On the retrospective analysis of COVID-19 macroscopic epidemiological data

Manuel Graña
University of the Basque Country

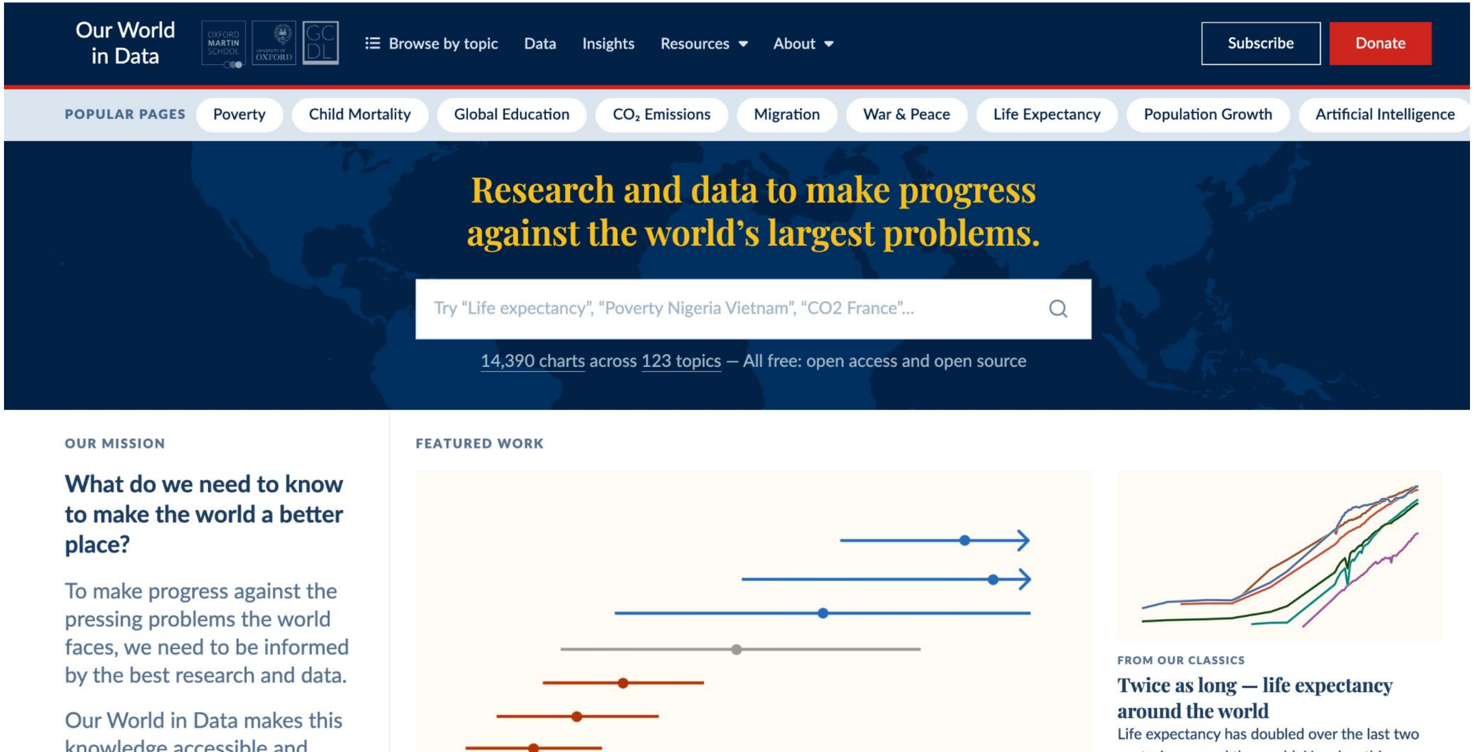
Contents

Motivation

- After five years we can look back, can’t we?
- Posing question rather than giving answers
- Preparedness for coming pandemics
- The war on virus

Data source

Mostly the “Our World in Data” site of the University of Oxford



Research and data to make progress against the world’s largest problems.

Try “Life expectancy”, “Poverty Nigeria Vietnam”, “CO2 France”...

14,390 charts across 123 topics — All free: open access and open source

OUR MISSION

What do we need to know to make the world a better place?

To make progress against the pressing problems the world faces, we need to be informed by the best research and data.

Our World in Data makes this knowledge accessible and

FEATURED WORK

FROM OUR CLASSICS

Twice as long — life expectancy around the world

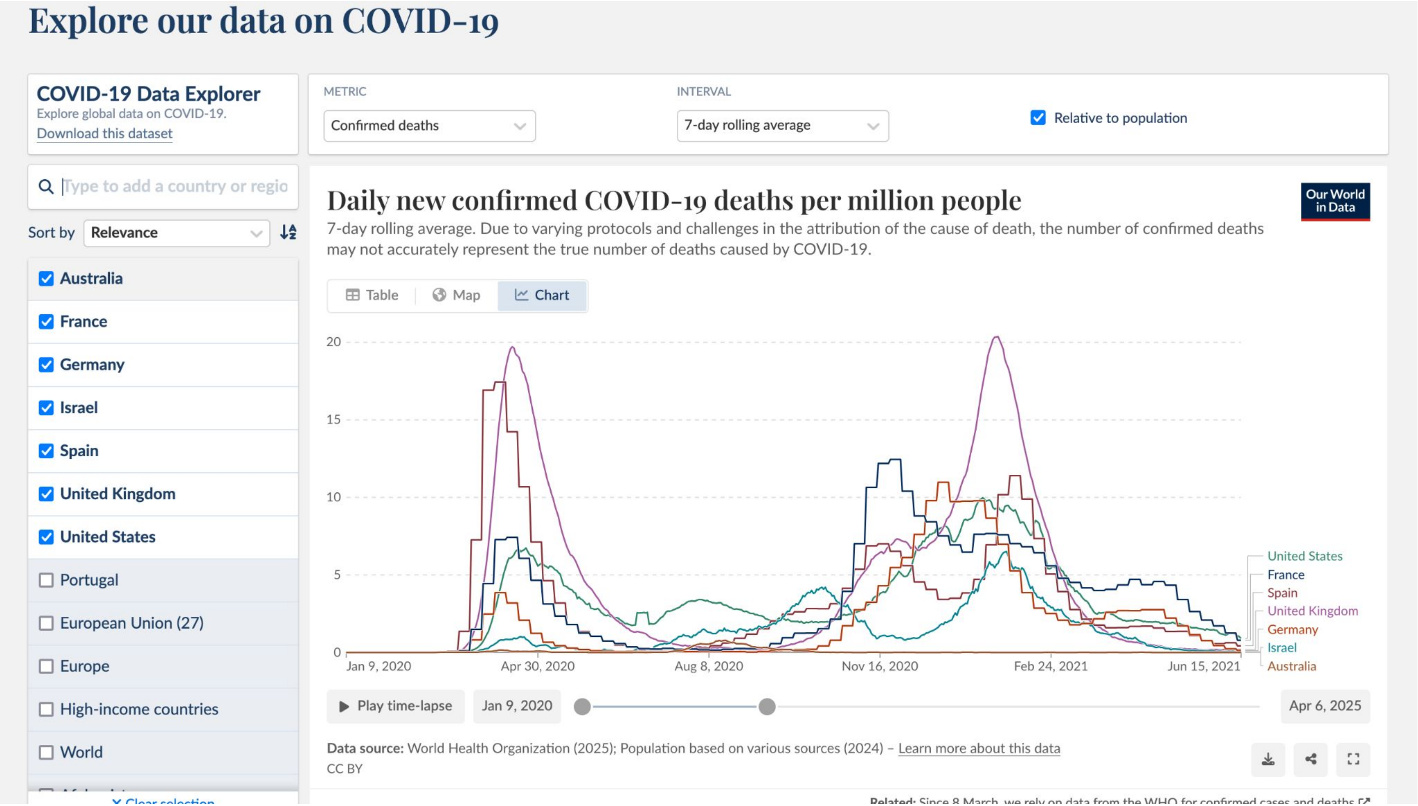
Life expectancy has doubled over the last two

The first wave

The first wave (mortality)

- In the spring of 2020 several countries suffered (strong) mortality peaks,
- Others didn’t
- synchronization of peaks seemed independent of geographical distance

first wave



Explore our data on COVID-19

COVID-19 Data Explorer
Explore global data on COVID-19.
Download this dataset

METRIC: Confirmed deaths

INTERVAL: 7-day rolling average

☑ Relative to population

Type to add a country or regio

Sort by: Relevance

☑ Australia
☑ France
☑ Germany
☑ Israel
☑ Spain
☑ United Kingdom
☑ United States
☐ Portugal
☐ European Union (27)
☐ Europe
☐ High-income countries
☐ World

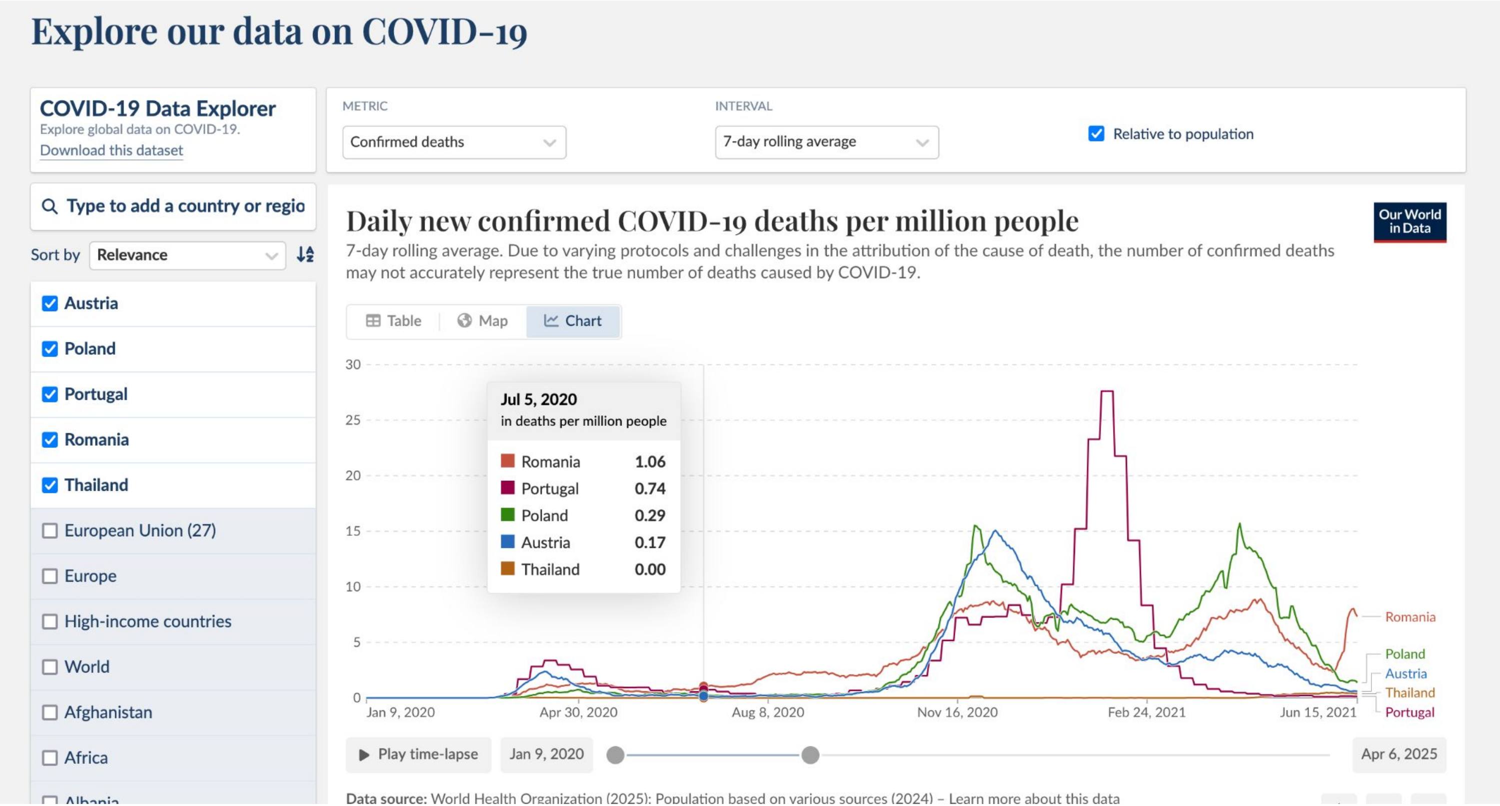
Daily new confirmed COVID-19 deaths per million people

7-day rolling average. Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.

Data source: World Health Organization (2025); Population based on various sources (2024) – Learn more about this data
CC BY

Related: Since 8 March, we rely on data from the WHO for confirmed cases and deaths

first wave



Explore our data on COVID-19

COVID-19 Data Explorer
Explore global data on COVID-19.
Download this dataset

METRIC: Confirmed deaths

INTERVAL: 7-day rolling average

☑ Relative to population

Type to add a country or regio

Sort by: Relevance

- ☑ Austria
- ☑ Poland
- ☑ Portugal
- ☑ Romania
- ☑ Thailand
- ☐ European Union (27)
- ☐ Europe
- ☐ High-income countries
- ☐ World
- ☐ Afghanistan
- ☐ Africa
- ☐ Albania

Daily new confirmed COVID-19 deaths per million people

7-day rolling average. Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.

Jul 5, 2020
in deaths per million people

🟥 Romania	1.06
🟥 Portugal	0.74
🟩 Poland	0.29
🟦 Austria	0.17
🟧 Thailand	0.00

Data source: World Health Organization (2025); Population based on various sources (2024) – Learn more about this data

First wave

In January 2020, the World Health Organization introduced a disease code specific for COVID-19

Some countries did use this codification since the early stages of the pandemic, others lagged behind

In some countries, COVID-19 cases were paid more than other respiratory diseases, the hospital administration was rewarded for using the COVID-19 disease code

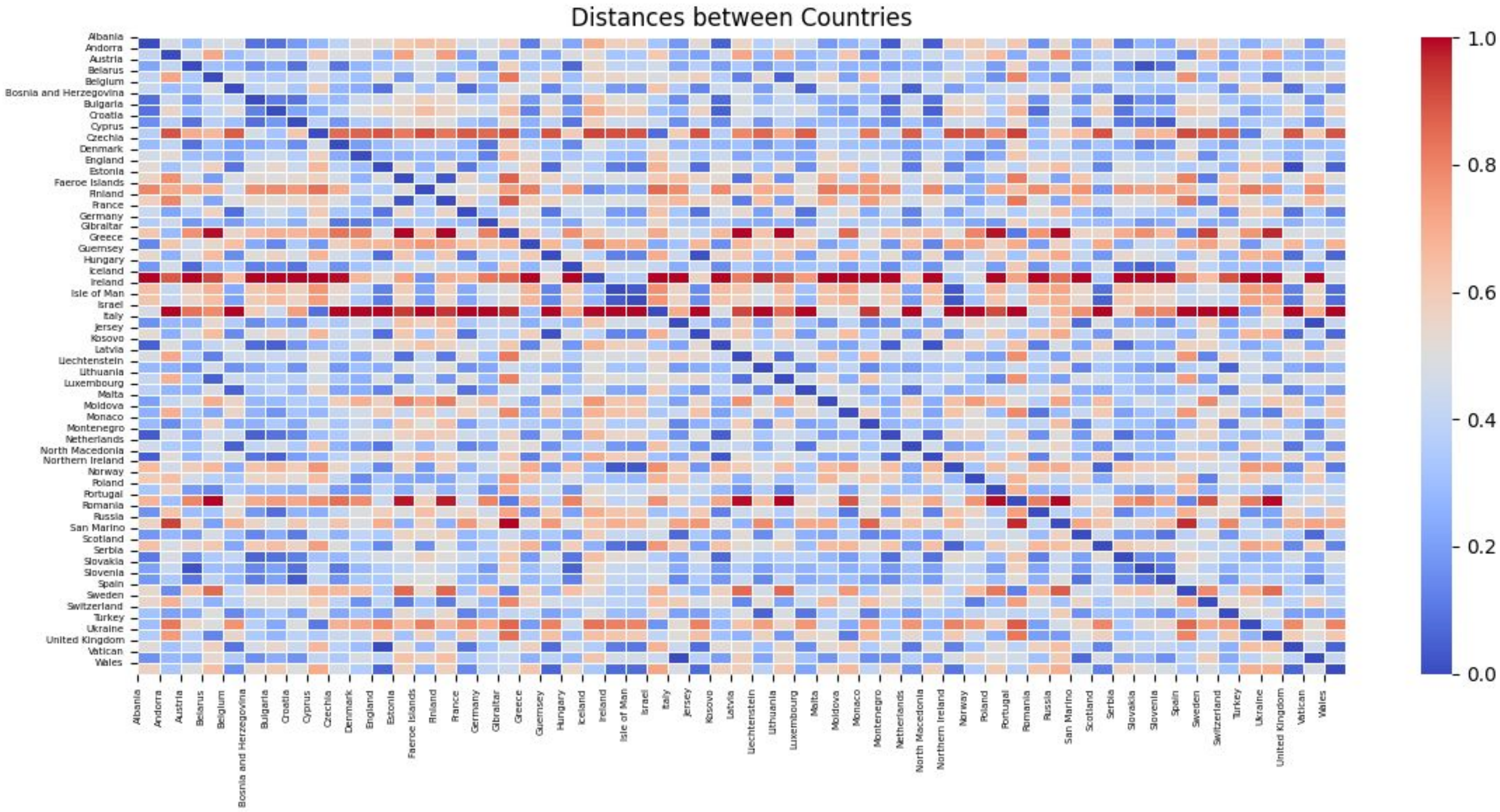
Propagation of the virus in Europe in 2020

Propagation of the virus in Europe in 2020

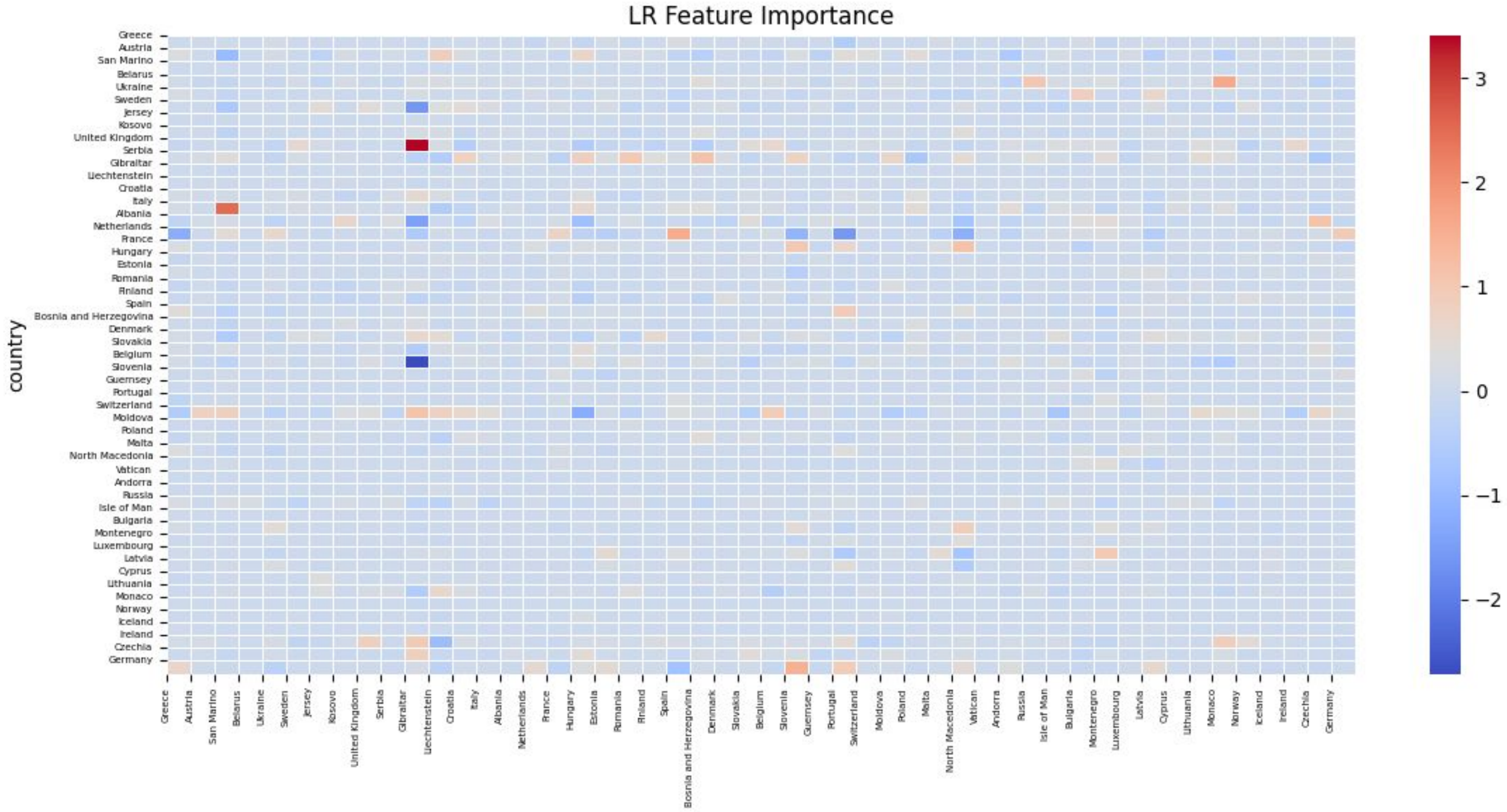
We have considered the mortality in European countries

We correlate similarity measures among mortality time series with distances among country capital cities

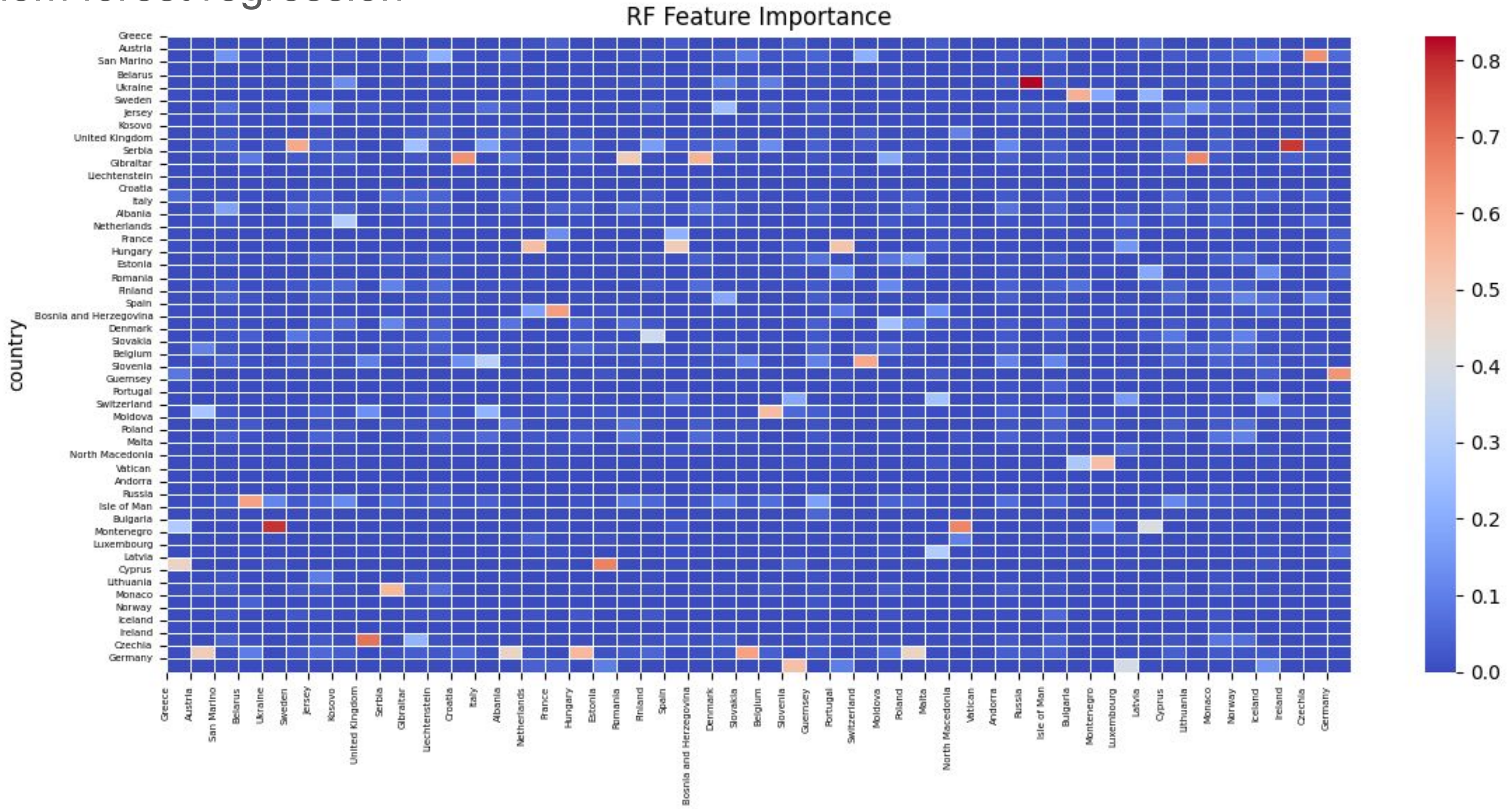
The idea is that more distant countries should be less similar


Distances between Countries
Albania
Andorra
Austria
Belarus
Belgium
Bosnia and Herzegovina
Bulgaria
Croatia
Cyprus
Czechia
Denmark
England
Estonia
Faeroe Islands
Finland
France
Germany
Gibraltar
Greece
Guernsey
Hungary
Iceland
Ireland
Isle of Man
Israel
Italy
Jersey
Kosovo
Latvia
Liechtenstein
Lithuania
Luxembourg
Malta
Moldova
Monaco
Montenegro
Netherlands
North Macedonia
Northern Ireland
Norway
Poland
Portugal
Romania
Russia
San Marino
Scotland
Serbia
Slovakia
Slovenia
Spain
Sweden
Switzerland
Turkey
Ukraine
United Kingdom
Vatican
Wales
1.0
0.8
0.6
0.4
0.2
0.0

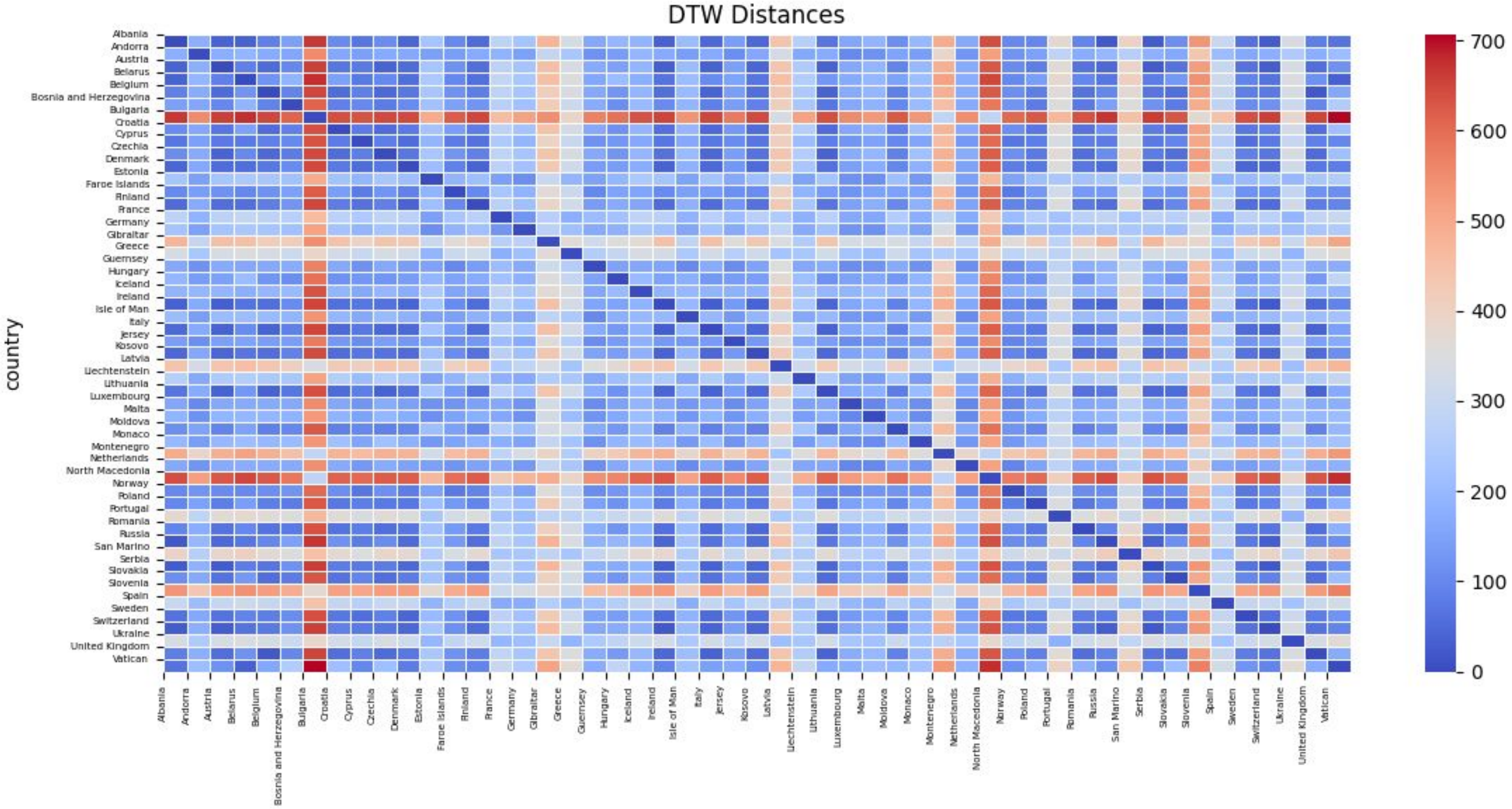
significance of each country when trying to predict deaths in other countries by Linear regression



significance of each country when trying to predict deaths in other countries by random forest regression



dynamic time warping distances among death time series



Propagation of the virus in Europe in 202

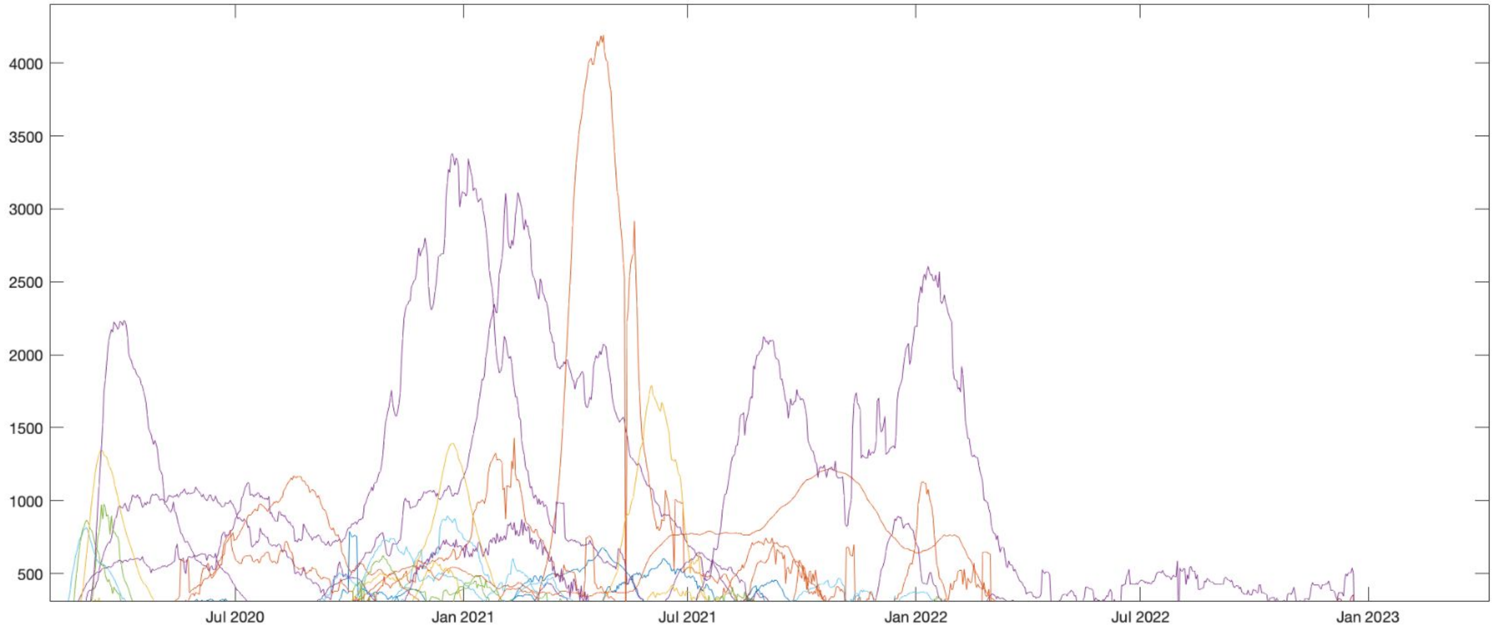
- Pearson’s Correlation Coefficients found against geographical distance:
    - DTW distance r ≈ -0.60
    - Random Forest regression importance r ≈ -0.50
    - Linear regression importance r ≈ -0.35
- Conclussion: there is no influence of geographical distance on the similarity of the death time series
- Is there any viral diffusion process?

Clustering of countries

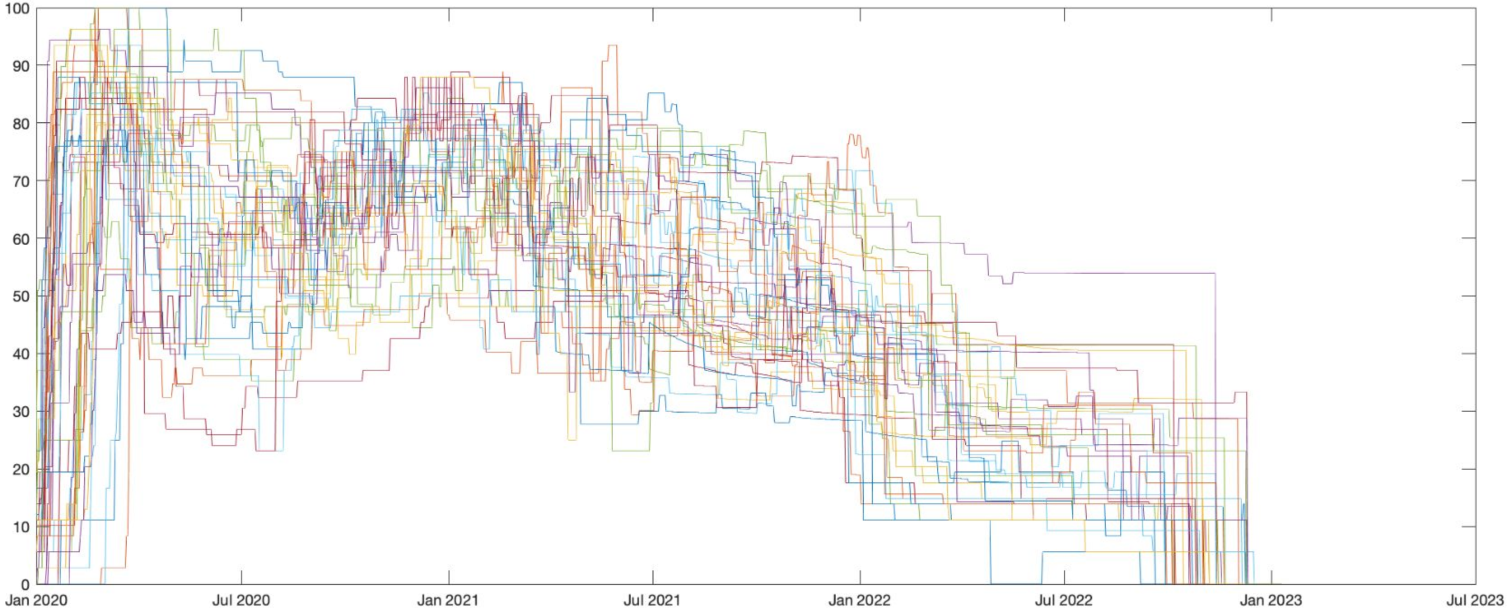
Clustering of countries

- We compute the Dynamic Time Warping (DTW) similarity among the main epidemiological variables of countries
    - death time series
    - stringency index
    - vaccination
    - we select countries with the highest relative accumulative death toll
- We carry out hierarchical clustering of DTW distances
- we find relevant clusters of countries with similar policies and results

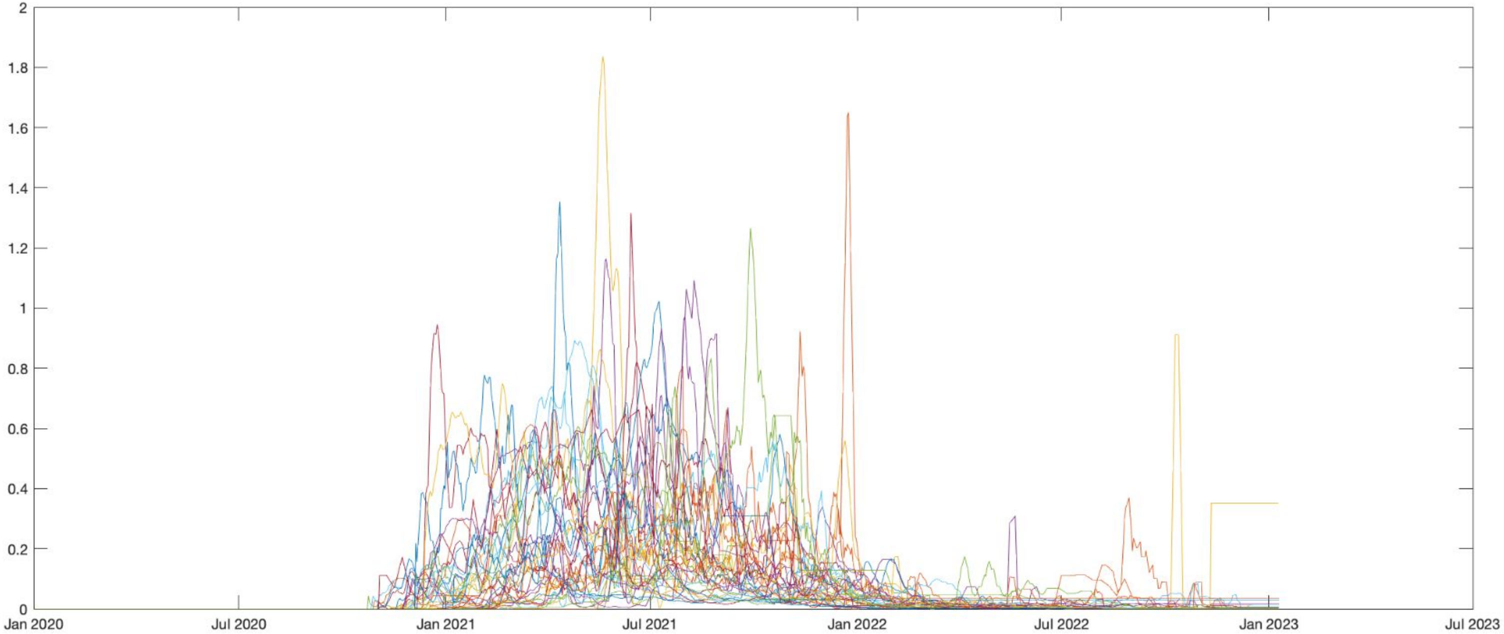
death time series of selected countries

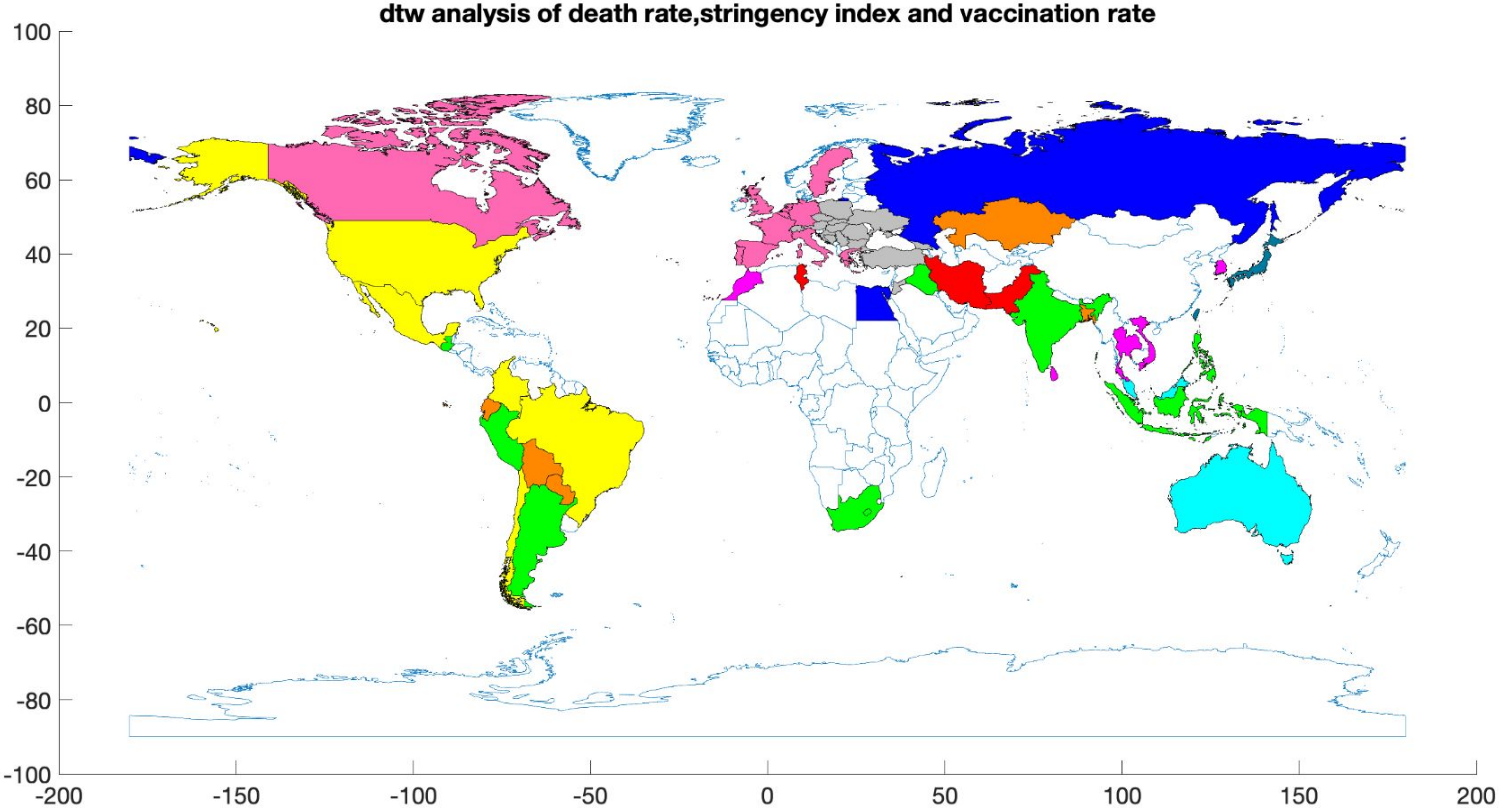


stringency index time series of selected countries



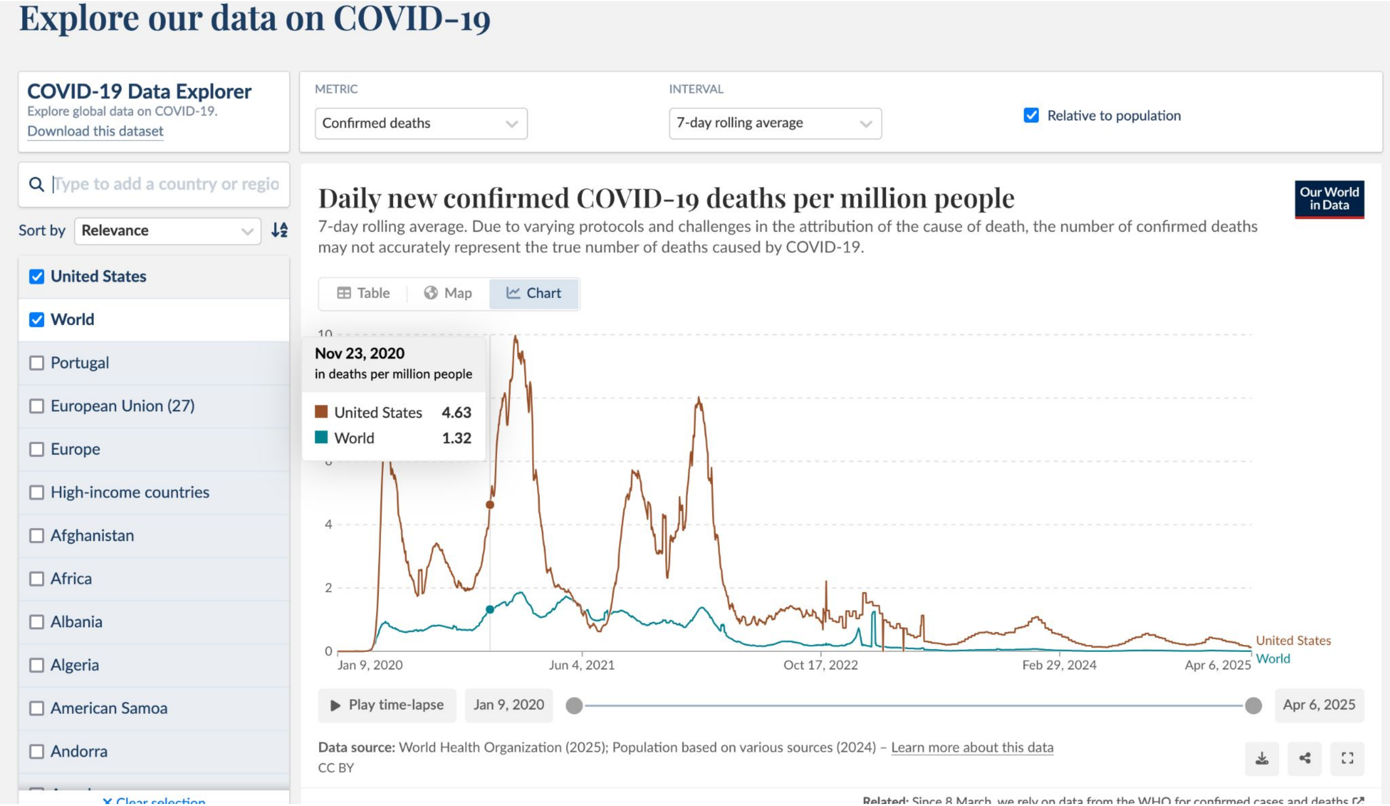
vaccine doses time series of selected countries




dtw analysis of death rate,stringency index and vaccination rate
100
80
60
40
20
0
-20
-40
-60
-80
-100
-200
-150
-100
-50
0
50
100
150
200

The case of USA

The case of USA



Explore our data on COVID-19

COVID-19 Data Explorer
Explore global data on COVID-19.
Download this dataset

METRIC: Confirmed deaths

INTERVAL: 7-day rolling average

☑ Relative to population

Daily new confirmed COVID-19 deaths per million people

7-day rolling average. Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.

Nov 23, 2020
in deaths per million people

United States	4.63
World	1.32

Data source: World Health Organization (2025); Population based on various sources (2024) – Learn more about this data
CC BY

Related: Since 8 March, we rely on data from the WHO for confirmed cases and deaths

Health, Pharma & Medtech › Pharmaceutical Products & Market

Prescription drug expenditure in the United States from 1960 to 2024

(in billion U.S. dollars)

Details: United States; CMS (Office of the Actuary)

© Statista 2025

DOWNLOAD

Source
CMS

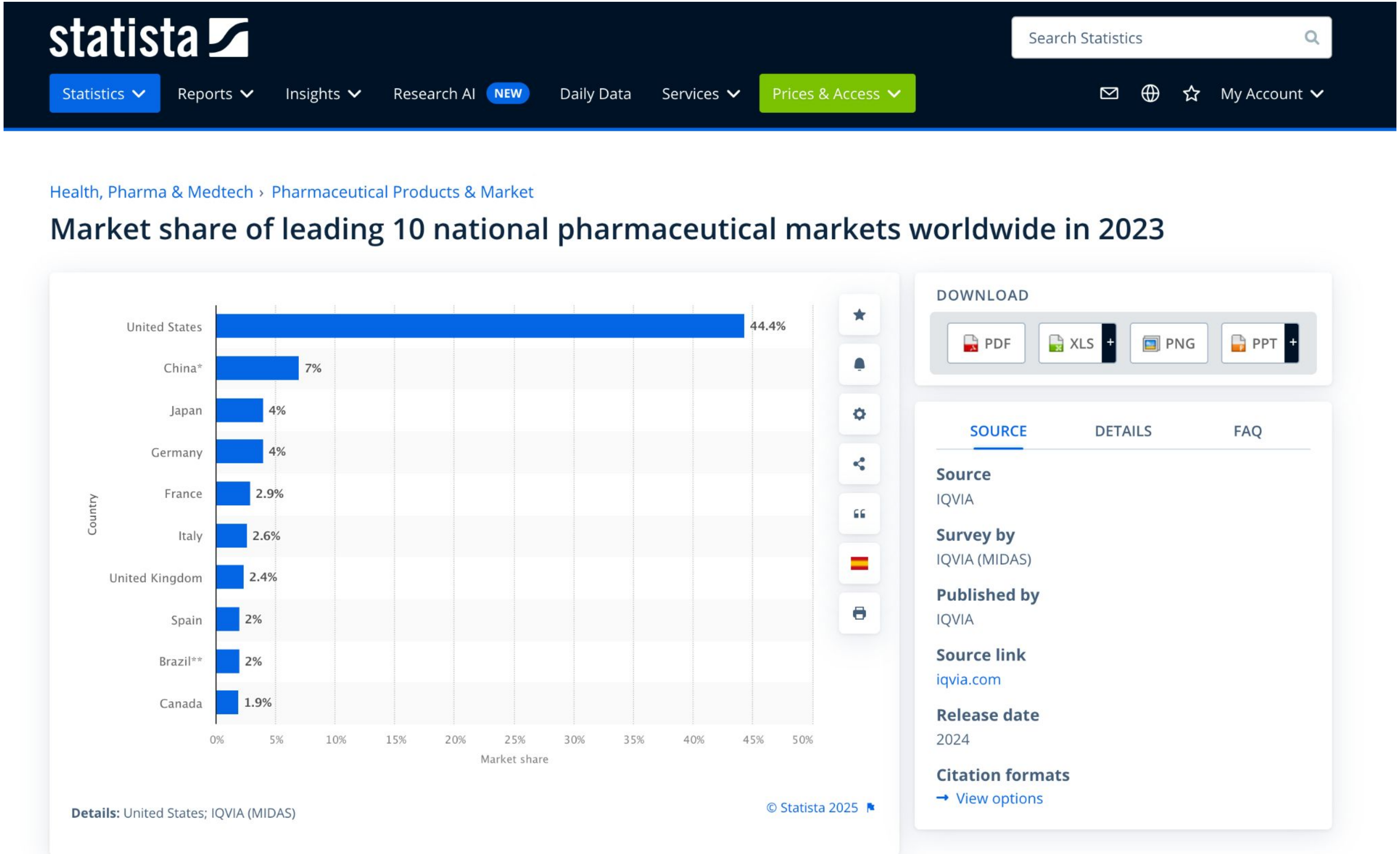
Survey by
CMS (Office of the Actuary)

Published by
CMS

Source link
cms.gov

Release date
2024

Citation formats
→ View options



Health, Pharma & Medtech › Pharmaceutical Products & Market

Market share of leading 10 national pharmaceutical markets worldwide in 2023

Details: United States; IQVIA (MIDAS)

Source
IQVIA

Survey by
IQVIA (MIDAS)

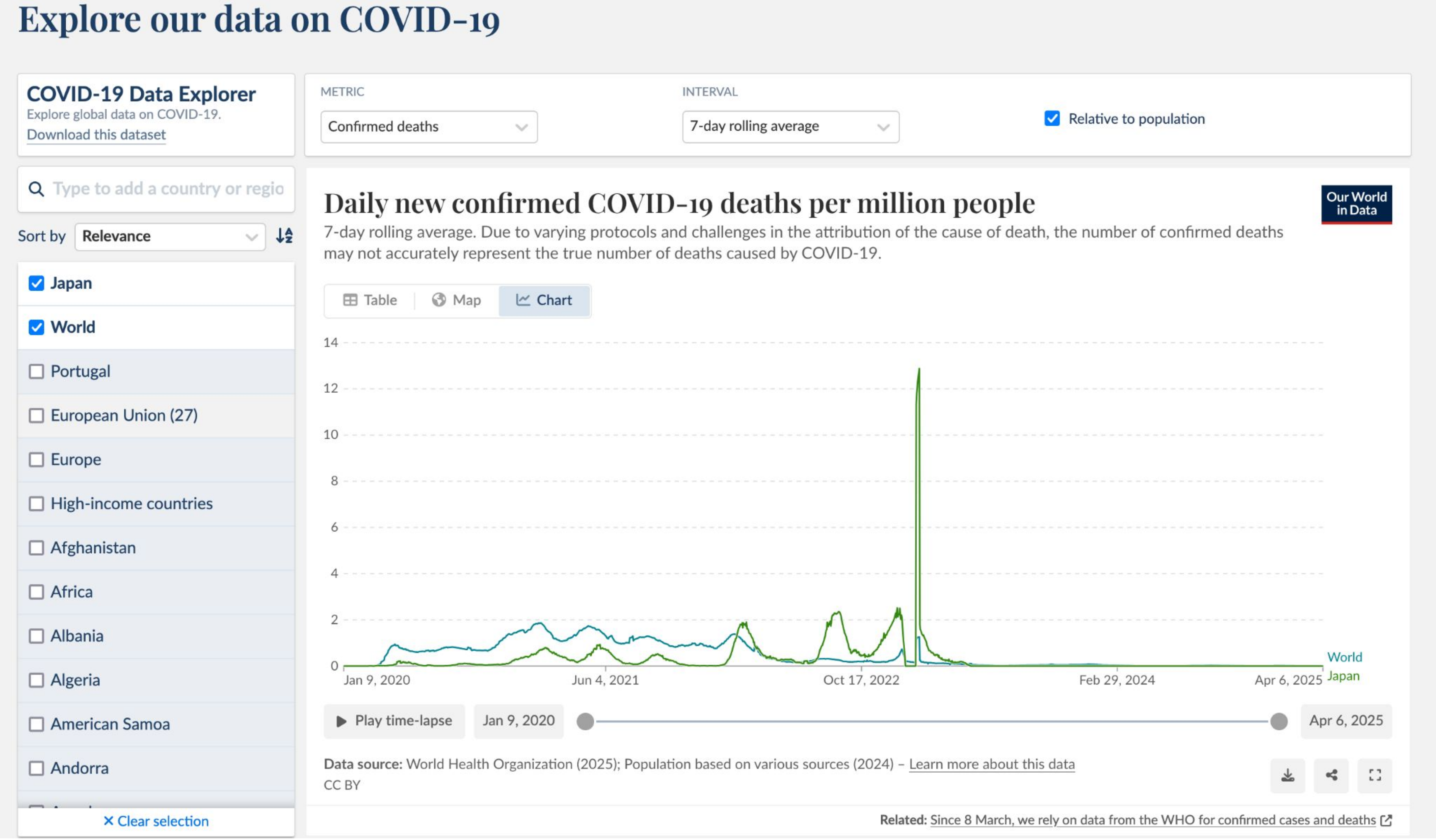
Published by
IQVIA

Source link
iqvia.com

Release date
2024

Citation formats
→ View options

The case of Japan



Explore our data on COVID-19

COVID-19 Data Explorer

Explore global data on COVID-19.

Download this dataset

METRIC: Confirmed deaths

INTERVAL: 7-day rolling average

☑ Relative to population

Type to add a country or regio

Sort by Relevance

- ☑ Japan
- ☑ World
- ☐ Portugal
- ☐ European Union (27)
- ☐ Europe
- ☐ High-income countries
- ☐ Afghanistan
- ☐ Africa
- ☐ Albania
- ☐ Algeria
- ☐ American Samoa
- ☐ Andorra

✕ Clear selection

Daily new confirmed COVID-19 deaths per million people

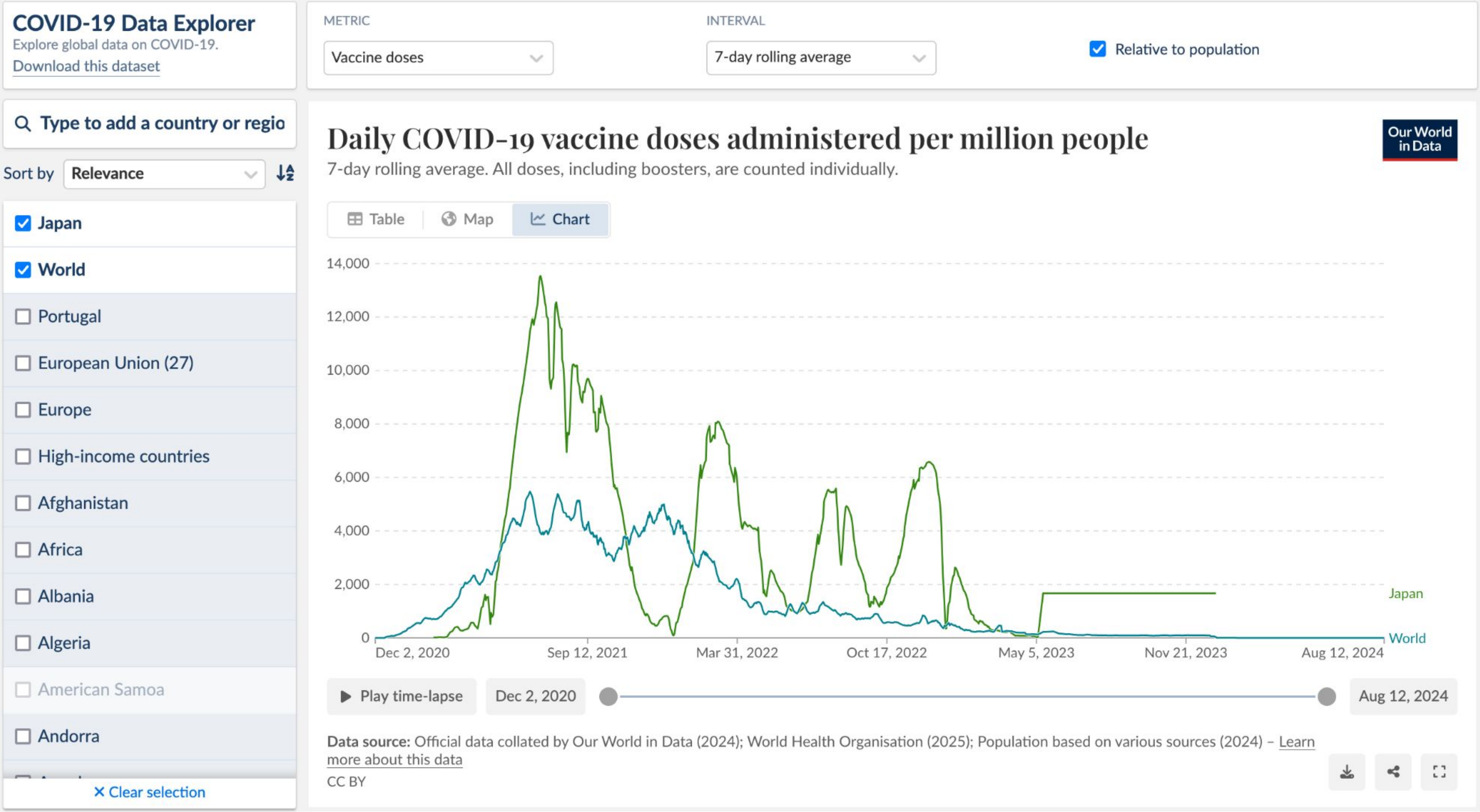
7-day rolling average. Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.

Data source: World Health Organization (2025); Population based on various sources (2024) – Learn more about this data

CC BY

Related: Since 8 March, we rely on data from the WHO for confirmed cases and deaths

Explore our data on COVID-19



COVID-19 Data Explorer

Explore global data on COVID-19.

Download this dataset

METRIC: Vaccine doses

INTERVAL: 7-day rolling average

☑ Relative to population

Type to add a country or regio

Sort by: Relevance

☑ Japan
☑ World
☐ Portugal
☐ European Union (27)
☐ Europe
☐ High-income countries
☐ Afghanistan
☐ Africa
☐ Albania
☐ Algeria
☐ American Samoa
☐ Andorra

Daily COVID-19 vaccine doses administered per million people

7-day rolling average. All doses, including boosters, are counted individually.

Data source: Official data collated by Our World in Data (2024); World Health Organisation (2025); Population based on various sources (2024) – Learn more about this data

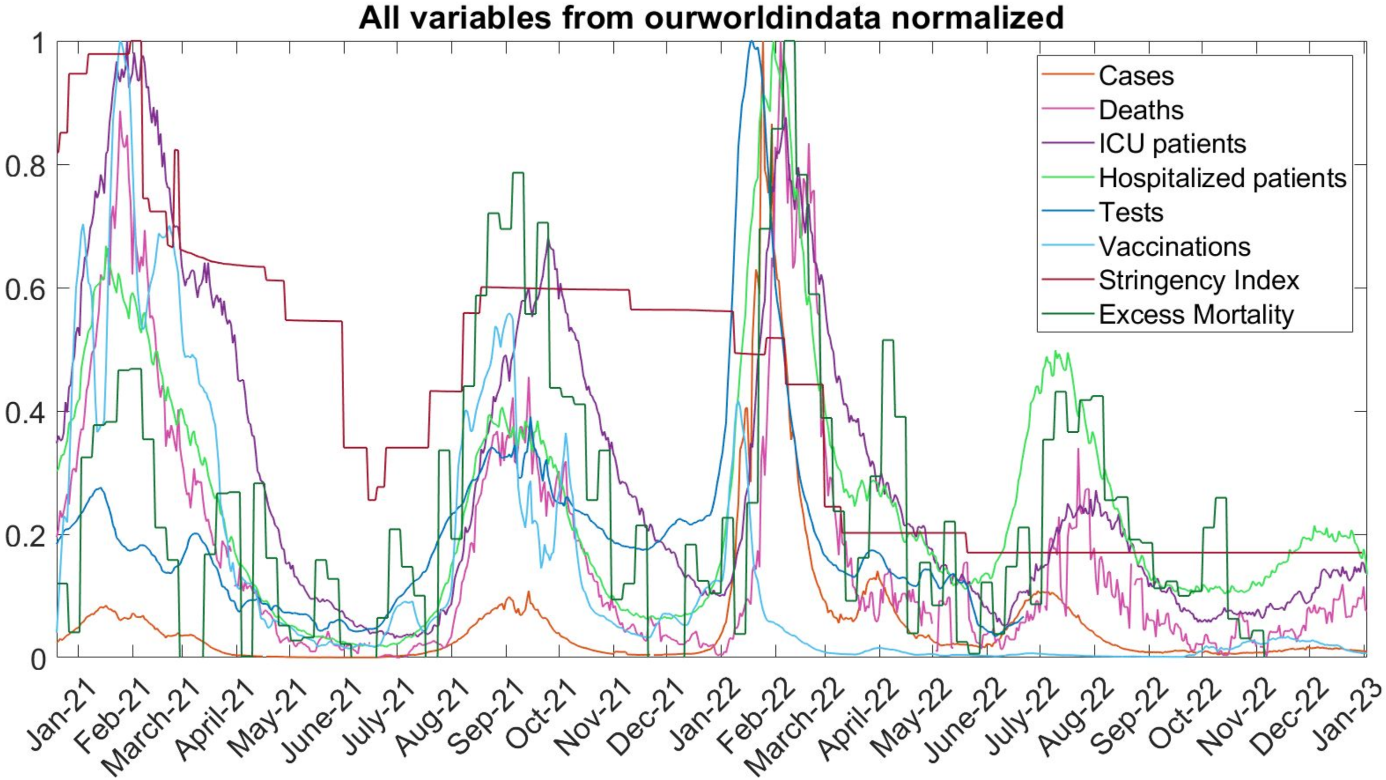
CC BY

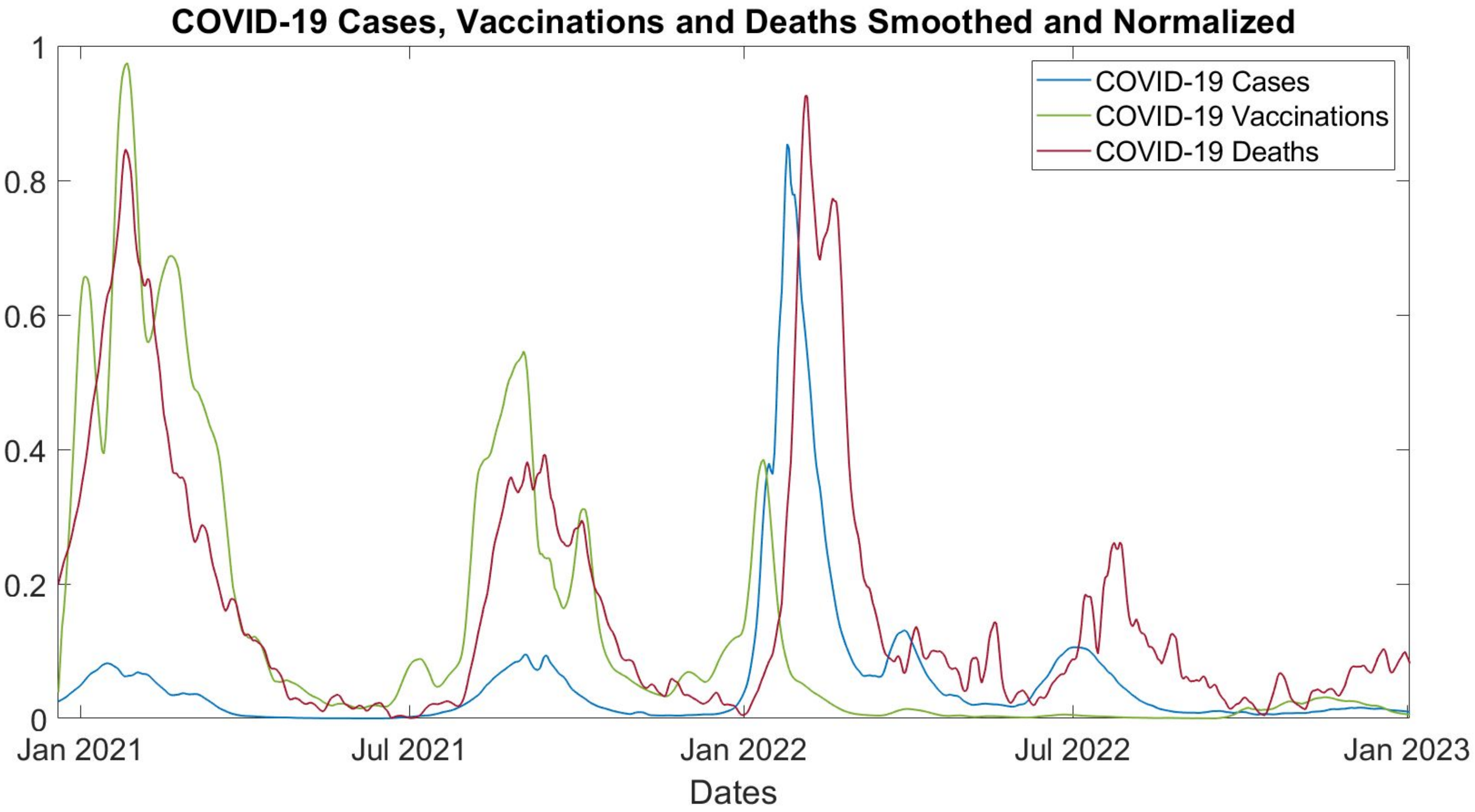
The case of Israel

The case of Israel

- Israel was the showcase of Pfizer-Biontech vaccine
- Israel was in the forefront of non-pharmaceutical measures

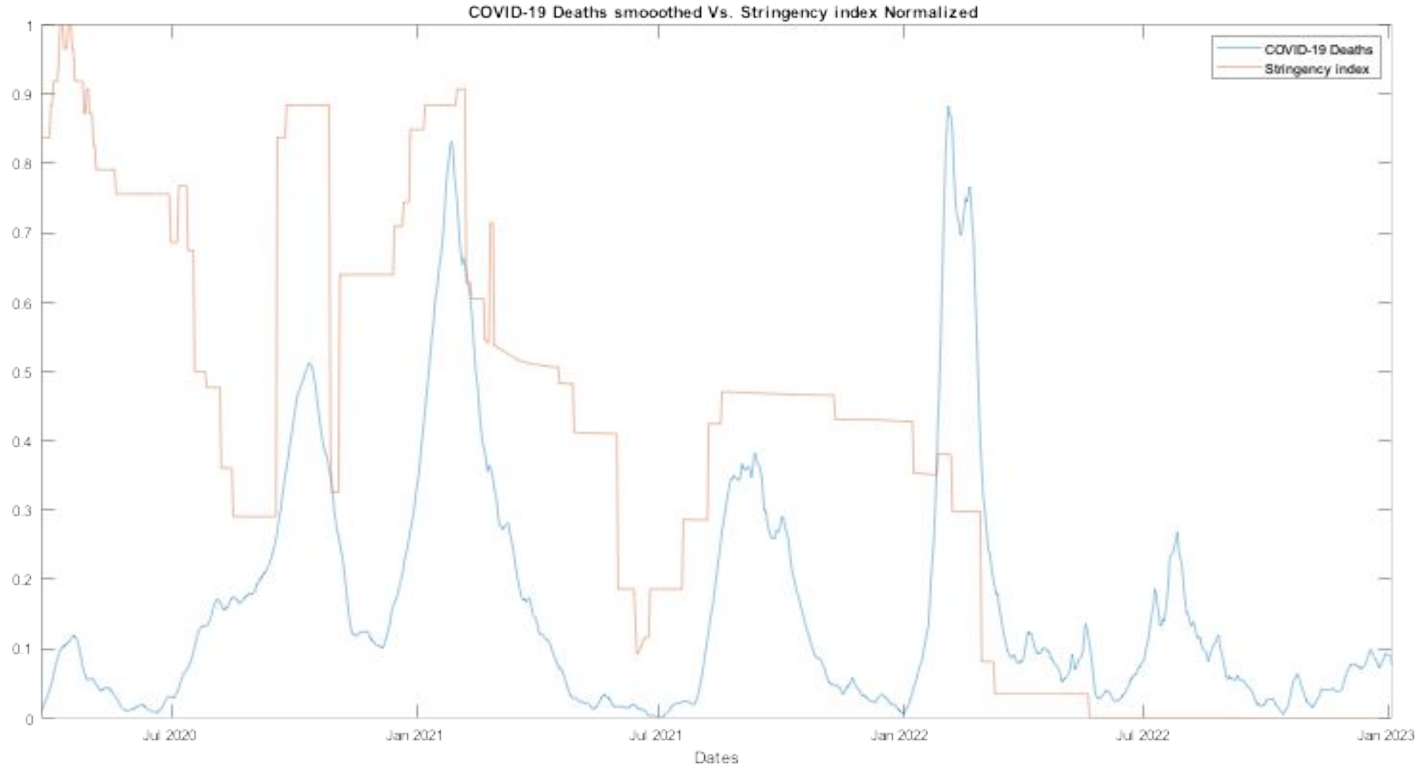
The case of israel



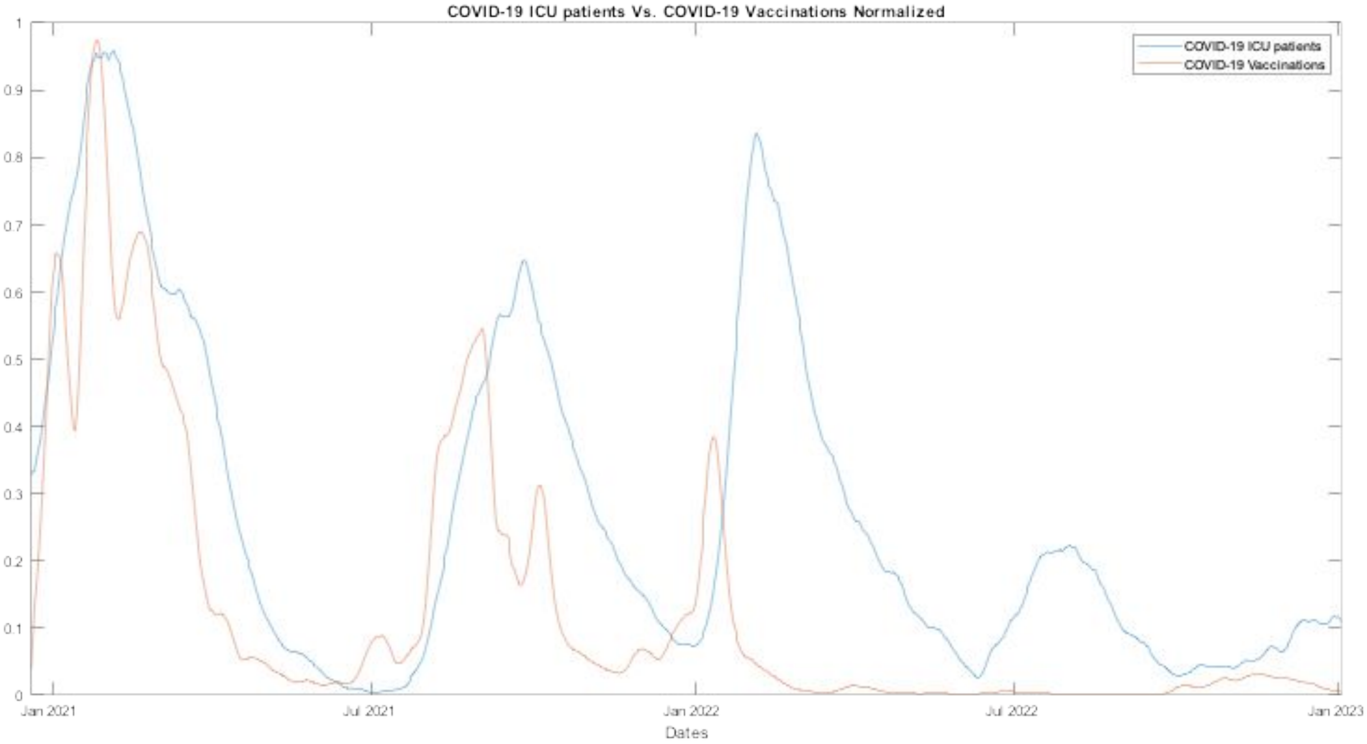

COVID-19 Cases, Vaccinations and Deaths Smoothed and Normalized
1
0.8
0.6
0.4
0.2
0
Jan 2021
Jul 2021
Jan 2022
Jul 2022
Jan 2023
Dates
COVID-19 Cases
COVID-19 Vaccinations
COVID-19 Deaths

Stringency index, deaths

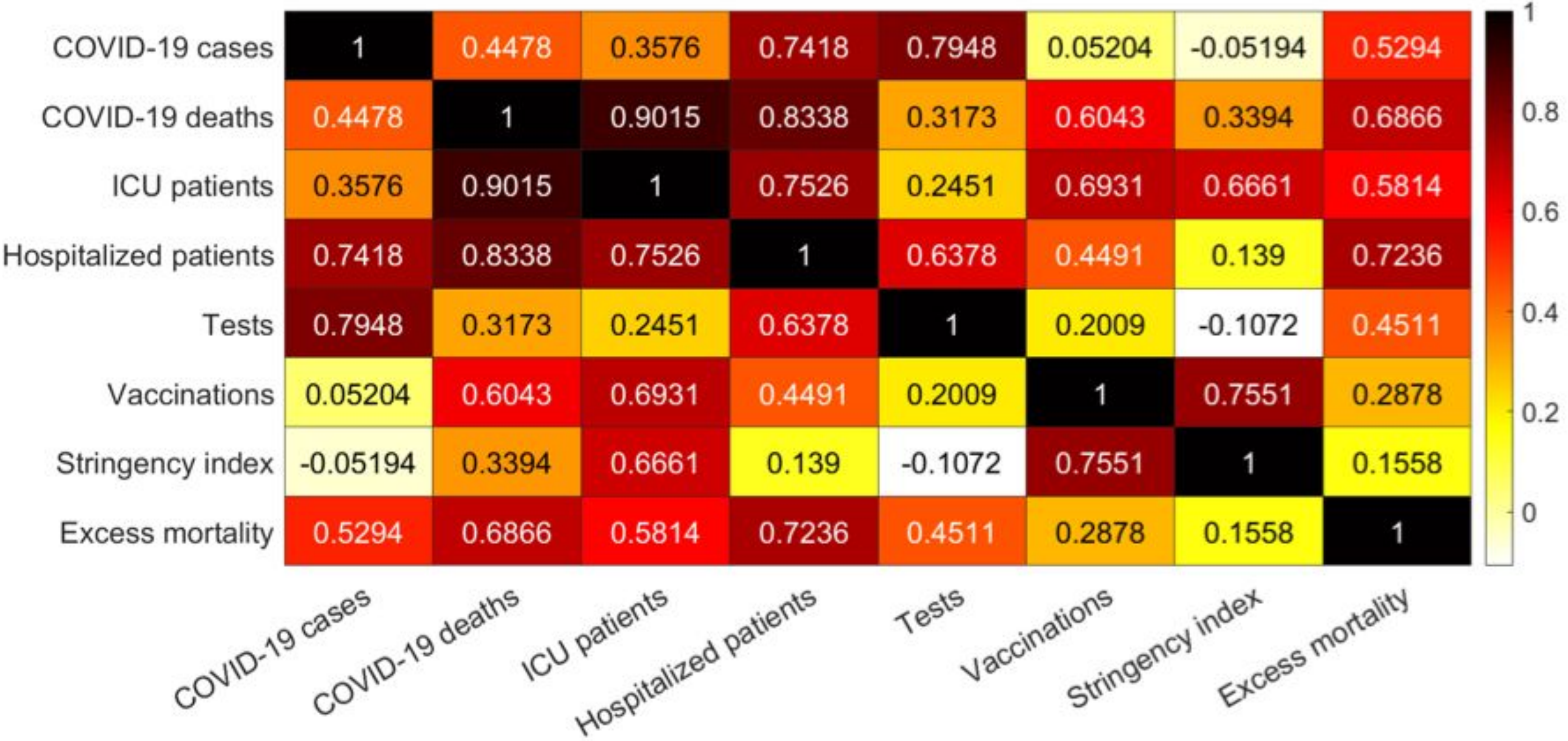
- Normalized to [0,1]



Vaccination doses, ICU patients· Normalized to [0,1]



Overall Correlation

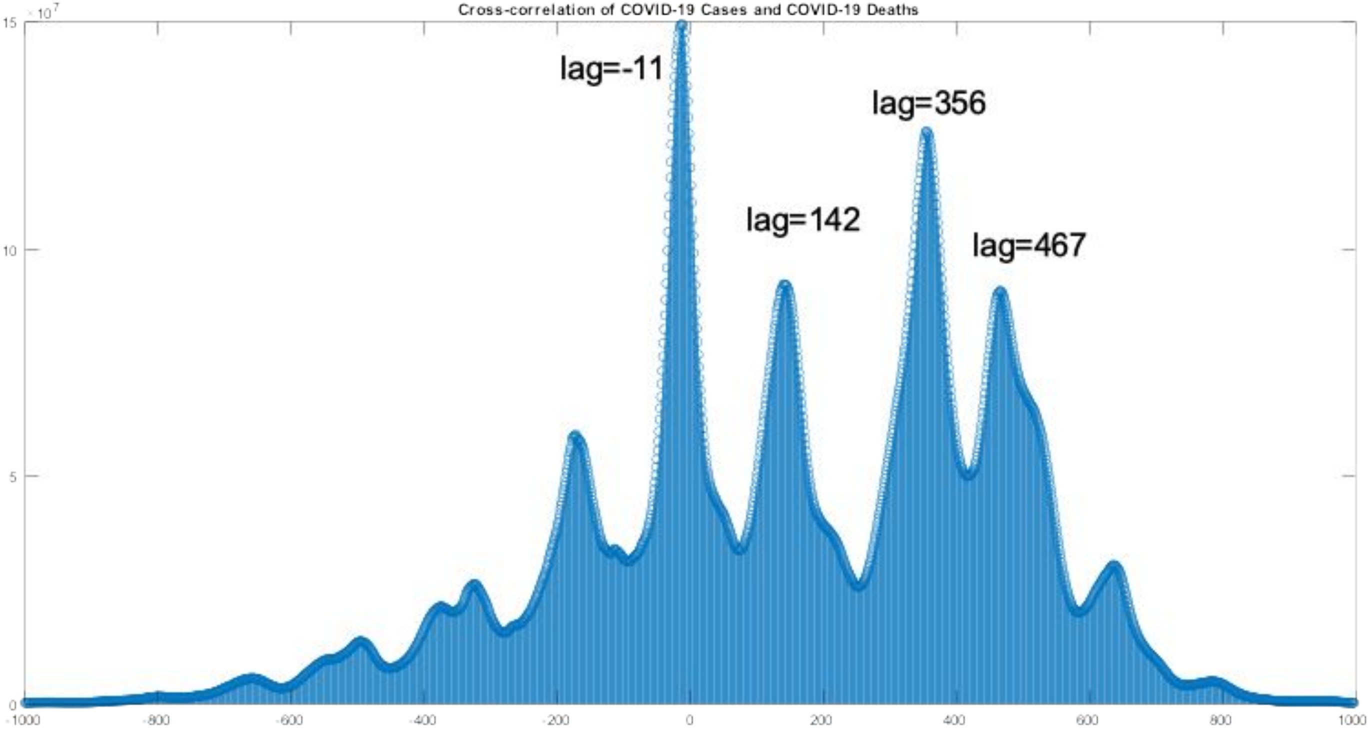


	COVID-19 cases	COVID-19 deaths	ICU patients	Hospitalized patients	Tests	Vaccinations	Stringency index	Excess mortality
COVID-19 cases	1	0.4478	0.3576	0.7418	0.7948	0.05204	-0.05194	0.5294
COVID-19 deaths	0.4478	1	0.9015	0.8338	0.3173	0.6043	0.3394	0.6866
ICU patients	0.3576	0.9015	1	0.7526	0.2451	0.6931	0.6661	0.5814
Hospitalized patients	0.7418	0.8338	0.7526	1	0.6378	0.4491	0.139	0.7236
Tests	0.7948	0.3173	0.2451	0.6378	1	0.2009	-0.1072	0.4511
Vaccinations	0.05204	0.6043	0.6931	0.4491	0.2009	1	0.7551	0.2878
Stringency index	-0.05194	0.3394	0.6661	0.139	-0.1072	0.7551	1	0.1558
Excess mortality	0.5294	0.6866	0.5814	0.7236	0.4511	0.2878	0.1558	1

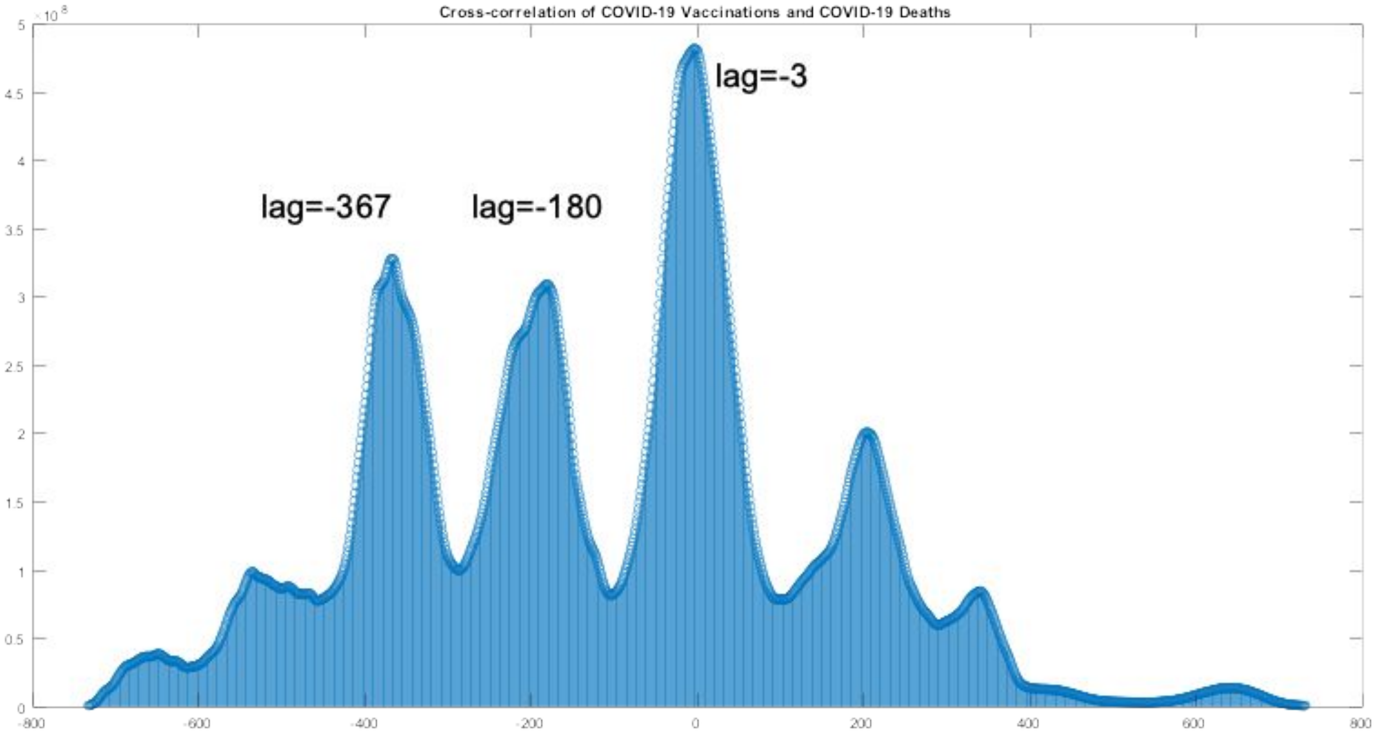
Overall signal correlations 2021-2022

- Non-surprising positive correlations:
  - Deaths – cases, Deaths – ICU patients
- Surprising positive correlation
  - Deaths – Vaccination
- Surprising null correlation
  - Stringency index – deaths and cases

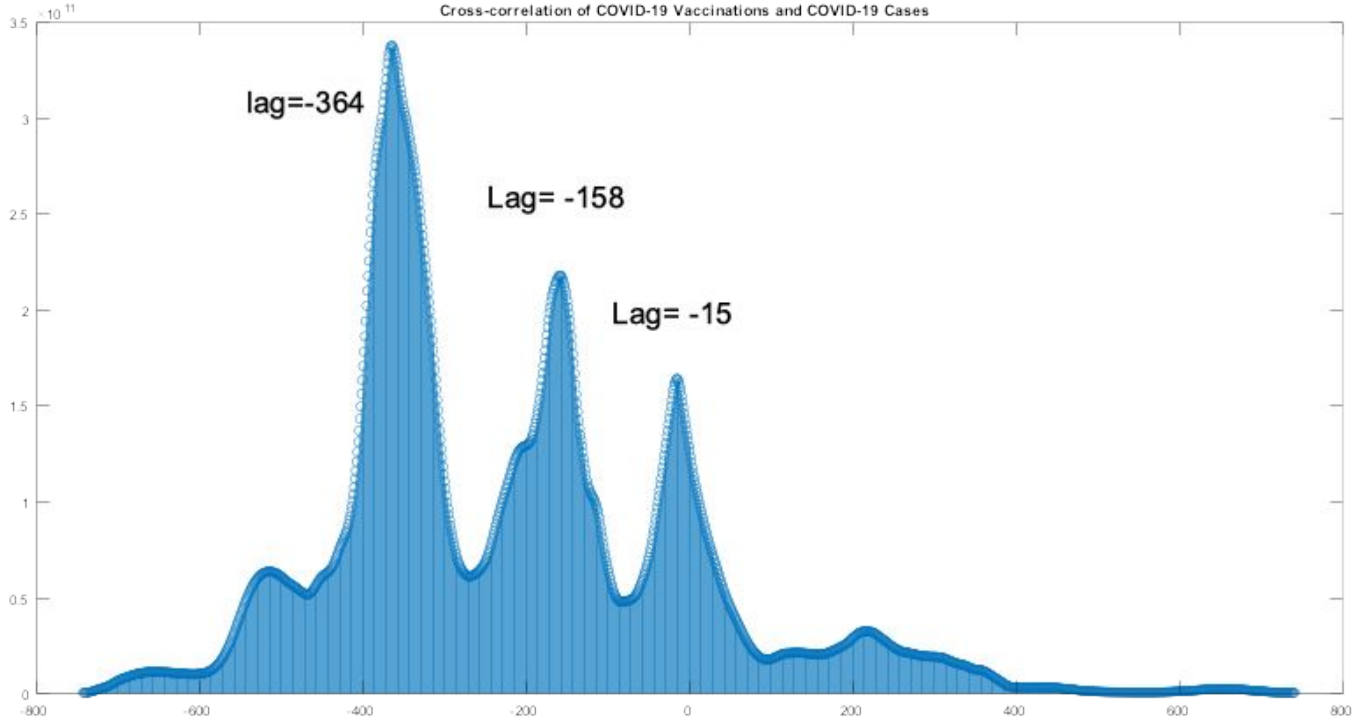
xcorr(new_cases_smoothed, new_deaths_smoothed)



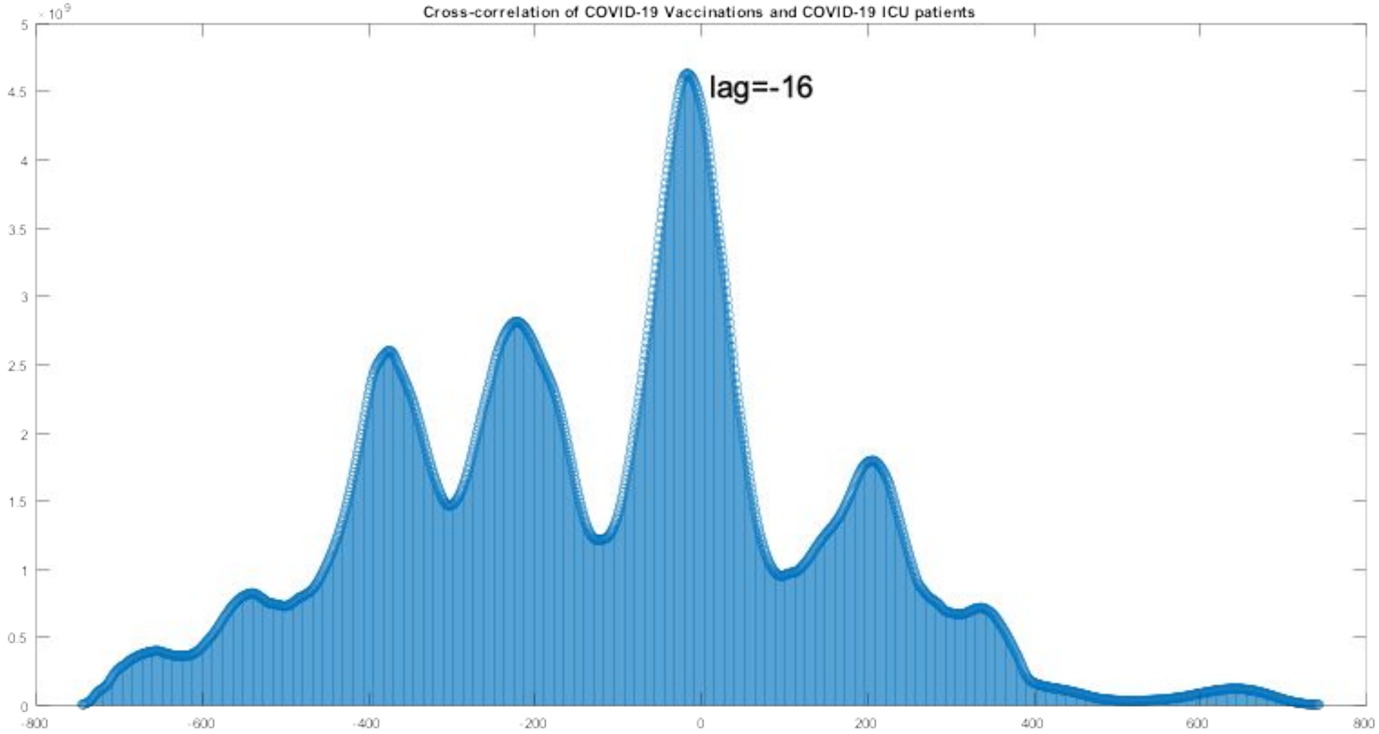
`xcorr(new_vaccinations_smoothed, new_deaths_smoothed)`



xcorr(new_vaccinations_smoothed , new_cases_smoothed)



xcorr(new_vaccinations_smoothed , icu_patients)



Results from cross-correlation

- Non surprising
  - Cases match/predict deaths after 11 days
- Surprising
  - Vaccination doses match/predict ICU after 16 days
  - Vaccination doses match/predict deaths after 3 days
  - Vaccination doses match/predict cases after 15, 158 days
    - Maximal at 365 days??

The origins of the virus

The origin of the virus

https://www.nature.com/articles/s41586-020-2008-3

Case Reports > Nature. 2020 Mar;579(7798):265-269. doi: 10.1038/s41586-020-2008-3. Epub 2020 Feb 3.

A new coronavirus associated with human respiratory disease in China

Fan Wu[#, 1], Su Zhao[#, 2], Bin Yu[#, 3], Yan-Mei Chen[#, 1], Wen Wang[#, 4], Zhi-Gang Song[#, 1], Yi Hu[#, 2], Zhao-Wu Tao[2], Jun-Hua Tian[3], Yuan-Yuan Pei[1], Ming-Li Yuan[2], Yu-Ling Zhang[1], Fa-Hui Dai[1], Yi Liu[1], Qi-Min Wang[1], Jiao-Jiao Zheng[1], Lin Xu[1], Edward C Holmes[1, 5], Yong-Zhen Zhang[6, 7, 8]

Affiliations + expand

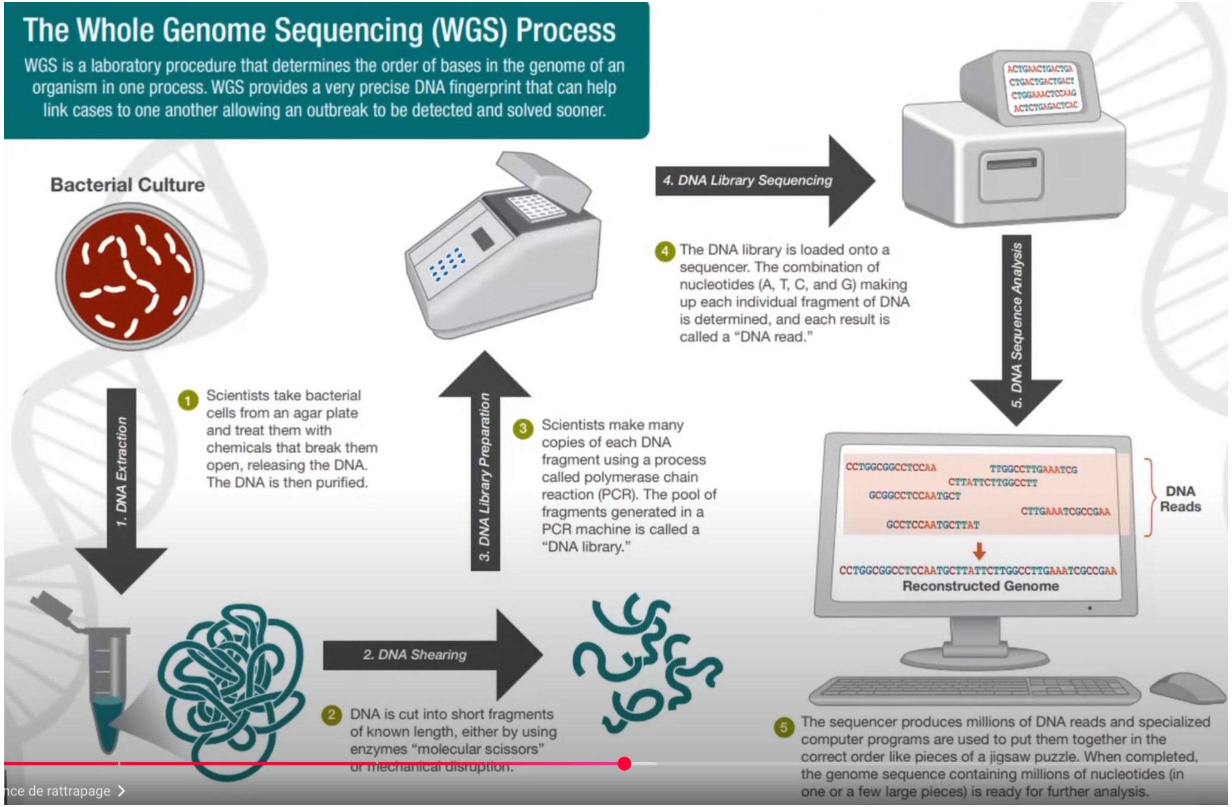
PMID: 32015508 PMCID: PMC7094943 DOI: 10.1038/s41586-020-2008-3

Erratum in

Author Correction: A new coronavirus associated with human respiratory disease in China.
Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ.
Nature. 2020 Apr;580(7803):E7. doi: 10.1038/s41586-020-2202-3.

sample size
n=1
analysis based on a sample of pulmonary tissue of a person with Pneumonia

The origin of the virus



The Whole Genome Sequencing (WGS) Process

WGS is a laboratory procedure that determines the order of bases in the genome of an organism in one process. WGS provides a very precise DNA fingerprint that can help link cases to one another allowing an outbreak to be detected and solved sooner.

① Scientists take bacterial cells from an agar plate and treat them with chemicals that break them open, releasing the DNA. The DNA is then purified.

② DNA is cut into short fragments of known length, either by using enzymes "molecular scissors" or mechanical disruption.

③ Scientists make many copies of each DNA fragment using a process called polymerase chain reaction (PCR). The pool of fragments generated in a PCR machine is called a "DNA library."

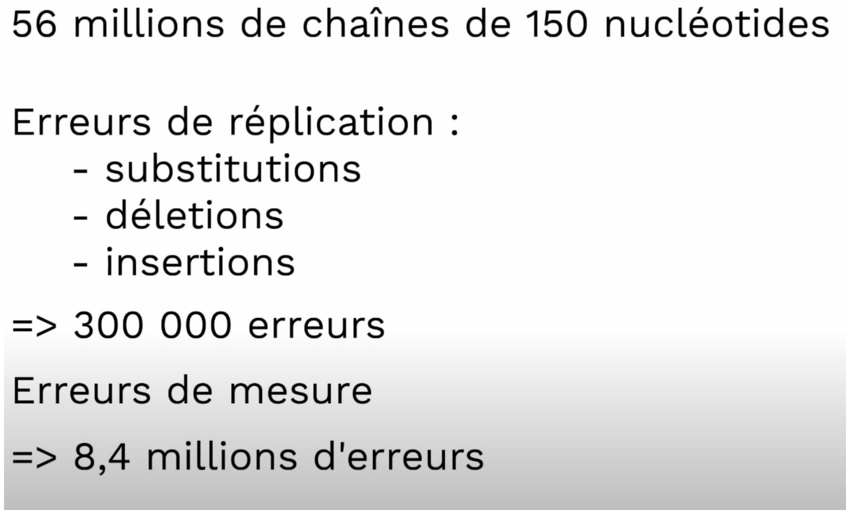
④ The DNA library is loaded onto a sequencer. The combination of nucleotides (A, T, C, and G) making up each individual fragment of DNA is determined, and each result is called a "DNA read."

⑤ The sequencer produces millions of DNA reads and specialized computer programs are used to put them together in the correct order like pieces of a jigsaw puzzle. When completed, the genome sequence containing millions of nucleotides (in one or a few large pieces) is ready for further analysis.

the sample biological material is the basis for a sequencing process, whose critical component is a software looking for the **most likely** sequence similar to a given pattern: i.e. SARS-COV-1

https://www.youtube.com/watch?v=EMMsdAanAJM

The origin of the virus



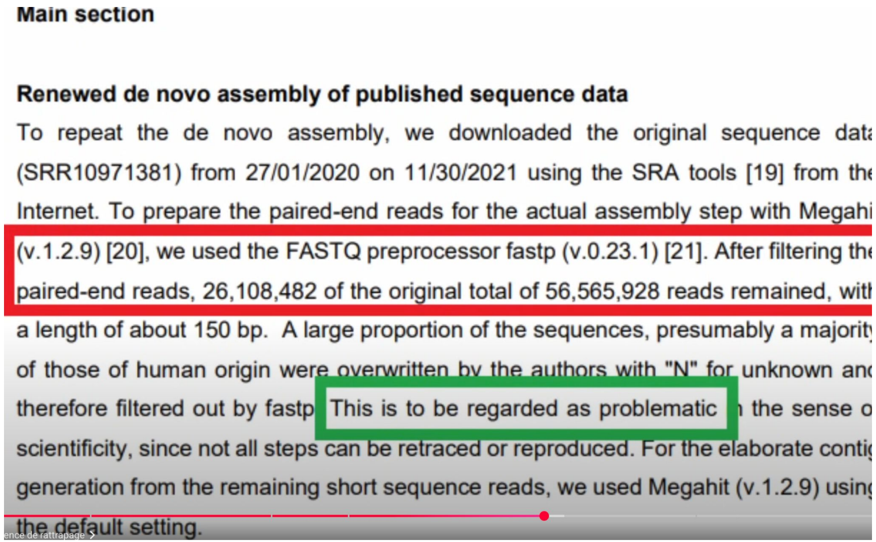
56 millions de chaînes de 150 nucléotides

Erreurs de réplication :
  - substitutions
  - délétions
  - insertions

=> 300 000 erreurs

Erreurs de mesure

=> 8,4 millions d'erreurs



Main section

Renewed de novo assembly of published sequence data

To repeat the de novo assembly, we downloaded the original sequence dat
(SRR10971381) from 27/01/2020 on 11/30/2021 using the SRA tools [19] from th
Internet. To prepare the paired-end reads for the actual assembly step with Megah
(v.1.2.9) [20], we used the FASTQ preprocessor fastp (v.0.23.1) [21]. After filtering th
paired-end reads, 26,108,482 of the original total of 56,565,928 reads remained, wit
a length of about 150 bp. A large proportion of the sequences, presumably a majorit
of those of human origin were overwritten by the authors with "N" for unknown an
therefore filtered out by fastp This is to be regarded as problematic the sense o
scientificity, since not all steps can be retraced or reproduced. For the elaborate conti
generation from the remaining short sequence reads, we used Megahit (v.1.2.9) usin
the default setting.

Some uncertainties and errors in the process

The origin of the virus

Two sequencing softwares were applied to find the “most likely” sequence similar to SARS-COV-1

Trinity	Megahit
1 million de séquences Max : 11 760 nucléotides	400 000 séquences Max : 30 474 nucléotides

the chosen one

The origin of the virus

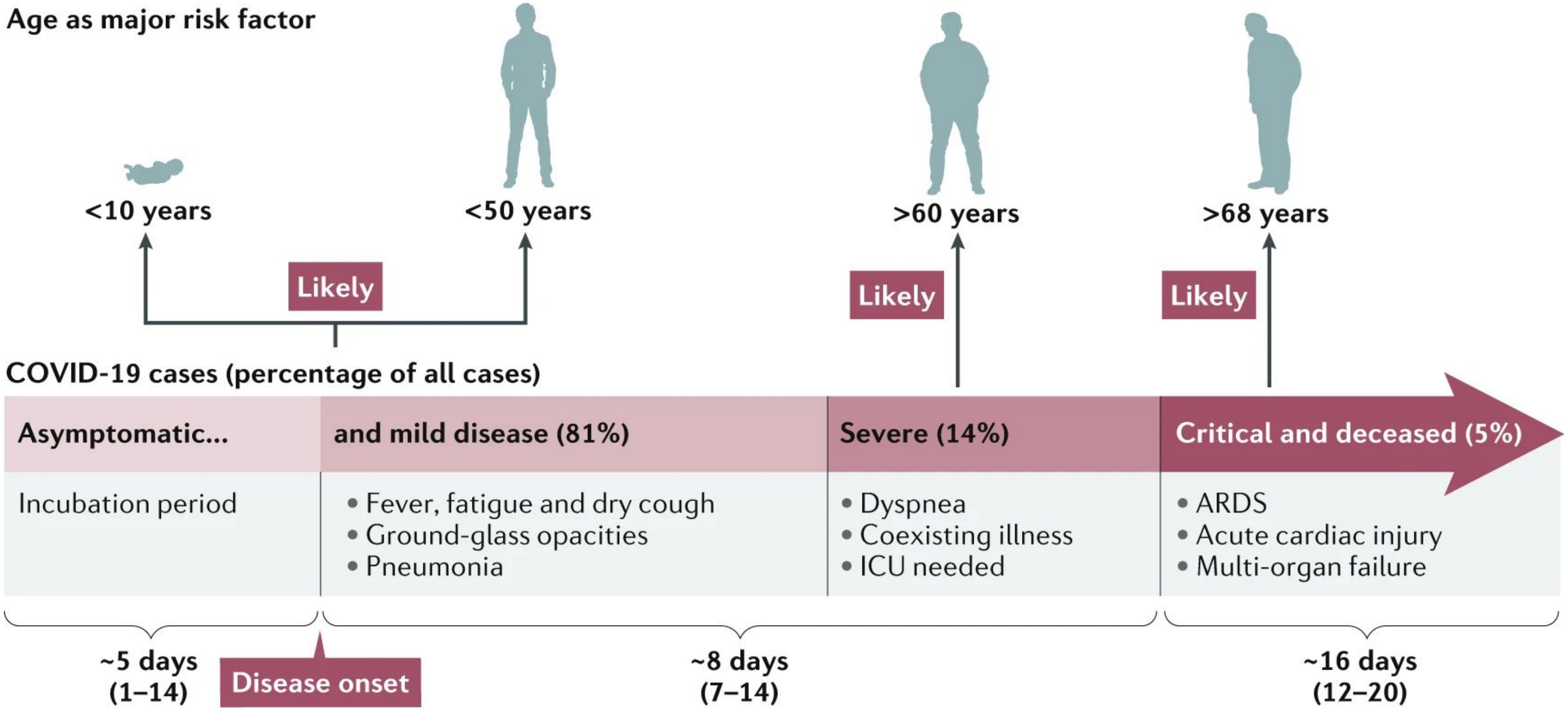
The sequence published in Nature has never been replicated by any researcher in the world

More than 6 M variants

It has been corrected in the GISAID (https://gisaid.org/) repository

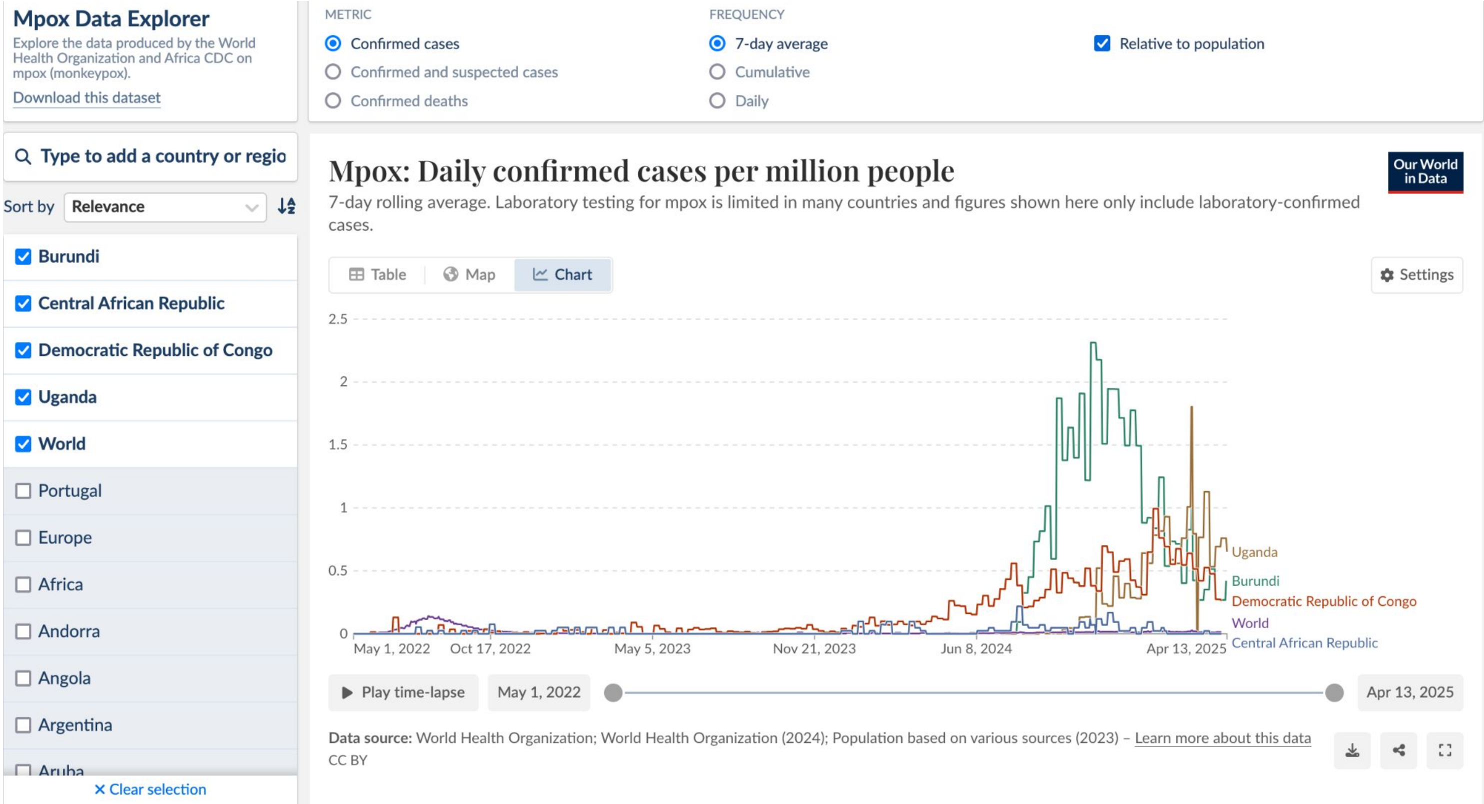
The origin of the virus

https://www.nature.com/articles/s41579-020-00459-7



The next pandemic

The next pandemic



Mpox Data Explorer

Explore the data produced by the World Health Organization and Africa CDC on mpox (monkeypox).

Download this dataset

METRIC

● Confirmed cases
○ Confirmed and suspected cases
○ Confirmed deaths

FREQUENCY

● 7-day average
○ Cumulative
○ Daily

☑ Relative to population

Type to add a country or regio

Sort by Relevance

☑ Burundi
☑ Central African Republic
☑ Democratic Republic of Congo
☑ Uganda
☑ World
☐ Portugal
☐ Europe
☐ Africa
☐ Andorra
☐ Angola
☐ Argentina
☐ Aruba

✕ Clear selection

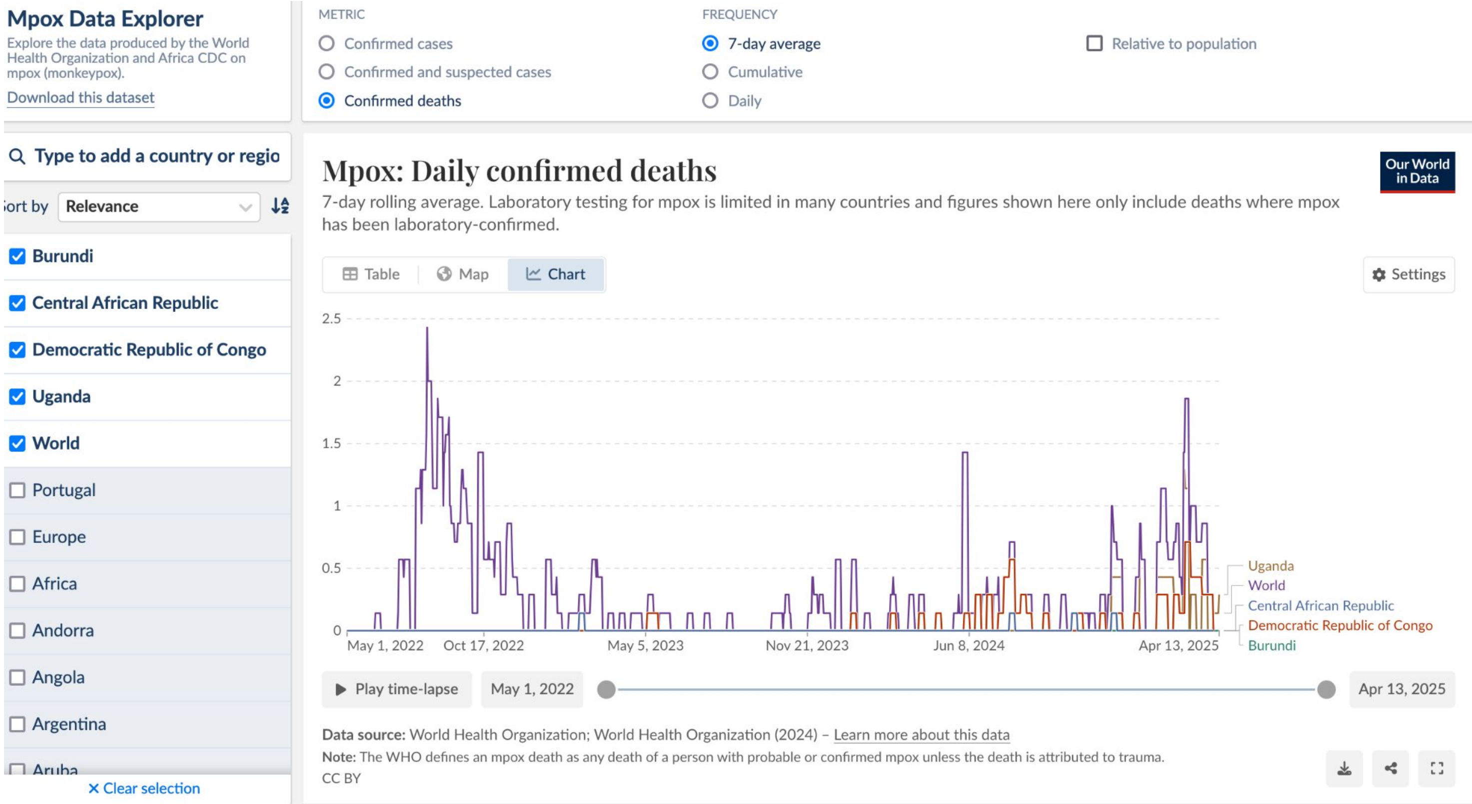
Mpox: Daily confirmed cases per million people

7-day rolling average. Laboratory testing for mpox is limited in many countries and figures shown here only include laboratory-confirmed cases.

Data source: World Health Organization; World Health Organization (2024); Population based on various sources (2023) – Learn more about this data

CC BY

The next pandemic



Mpox Data Explorer

Explore the data produced by the World Health Organization and Africa CDC on mpox (monkeypox).

Download this dataset

METRIC

○ Confirmed cases
○ Confirmed and suspected cases
● Confirmed deaths

FREQUENCY

● 7-day average
○ Cumulative
○ Daily

☐ Relative to population

Type to add a country or regio

Sort by Relevance

☑ Burundi
☑ Central African Republic
☑ Democratic Republic of Congo
☑ Uganda
☑ World
☐ Portugal
☐ Europe
☐ Africa
☐ Andorra
☐ Angola
☐ Argentina
☐ Aruba

Mpox: Daily confirmed deaths

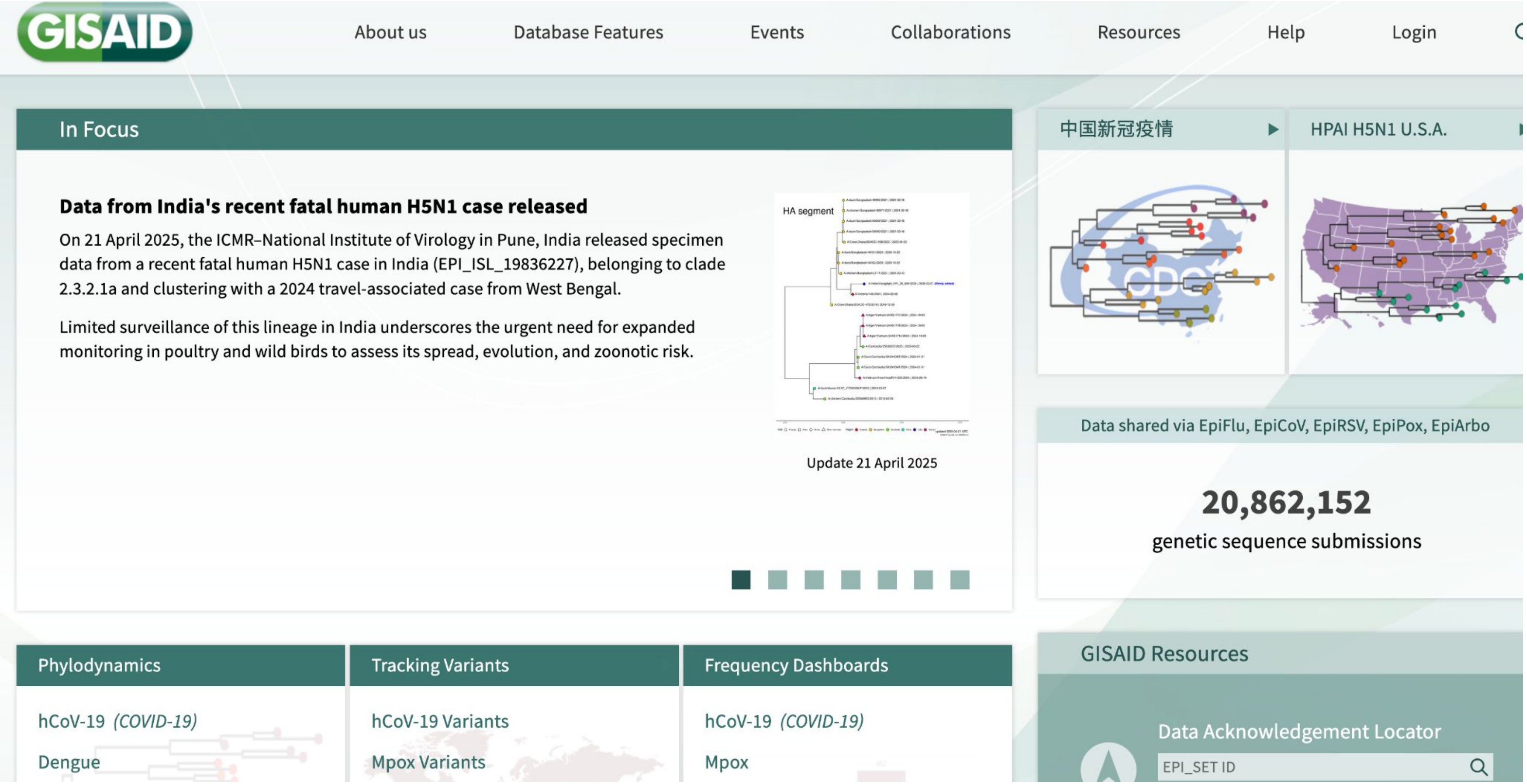
7-day rolling average. Laboratory testing for mpox is limited in many countries and figures shown here only include deaths where mpox has been laboratory-confirmed.

Data source: World Health Organization; World Health Organization (2024) – Learn more about this data

Note: The WHO defines an mpox death as any death of a person with probable or confirmed mpox unless the death is attributed to trauma.

CC BY

The next pandemic



Data from India's recent fatal human H5N1 case released

On 21 April 2025, the ICMR–National Institute of Virology in Pune, India released specimen data from a recent fatal human H5N1 case in India (EPI_ISL_19836227), belonging to clade 2.3.2.1a and clustering with a 2024 travel-associated case from West Bengal.

Limited surveillance of this lineage in India underscores the urgent need for expanded monitoring in poultry and wild birds to assess its spread, evolution, and zoonotic risk.

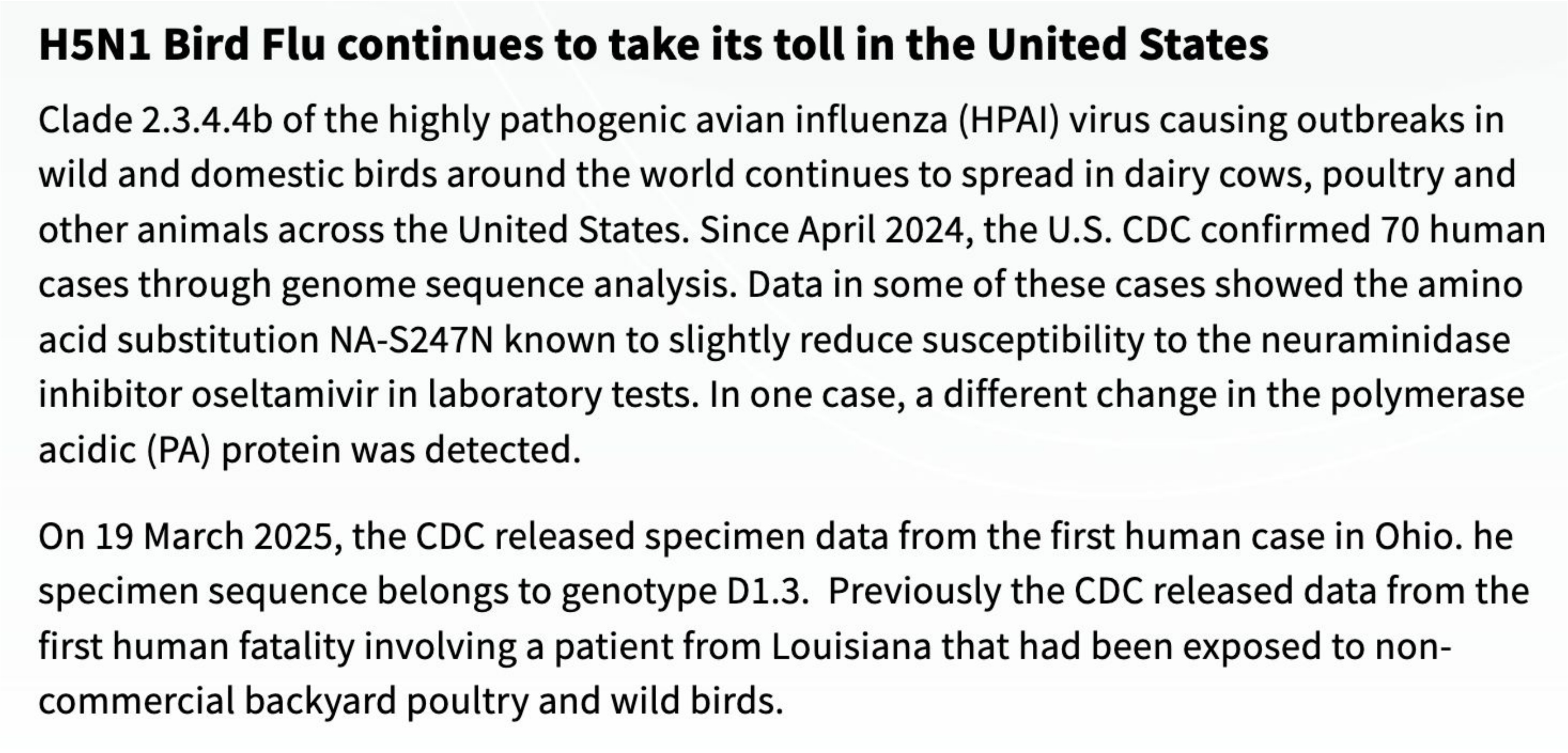
Update 21 April 2025

Data shared via EpiFlu, EpiCoV, EpiRSV, EpiPox, EpiArbo

20,862,152

genetic sequence submissions

The next pandemic



H5N1 Bird Flu continues to take its toll in the United States

Clade 2.3.4.4b of the highly pathogenic avian influenza (HPAI) virus causing outbreaks in wild and domestic birds around the world continues to spread in dairy cows, poultry and other animals across the United States. Since April 2024, the U.S. CDC confirmed 70 human cases through genome sequence analysis. Data in some of these cases showed the amino acid substitution NA-S247N known to slightly reduce susceptibility to the neuraminidase inhibitor oseltamivir in laboratory tests. In one case, a different change in the polymerase acidic (PA) protein was detected.

On 19 March 2025, the CDC released specimen data from the first human case in Ohio. he specimen sequence belongs to genotype D1.3. Previously the CDC released data from the first human fatality involving a patient from Louisiana that had been exposed to non-commercial backyard poultry and wild birds.