.100.

Table 2: Cumulative Covid cases and deaths per million inhabitants for 19 European countries from the start of full vaccination.

Country	Cases/million		Deaths/million	
	05.07.22	05.05.23–05.07.22	05.07.22	05.05.23–05.07.22
Belgium	366817	45237	2742	202
Czechia	402000	40071	3849	226
Denmark	540220	1847	1104	360
Estonia	426649	38837	1863	398
Finland	213063	53542	950	798
France	469792	150138	2327	256
Germany	341617	118782	1738	340
Hungary	199204	28056	4815	227
Iceland	513539	49299	315	301
Ireland	319624	22179	1524	267
Italy	316309	120251	2850	362
Netherlands	468165	23030	1277	34
Norway	267808	6288	628	381
Portugal	504755	38026	2358	234
Slovakia	330917	12220	3705	187
Spain	272825	18168	2324	218
Switzerland	433883	71567	1532	127
U.K.	338308	26643	2983	385

Table 3 gives the r^2 values of correlations of Covid-19 cases and deaths with each socio-demographic factor separately.

Table 4 gives the excess all-cause mortality (EACM) values for the 19 European countries from seven different sources.

Table 5 gives the correlation matrix of each set of values in Table 4.

Table 4: EACM figures for the 19 European countries, from seven different sources.

Country	ONS	OWID	Euro	OECD	Coll	Econ	WHO
	%	%	%	/100K	/100K	/100K	/100K
Belgium	8.0	20.6	8.7	137.4	146.6	139	77
Czechia	15.5	43.1	19	346.5	244.8	253	173
Denmark	5.4	12.9	6.1	19.5	94.1	57	32
Estonia	11.1	28.6	11.5	139.6	226.7	199	127
Finland	8.7	17.6	8.8	34.3	80.8	114	26
France	9.2	24.3	10.1	137.4	124.2	102	63
Germany	8.6	18.3	9.7	92.5	120.5	122	116
Hungary	10.1	28.6	11.0	242.4	297.8	262	189
Iceland	8	16.8	8.1	18.8	-47.8	59	-2
Ireland	8.9	*	*	*	12.5	122	29
Italy	12.3	27.9	12.2	215.1	227.4	185	133
Netherlands	11.6	29.6	12.3	138.4	140	148	85
Norway	5	6.9	5.1	-27.7	7.2	87	-1
Portugal	11.3	27.8	11.5	202.5	202.2	154	100
Slovakia	18.7	49.7	19.0	313.3	250.4	356	223
Spain	11.3	28.6	12.3	184.1	186.7	153	111
Sweden	4.4	6.7	1.7	54.5	91.2	85	56
Switzerland	9.4	19.4	12.1	106.9	93.1	119	47
U.K.	7	24.5	*	155.5	126.8	168	109

Table 3: Single correlations (r^2) of cases/million and deaths/million with socio-demographic descriptors (signs in brackets indicate a positive or negative correlation).

Descriptors	09 May 2020		07 November 2021		05 May 2023	
	Cases/ M	Deaths/M	Cases/M	Deaths/M	Cases/M	Deaths/M
Cases/M	(+) 1.000	(+) 0.417	(+) 1.000	(+) 0.423	(+) 1,000	(-) 0.109
Lockdown 1	(+) 0.158	(+) 0.332	(+) 0.151	(+) 0.069	(+) 0.017	(+) 0.017
Lockdown 2	(-) 0.016	(-) 0.004	(+) 0.097	(+) 0.012	(+) 0.164	(-) 0.000
VaccJ22	*	*	(-) 0.098	(-) 0.142	(+) 0.037	(-) 0.204
VaccD22	*	*	(-) 0.128	(-) 0.147	(+) 0.043	(-) 0.211
Vacc22/23	*	*	(-) 0.065	(-) 0.088	(+) 0.022	(-) 0.147
Vitamin D	(-) 0.345	(-) 0.177	(-) 0.044	(-) 0.013	(-) 0.018	(+) 0.004
CVD	(-) 0.401	(-) 0.264	(+) 0.019	(+) 0.289	(-) 0.146	(+) 0.501
CHD	(-) 0.380	(-) 0.263	(+) 0.070	(+) 0.182	(-) 0.103	(+) 0.353
Stroke	(-) 0.413	(-) 0.208	(+) 0.026	(+) 0.294	(-) 0.074	(+) 0.507
Respiratory	(+) 0.105	(+) 0.051	(+) 0.009	(+) 0.014	(-) 0.000	(+) 0.001
Dementia	(+) 0.022	(+) 0.006	(-) 0.178	(-) 0.191	(-) 0.063	(-) 0.156
Diabetes	(-) 0.073	(-) 0.157	(-) 0.049	(+) 0.000	(+) 0.062	(+) 0.006
Obesity	(-) 0.001	(+) 0.003	(+) 0.059	(+) 0.187	(-) 0.210	(+) 0.204
Smoking	(-) 0.179	(-) 0.012	(+) 0.261	(+) 0.267	(-) 0.009	(+) 0.264
Poverty	(+) 0.076	(+) 0.219	(+) 0.120	(+) 0.067	(-) 0.018	(+) 0.028
Inactivity	(+) 0.169	(+) 0.324	(-) 0.000	(+) 0.020	(-) 0.098	(+) 0.008
Exercise	(-) 0.003	(-) 0.122	(-) 0. 426	(-) 0.412	(-) 0.042	(-) 0.245
Vegetarian	(+) 0.086	(+) 0.097	(-) 0.007	(-) 0.017	(+) 0.015	(-) 0.065
Alcohol	(-) 0.070	(-) 0.016	(+) 0.235	(+) 0.362	(+) 0.019	(+) 0.365
%>55	(-) 0.036	(+) 0.035	(+) 0.008	(+) 0.023	(+) 0.006	(+) 0.028
%Afro	(+) 0.113	(+) 0.494	(+) 0.044	(+) 0.009	(+) 0.045	(-) 0.007
Popn Density	(+) 0.037	(+) 0.269	(+) 0.128	(+) 0.117	(+) 0.062	(+) 0.023
HDI	(+) 0.153	(+) 0.009	(-) 0.123	(-) 0.469	(+) 0.001	(-) 0.584
Latitude N	(-) 0.016	(-) 0.142	(-) 0.109	(-) 0.415	(-) 0.007	(-) 0.276
Pollution index	(+) 0.015	(+) 0.336	(+) 0.153	(+) 0.680	(-) 0.004	(+) 0.536
Exp. index	(-) 0.001	(+) 0.016	(+) 0.000	(+) 0.090	(+) 0.002	(+) 0.088
UHC index	(+) 0.402	(+) 0.164	(-) 0.061	(-) 0.275	(+) 0.074	(-) 0.486

Table 5: Correlation matrix (r^2) of seven sets of EACM data.

	ONS	OWID	OECD	Euro	Coll	Econ
OWID	0.92					
OECD	0.734	0.846				
Euro	0.929	0.914	0.383			
Coll	0.449	0.573	0.762	0.432		
Econ	0.712	0.787	0.78	0.656	0.664	
WHO	0.59	0.716	0.217	0.566	0.839	0.863

Table 6 gives the Covid vaccination figures for the three sets of vaccination data that we used.

Table 7 gives the r^2 values of correlations of EACM with each socio-demographic factor separately.

The best QSARs for cumulative cases and deaths on each of the three selected dates are given below. The statistical data are: n=number of countries; r^2=coefficient of determination; q^2=predictive coefficient of determination; s=standard error of prediction; *F*=variance ratio or Fisher coefficient (a measure of goodness of fit); p=probability of chance correlation (range 0 to 1).

Table 6: Covid vaccination data.

Country	VaccJ22 (%)	VaccD22 (%)	Vacc22/23 (/100)
Belgium	79	79.3	253.9
Czechia	64	64.4	177.3
Denmark	82	81.7	223.8
Estonia	64	63.7	158.7
Finland	78	78.6	237.7
France	78	78	227.0
Germany	76	76.3	228.7
Hungary	64	63.7	167.6
Iceland	79	81.3	216.0
Ireland	81	81.2	221.0
Italy	79	80.6	243.2
Netherlands	70	68.6	205.6
Norway	74	75.5	223.5
Portugal	87	86.2	272.8
Slovakia	51	51.1	102.0
Spain	87	85.9	219.9
Sweden	75	73.7	242.7
Switzerland	69	69.6	193.3
U.K.	73	75.7	224

*(%) means percentage of people fully vaccinated; (/100) means average number of vaccinations per 100 of population.

Cases/million 9.5.20 = 12773 – 75.7 vitamin D – 76.2 stroke – 217.5 >65 **(1)**

n = 19 r^2 = 0.677 q^2 = 0.502 s = 938.3 *F* = 10.5 All p≤0.047

Deaths/M 9.5.20 = – 3325 + 13.42 pollution index + 37.47 UHC index **(2)**

n = 19 r^2 = 0.737 q^2 = 0.626 s = 121.6 *F* = 22.4 All p < 0.001

Deaths/M 9.5.20 = 242 – 15.89 stroke + 12.70 pollution index **(3)**

n = 19 r^2 = 0.722 q^2 = 0.632 s = 125.0 *F* = 20.8 All p < 0.001

Deaths/M 9.5.20 = 73 – 13.51 stroke + 10.56 pollution index + 4.82 inactivity **(4)**

n = 19 r^2 = 0.797 q^2 = 0.701 s = 110.4 *F* = 19.7 All p≤0.033

Cases/M 31.1.21 No good models with $r^2 \geq 0.6$

Deaths/M 31.1.21 = – 142 + 42.06 pollution index – 5.62 exponential index **(5)**

n = 19 r^2 = 0.710 q^2 = 0.000 s = 324.8 *F* = 19.1 All p≤0.039

Cases/M 7.11.21 No good models with $r^2 \geq 0.6$

Deaths/M 7.11.21 = – 769 + 7.12 CHD + 56.12 pollution index **(6)**

n = 19 r^2 = 0.793 q^2 = 0.723 s = 446.7 *F* = 30.7 All p≤0.008

Table 7: Single correlations (r^2) of all-cause mortality values with socio-demographic descriptors.

Descriptor	ONS	OWID	Euro	OECD	Coll	Econ	WHO
Lockdown 1	(-) 0.008	(+) 0.002	0.020	0.014	(+) 0.001	(-) 0.006	(+) 0.005
Lockdown 2	(+) 0.038	0.049	0.097	0.048	0.010	(+) 0.000	(+) 0.025
VaccJ22	(-) 0.294	0.316	0.282	0.224	0.196	(-) 0.546	(-) 0.375
VaccD22	(-) 0.312	(-) 0.330	(-) 0.281	(-) 0.243	0.235	(-) 0.558	(-) 0.402
Vacc22/23	(-) 0.393	(-) 0.4 10	0.235	(-) 0.377	0.178	(-) 0.537	(-) 0.360
Vitamin D	(+) 0.008	0.003	0.003	0.002	0.005	(+) 0.041	(+) 0.004
CVD	(+) 0.307	0.329	0.260	0.376	0.359	0.613	(+) 0.536
CHD	(+) 0.222	0.254	0.185	0.247	0.343	0.520	(+) 0.441
Stroke	(+) 0.370	0.403	0.266	0.439	0.513	0.652	(+) 0.565
Respiratory	(-) 0.034	(-) 0.000	0.005	0.003	0.028	(-) 0.006	(-) 0.008
Dementia	(-) 0.128	(-) 0.123	(-) 0.117	(-) 0.222	0.310	(-) 0.140	(-) 0.262
Diabetes	(+) 0.025	(+) 0.008	0.022	0.014	0.061	0.003	(+) 0.058
Obesity	(-) 0.000	(+) 0.035	0.048	0.089	0.009	(+) 0.061	(+) 0.047
Smoking	(+) 0.400	0.393	0.419	0.413	0.477	(+) 0.405	(+) 0.498
Poverty	(+) 0.013	(+) 0.015	0.007	(+) 0.036	(+) 0.142	(+) 0.033	(+) 0.087
Inactivity	(-) 0.108	(-) 0.066	(-) 0.113	(-) 0.004	(-) 0.012	(-) 0.042	(-) 0.029
Exercise	(-) 0.350	(-) 0.405	(-) 0.390	(-) 0.511	0.420	(-) 0.284	(-) 0.294
Vegetarian	(-) 0.151	(-) 0.206	0.109	(-) 0.110	(-) 0.090	(-) 0.150	(-) 0.081
Alcohol	(+) 0.242	(+) 0.370	0.344	(+) 0.452	(+) 0.286	(+) 0.261	(+) 0.331
% >65	(-) 0.009	(-) 0.031	(-) 0.028	(+) 0.000	(+) 0.167	(-) 0.008	(+) 0.026
% Afro	(-) 0.065	(-) 0.025	(-) 0.039	(-) 0.018	(-) 0.032	(-) 0.073	(-) 0.059
Popn.Density	(+) 0.018	(+) 0.044	0.068	(+) 0.058	0.054	(+) 0.015	(+) 0.055
HDI	(-) 0.428	(-) 0.509	(-) 0.330	0.603	(-) 0.723	(-) 0.548	(-) 0.606
Latitude N	(-) 0.249	(-) 0.262	(-) 0.271	(-) 0.439	(-) 0.391	(-) 0.181	(-) 0.286
Poll. index	(+) 0.165	(+) 0.262	0.204	0.468	(+) 0.347	0.263	(+) 0.354
Exp. index	(+) 0.180	0.100	0.201	0.158	0.160	0.081	(+) 0.097
UHC index	(-) 0.345	(-) 0.390	(-) 0.296	(-) 0.395	(-) 0.477	(-) 0.612	(-) 0.494

Deaths/M 7.11.21 = – 1000 +33.1 stroke + 52.00 pollution index **(7)**

n = 19 r^2 = 0.782 q^2 = 0.708 s = 458.3 *F* = 28.8 All p≤0.013

Cases/M 5.5.23 No good models with $r^2 \geq 0.6$

Deaths/M 5.5.23 = – 872 + 64.6 stroke + 50.24 pollution index **(8)**

n = 19 r^2 = 0.811 q^2 = 0.744 s = 516.6 *F* = 34.4 All p<0.001

Deaths/M 5.5.23 = – 367 + 13.01 CHD + 58.58 pollution index **(9)**

n = 19 r^2 = 0.802 q^2 = 0.718 s = 528.8 *F*=32.1 All p<0.001

For both (cases/M 5.5.23 – cases/M 5.7.22) and (deaths/M 5.5.23 – deaths/M 5.7.22) no satisfactory models were found.

The best QSARs for excess mortality are given below.

ONS = 35.76 – 0.0583 Vacc22/23 – 0.264 Lat N **(10)**

n = 19 r^2 = 0.680 q^2 = 0.528 s = 2.085 *F* = 17.0 All p≤0.002

$$ONS = 42.31 – 0.240\ VaccJ22 – 0.286\ Lat\ N \quad \textbf{(11)}$$

$n = 19$ $r^2 = 0.624$ $q^2 = 0.421$ $s = 2.260$ $F = 13.3$ All $p \leq 0.002$

$$OWID = 105.6 – 0.182\ Vacc22/23 – 0.832\ Lat\ N \quad \textbf{(12)}$$

$n = 18$ $r^2 = 0.713$ $q^2 = 0.608$ $s = 6.217$ $F = 18.7$ All $p \leq 0.001$

$$OWID = 128.9 – 0.784\ VaccJ22 – 0.912\ Lat\ N \quad \textbf{(13)}$$

$n = 18$ $r^2 = 0.674$ $q^2 = 0.517$ $s = 6.632$ $F = 15.5$ All $p \leq 0.001$

$$EURO = 41.45 – 0.0674\ Vacc22/23 – 0.323\ Lat\ N \quad \textbf{(14)}$$

$n = 17$ $r^2 = 0.688$ $q^2 = 0.574$ $s = 2.576$ $F = 15.4$ All $p \leq 0.002$

$$EURO = 39.9 – 0.0690\ Vacc22/23 + 0.0192\ Lockdown2 – 0.303\ Lat\ N \quad \textbf{(15)}$$

$n = 17$ $r^2 = 0.769$ $q^2 = 0.599$ $s = 2.301$ $F = 14.4$ All $p \leq 0.050$

$$OECD = 1158 – 6.51\ VaccJ22 – 10.43\ Lat\ N \quad \textbf{(16)}$$

$n = 18$ $r^2 = 0.769$ $q^2 = 0.690$ $s = 51.81$ $F = 24.9$ All $p < 0.001$

$$OECD = 1136 – 6.39\ VaccD22 – 10.15\ Lat\ N \quad \textbf{(17)}$$

$n = 18$ $r^2 = 0.763$ $q^2 = 0.657$ $s = 52.45$ $F = 24.1$ All $p < 0.001$

$$Collaborators = 1611 + 6.51\ \%\ smoking – 17.96\ HDI \quad \textbf{(18)}$$

$n = 19$ $r^2 = 0.820$ $q^2 = 0.749$ $s = 41.04$ $F = 36.4$ All $p \leq 0.010$

$$Collaborators = 504.2 + 1.336\ CHD – 8.93\ Lat\ N \quad \textbf{(19)}$$

$n = 19$ $r^2 = 0.815$ $q^2 = 0.747$ $s = 41.60$ $F = 35.2$ All $p < 0.001$

$$Economist = 923.1 – 6.646\ VaccD22 – 5.35\ Lat\ N \quad \textbf{(20)}$$

$n = 19$ $r^2 = 0.808$ $q^2 = 0.737$ $s = 34.75$ $F = 73.7$ All $p < 0.001$

$$Economist = 945.9 – 6.775\ VaccJ22 – 5.64\ Lat\ N \quad \textbf{(21)}$$

$n = 19$ $r^2 = 0.822$ $q^2 = 0.753$ $s = 33.50$ $F = 36.9$ All $p < 0.001$

$$WHO = 728.2 – 4.852\ VaccD22 – 5.37\ Lat\ N \quad \textbf{(22)}$$

$n = 19$ $r^2 = 0.761$ $q^2 = 0.699$ $s = 35.55$ $F = 25.5$ All $p < 0.001$

$$WHO = 739.3 – 4.879\ VaccJ22 – 5.57\ Lat\ N \quad \textbf{(23)}$$

$n = 19$ $r^2 = 0.758$ $q^2 = 0.691$ $s = 32.75$ $F = 25.1$ All $p < 0.001$

Discussion

Correlation between cases and deaths

If figures for cases and deaths were both recorded accurately, it would be reasonable to expect these to be positively correlated. The negative correlation seen for the 2023 data strongly suggests that by that year, one, or possible both, sets of data were no longer being reliably and consistently recorded in European nations.

Correlation of cases with potential contributory factors

Among equations 1-9 linking cases and deaths to possible factors, there are only two cases where an acceptable relationship for cases could be established. In contrast, eight such equations emerged for deaths. The correlation coefficients between cases and deaths had already created doubt about one or both of these measures, and this finding casts particular suspicion on the data for cases.

In our previous study of the modelling of Covid-19 cases and deaths in 20 European countries, we used the Box and Cox approach [27] to transform the Covid-19 data and the potential descriptors. Best sub-sets regression analysis yielded two models, one selecting vitamin D levels, stroke deaths/100K and respiratory deaths/100K, and the other selecting vitamin D levels, smoking % prevalence and HDI.

In the present work, we studied only 19 European countries, and found (Equation 1) that whilst vitamin D level was again selected as a good descriptor for the 9th May 2020 cases data, other selected descriptors were different from those previously selected. Nevertheless, stroke deaths/100K correlated well (r^2=0.750) with CVD deaths/100K, so the contributions of those two descriptors could be similar. However, in equation 1 both the stroke and >65 terms are negative, which is contrary to expectation. No reliance can therefore be placed on Equation 1.

For all dates other than 9th May 2020, no good or even satisfactory models could be obtained for cases. The reason is likely to be inaccuracies in the data; on later dates infections by the Covid-19 Omicron variant were generally much milder, which led to many fewer cases being reported. Several studies have examined the effects of inaccurate data on Covid-19 policies [28-30].

From the above, it is clearly necessary that great improvements in the accuracy of the numbers of both cases and deaths from a pandemic need to be made.

Correlation of deaths with potential contributory factors

In stark contrast to the modelling of Covid-19 cases, that of Covid-19 deaths shows that very similar descriptors were selected for each of the three dates (Equations 2-9). Pollution index was selected in each of those QSARs; CVD deaths/100K and stroke deaths/100K were also selected several times. It can be inferred that Covid-19 deaths in European countries were much better reported than were cases, although it is claimed [31] that there was under-reporting of Covid-19 deaths. Nevertheless, it can be concluded that the main factors affecting Covid-19 deaths in European countries were pollution and cardiovascular problems. This confirms what has observed elsewhere for pollution [32] and for cardiovascular disease [33,34]. It should be noted that the study by Ssentongo et al. [34] found, when examining the effects of 11 comorbidities on Covid-19 mortality, that cardiovascular disease had the greatest effect, increasing mortality by 125%.

If the effects of any future similar pandemic are to be contained, it is clear that much work needs to be done in reducing levels of pollution and cardiovascular disease.

Correlation of excess all-cause mortality with potential contributory factors

The strong positive correlations among most of the measures of excess deaths (Table 5) suggest that they are all measuring essentially the same thing, although the data from The Collaborators [15] correlate less well with the other measures, leaving the possibility that this data set is measuring something slightly different.

In contrast to numbers of cases and deaths, the data on excess deaths are well modelled using a limited number of variables (only eight appear in equations 10-23), and these are dominated by various measures of vaccination levels, all of which show negative correlation with excess deaths, as might be expected, and by latitude N, which also has a negative sign.

Latitude N must be a proxy for one or more other factors. Of the other factors that we have used, latitude N correlates best with exercise levels (r^2=0.537) and HDI (r^2=0.357), which is understandable. However, it occurred to us that two other factors, happiness and temperature, could possibly affect excess all-cause mortality. The Bible tells us [35] that: "A cheerful heart is good medicine, but a crushed spirit dries up the bones". Happiness levels are available [36], and these correlated well with latitude N (r^2=0.543) for our 19 European countries. Temperature also is known to affect excess mortality [37], although there are very few studies in this area. Mean national temperatures were taken from [38].

Table 8 gives the data for these factors.

Table 8: Latitude N and factors for which it may be a proxy.

Country	Latitude N	Exercise	Temperature	Happiness	HDI	Smoking
Belgium	50.85	24	9	6.86	91.9	23.3
Czechia	50.08	28.4	6.8	6.85	89.1	33.2
Denmark	55.68	54.6	7.5	7.59	93	17
Estonia	59.37	23.2	5.5	6.46	88.2	33.1
Finland	60.25	54.6	2.7	7.8	92.5	20.9
France	48.83	25	11.2	6.66	89.1	27.7
Germany	52.5	48.3	7.8	6.89	93.9	30.4
Hungary	47.48	28.6	10	6.04	84.5	28.4
Iceland	64.17	50.8	3.4	7.53	93.8	16.1
Ireland	53.35	29.1	9.6	6.91	94.2	22.2
Italy	41.9	18.2	13.5	6.4	88.3	24
Netherlands	52.38	25	9.3	7.4	93.3	25.1
Norway	59.92	56.8	4.3	7.32	95.4	22.3
Portugal	38.7	18.4	15.7	5.97	85	22.6
Slovakia	48.17	29.4	6.2	6.47	85.7	28.7
Spain	40.42	34	15.5	6.44	89.5	29.2
Sweden	59.53	54.1	4.7	7.4	93.7	20.6
Switzerland	46.95	21.8	6	7.24	94.6	23.3
U.K.	51.8	36.7	9.3	6.8	92	19.2

The only measures of excess deaths that were not modelled by vaccination levels and latitude N were those from The Collaborators [15]. However, it has already been commented that this data-set may be measuring something slightly different from the others.

We checked whether any of the above factors could be important by best sub-sets modelling of EACM without using latitude N as a factor. The best resulting QSARs were:

EACM (ONS) = 24.1 – 0.0480 Vacc22/23 – 0.119 Exercise **(24)**

n = 19 r^2 = 0.634 q^2 = 0.384 s = 2.230
F = 13.8 All p≤0.005

EACM (ONS) = 20.63 – 0.0691 Vacc22/23 + 0.451 Temperature **(25)**

n = 19 r^2 = 0.609 q^2 = 0.378 s = 2.304
F = 12.5 All p≤0.009

EACM (OWID) = 57.62 – 0.220 Vacc22/23 + 1.588 Temperature **(26)**

n = 18 r^2 = 0.695 q^2 = 0.540 s = 6.414
F = 17.1 All p≤0.002

EACM (OWID) = 69.3 – 0.147 Vacc22/23 – 0.393 Exercise **(27)**

n = 18 r^2 = 0.691 q^2 = 0.511 s = 6.460
F = 16.8 All p≤0.002

EACM (EURO) = 27.3 – 0.0540 Vacc22/23 – 0.150 Exercise **(28)**

n = 17 r^2 = 0.649 q^2 = 0.491 s = 2.729
F = 13.0 All p≤0.006

EACM (EURO) = 22.97 – 0.0794 Vacc22/23 + 0.534 Temperature **(29)**

n = 17 r^2 = 0.595 q^2 = 0.423 s = 2.932
F = 10.3 All p≤0.016

EACM (OECD) = 644 – 9.21 VaccJ22 + 21.29 Temperature **(30)**

n = 18 r^2 = 0.760 q^2 = 0.675 s = 52.74
F = 23.8 All p<0.001

EACM (OECD) = 629 – 8.85 VaccD22 + 20.17 Temperature **(31)**

n = 18 r^2 = 0.743 q^2 = 0.652 s = 54.62
F = 21.7 All p<0.001

EACM (COLLAB) = 1611 – 17.96 HDI + 6.51 Smoking **(32)**

n = 19 r^2 = 0.820 q^2 = 0.749 s = 41.04
F = 36.4 All p≤0.010

EACM (ECON) = 1652 – 4.803 VaccJ22 – 12.58 HDI **(33)**

n = 19 r^2 = 0.848 q^2 = 0.782 s = 30.96
F = 44.6 All p<0.001

EACM (ECON) = 1103 – 5.507 VaccJ22 – 77.6 Happiness **(34)**

n = 19 r^2 = 0.838 q^2 = 0.746 s = 31.92
F = 41.5 All p<0.001

EACM (WHO) = 473.5 – 6.550 VaccJ22 + 12.22 Temperature **(35)**

n = 19 r^2 = 0.808 q^2 = 0.753 s = 29.21
F = 41.5 All p<0.001

EACM (WHO) = 567.0 – 6.356 VaccD22 + 11.49 Temperature **(36)**

n = 19 r^2 = 0.799 q^2 = 0.744 s = 29.86
F = 31.8 All p<0.001

The above models indicate that latitude N is a proxy mainly for exercise and temperature, and possibly for happiness, all of which correlate well with latitude N (r=0.733, -0.874 and 0.737 respectively). In order to reduce EACM, governments and other bodies must stress the vital importance of exercise for everyone. That would, or course, lead also to better health in general, thereby lowering the cost of illness and other morbidities.

It can be seen from equation 32 that, the Collaborators data yielded a different model, again indicating that those data are measuring something different from the other data-sets. Note that we were able to obtain only one good QSAR model with The Collaborators data-set.

The Collaborators [15] used a Least Absolute Shrinkage and Selection Operator (LASSO) regression to select a list of contributory factors that "have sensible direction of effect on the excess mortality rate". They selected 15 such factors, seven of which are similar or identical to the factors that we have used, namely CVD death rate (positive), healthcare access and quality index (negative), mobility (positive), proportion of population over age 75 (positive), smoking prevalence (positive), UHC (negative), and average absolute latitude (positive). These yielded a model that accounted for 69.1% (i.e. r^2=0.691) of the variation in excess all-cause mortality, which is good considering that their study involved many more countries and regions than does our study. Our latitude N factor had a negative sign, but the difference in sign is probably because The Collaborators' study included a large number of countries covering a wide range of latitudes, some of which were south of the equator.

A review published by the British Office for National Statistics [39] reviewed a number of published studies of possible factors affecting excess all-cause mortality during the Covid-19 pandemic. They reported that cardiovascular diseases, diabetes, chronic obstructive pulmonary disease, dementia, obesity, smoking, age and ethnicity had all been found to affect excess all-cause mortality. Whilst our models for The Collaborators' data-set (equations 18 and 19) incorporate some of the above factors, it is striking that our models for the other excess all-cause mortality data-sets do not incorporate any of them.

Conclusion

The three outcomes that we have attempted to model have yielded very variable levels of success. Numbers of cumulative cases produced only two acceptable equations, but numbers of cumulative deaths generated far more, although the factors entering into these were rather variable, and in a number of models the signs on coefficients were counter-intuitive, perhaps indicating that data error could be responsible. Excess deaths were modelled in a far more satisfactory way, and showed a strong and consistent influence of vaccination levels, as well as of the proxy factor latitude N. Intuitively, this pattern probably accords with the relative difficulties in obtaining reliable and consistent measures of the three outcomes. Numbers of cumulative cases are likely to be particularly problematic at later time points, as governments had little need to record these once the lethality of the disease had declined. Numbers of cumulative deaths would have been recorded more accurately, but ascertaining numbers specifically due to Covid-19 is more problematic. Excess deaths data should be the most reliable as they only require total numbers of deaths in the Covid-affected years and in preceding years. Our analysis leads us to conclude that excess all-cause mortality is controlled largely by vaccination levels and factors such as exercise and temperature.

To reduce the impact of any future respiratory pandemics, great improvements should be made in levels of exercise, and in reduction of cardiovascular diseases and levels of pollution. Data collection needs to be improved, in order to follow the progress of a pandemic more accurately.

Declarations

Disclaimer: The submitted article is our own, and is not an official position of our institution.

References

1. WHO declares Covid-19 a pandemic: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-a-pandemic

2. Wikipedia: Spanish flu: https://en.wikipedia.org/wiki/Spanish_flu.

3. Ilie PC, Stefanescu S, Smith L. The role of vitamin D in the prevention of coronavirus disease-2019 infection and mortality. Aging Clin. Exp. Res. 2020; 32: 1195–1198.

4. Dearden JC. The history and development of quantitative structure-activity relationships (QSARs). Int J Quant Struct Prop Relat. 2016; 1: 1-43.

5. Dearden JC, Rowe PH. Correlation between vitamin D levels, individual and socio-demographic characteristics and COVID-19 infection and death rates in 20 European countries: a modelling study. J Health Soc Sci. 2020; 5: 513-524.

6. Starnini M, Aleta A, Tizzoni M, Moreno Y. Impact of data accuracy on the evaluation of COVID-19 mitigation policies. Data & Policy 2021; 3: e28.

7. Kobak D. Underdispersion: A statistical anomaly in reported Covid data. Significance. 2022: 10-13.

8. Karlinsky A, Kobak D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. eLife. 2021: 10e69336.

9. Giattino C, Ritchie H, Ortiz-Ospina E, Hasell J, Rodés-Guirao, Roser M. Excess mortality during the Coronavirus pandemic (COVID-19): https://ourworldindata.org/excess-mortality-covid.

10. Wikipedia Covid-19 cases and deaths: https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic.

11. Office for National Statistics (ONS): community/birthsdeathsandmarriages/deaths/articles/comparisonofallcausemortalitybetweeneuropeancountriesandregions/datauptowekending3September202110.

12. Our World in Data: https://ourworldindata.org/grapher/covid-vaccination-doses-per-capita?tab=table.

13. Eurostat: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/DDN-20230217-1

14. OECD: https://oecd.org/coronavirus/en/data-insights/excess-mortality-since-January-2020.

15. The Collaborators (98 authors). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. The Lancet. 2022; 399: 1513-1536.

16. The Economist: https://www.economist.com/graphic-detail/coronavirus-excess-deaths-Tracker.

17. World Health Organization: https://2022-wpds.prb.org/wp.World-Population-Data-Sheet-Booklet.pdf.

18. Health Development Index: https://hdr.undp.org/data-center/human-development-index#/indicies/HDI.

19. Universal Health Coverage: https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc).

20. Wikipedia Lockdown Lengths: https://en.wikipedia.org/wiki/COVID-19_lockdowns.

21. VaccJ22: https://www.bbc.co.uk/news/world-51235105.

22. VaccD22: https://ig.ft.com/coronavirus-vaccine-tracker.

23. Vacc22/23: https://ourworldindata.org/grapher/covid-vaccination-doses-per-capita?tab=table.

24. Minitab software: https://minitab.com.

25. Topliss JG, Costello RJ. Chance correlations in structure-activity studies using multiple linear regression. J Med Chem. 1972; 15: 1066-1068.

26. Vaccine Tracker: https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#uptake-tab.

27. Box GEP, Cox DR. An analysis of transformations. JR Stat Soc Ser B Stat Methodol. 1964; 26: 211-252.

28. Starnini M, Aleta A, Tizzoni M, Moreno Y. Impact of data accuracy on the evaluation of COVID-19 mitigation policies. Data & Policy. 2021; 3: e28.

29. Costa-Santos C, Neves AL, Correia R, Santos P, Monteiro-Soares M, Freitas A, et al. COVID-19 surveillance data quality issues: a national consecutive case series. BMJ Open. 2021; 11: e047623.

30. Balsari S, Buckee C, Khanna T. Which Covid-19 data can you trust? Harvard Bus. Rev: https://hbr.org/2020/05/which-covid-19-data-can-you-trust.

31. Whittaker C, Walker PGT, Alhaffar M, Hamlet A, Ghani A, Ferguson M, et al. Under-reporting of deaths limits our understanding of true burden of covid-19. BMJ. 2021; 375: n2239.

32. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Sci Adv. 2020; 6: eabd4049.

33. Li X, Guan B, Su T, Liu W, Chen M, Waleed KB, et al. Impact of cardiovascular disease and cardiac injury on in-hospital mortality in patients with COVID-19: a systematic review and meta-analysis. Heart. 2020; 106: 1142-1147.

34. Ssentongo P, Ssentongo AE, Heilbrunn ES, Ba DM, Chinchilli VM. Association of cardiovascular disease and 10 other pre-existing comorbidities with COVID-19 mortality: a systematic review and meta-analysis. PLoS One. 2020; 15: e0238215.

35. The Bible, Proverbs chapter 17, verse 22. New International Version, International Bible Society, 1984.

36. Happiness levels by country: https://www.theglobaleconomy.com/rankings/happiness/.

37. Gasparrini A, Guo Y, Sera F, Vicedo-Cabrera AM, Huber V, Coelho MSZS, et al. Projection of temperature-related excess mortality under climate change scenarios. Lancet Planet Health. 2017; 1: e360-e367.

38. Mean temperatures of European countries: https://www.bing.com/search?q=mean+temperatures+of+european+countries&form=ANSPH1&refig=f944b0af1be4472ebbb7e547010c59a&pc=EDGEXST.

39. British Office for National Statistics. International comparisons of possible factors affecting excess mortality: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/articles/internationalcomparisonsofpossiblefactorsaffectingexcessmortality/2022-12-20.