Research Article

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Abstract

Introduction: About a year and a half after the declaration of the COVID-19 pandemic, almost the entire planet has been affected by SARS-CoV-2 coronavirus and its variants, with serious public health consequences and other repercussions not yet thoroughly evaluated or foreseen in terms of economic, financial and social disruption throughout communities. Therefore, it is of utmost importance to understand the geography of the evolution of successive pandemic waves. Particularly in European countries, where, in recent decades, more advanced models for cohesion and competitiveness of a whole with more than 400 million inhabitants have been achieved, with ambitious challenges for horizon 2030 regarding this vast territory's economic, social, and environmental sustainability.

Objective: The main objective of this research is to describe the multivariate trajectories of COVID-19 incidence, mortality, hospital admissions, ICU admissions and testing, over three successive waves, covering all European Union (EU) countries with more than two million inhabitants, over 14-days periods before May 4 2020, until February 22 2021.

Methods: This research includes 22 European countries representing about 98.8% of the EU population, described by six epidemiological variables over 43 time periods from the ECDC database: the 14-day notification rate of new cases reported for 100,000 inhabitants; the 14-day notification rate of reported deaths per one million inhabitants; the mean and the rate for 100,000 population of hospital occupancy and ICU occupancy; the testing rate per 100,000 population; and the 14-days percentage of test positivity.

An exploratory data analysis of each epidemiological variable identified a typology of countries profiles evolution.

Multivariate exploratory statistical methods, namely a 3-way data analysis (double principal components and rank principal components analyses), were applied with software R version 4.1.0.

Results: The multivariate evolution profile of the COVID-19 pandemic in the EU over the studied period highlighted 3 phases: the first phase over 24 time periods, with a relatively low COVID-19 incidence, hitting only part of EU countries; a second phase at the beginning of the second wave, when COVID-19 spread to most countries, with a higher impact on national

José M. Tallon,[1,2] Paulo Gomes,[3] Leonor Bacelar-Nicolau,[3,4] Sérgio Bacelar[5]

[1]Sports Sciences Department, Exercise and Health, Universidade de Trás–os–Montes e Alto Douro, Portugal (UTAD)
[2]Medical Clinic Doctor Tallon
[3]NOVA IMS Information Management School, Portugal
[4]Faculdade de Medicina, Institute of Preventive Medicine and Public Health & ISAMB–Institute of Environmental Health, Universidade de Lisboa, Portugal
[5]Statistics Portugal

Correspondence: Jose Maria Tallon, Sports Sciences Department, Exercise and Health, Universidade de Trás–os–Montes e Alto Douro, Portugal (UTAD), Email josemariatallon@clinicadrtallon.com

health systems; lastly, a third phase coincident with the peak of the second wave and the onset of the third wave, a particularly reactive phase from the public authorities, with intensified testing of the population. These results are clear from the principal component analysis of the centres of gravity of the 43 time periods (interstructure). The multivariate statistical analysis of the global dataset of all countries over the 43 time periods additionally provides the main factorial representation of the trajectories of COVID-19 for each country in direct comparison with the global average ranked values reached by the six epidemiological variables over the whole period under study (intrastructure).

These trajectories make it possible to identify different country profiles throughout the successive pandemic waves and counter-cyclical behaviours, partly explained by the insufficient harmonisation of public policies to tackle the pandemic within the EU.

Keywords: COVID-19, epidemiological variables, three-way data analysis, principal components analysis, rank principal components, missing values

---

Abbreviations: European Centre for Disease Prevention and Control (ECDC), European Medicines Agency (EMA), European Union (EU), Intensive care unit (ICU), Principal Components Analysis (PCA), Ranks Principal Components Analysis (RPCA), World Health Organization (WHO)

Introduction

The acute respiratory syndrome triggered by the type 2 coronavirus SARS-CoV-2 was initially identified in Wuhan, China, spreading quickly throughout the rest of the world, leading to the World Health Organization (WHO) declaration of the COVID-19 pandemic on March 11 2020. This virus shows a high transmissibility rate, which explains the still quick and steep increase rate of infected people. The COVID-19 has two main predecessors: the 2002 Severe

Acute Respiratory Syndrome-SARS-Cov, and the 2012 Middle East Respiratory Syndrome-MERS-Cov.[1,2]

The average incubation time for SARS-CoV-2 is 4 to 6 days, and about 95% of cases are symptomatic within 14 days of infection.[3,4] The importance of this finding is the fact that the patient may transmit the virus in the asymptomatic phase of the infection.[5,6]

When in previous papers,[7,8] profiles of COVID-19 incidence were analysed in OECD countries from the beginning of the pandemic until the end of the period of confinement or the application of various restrictive measures with the consequent flattening of the epidemiological curve, the great unknown was if after these periods the pandemic would be relatively controlled or if new waves were to be expected. Today, the answer is very clear, having experienced a

---

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **99**

second and a third waves, driven by variants with greater transmission capacity. It is thus important to revisit in detail the evolution of countries' behaviours throughout successive moments in time, namely 43 periods, from May 4, 2020, until February 22, 2021, using indicators evaluated over fourteen days prior to each date. This study focuses on 22 countries of the European Union that currently account for about 43 million cases, representing more than 23% of all COVID-19 cases globally worldwide.

Methods

Data

Publicly available data from the European Centre for Disease Prevention and Control (ECDC) was used concerning 22 European Union (EU) countries from May 4, 2020, to February 22, 2021, regarding 43 time periods for epidemiologic variables and testing.[9]

The 22 countries under study were all EU countries with more than two million inhabitants: Austria, Belgium, Bulgaria, Croatia, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden.

The variables under study were evaluated each last fourteen days period, from the period fourteen days before May 4, 2020, to the period fourteen days before February 22, 2021: the number of new cases over the previous 14 days (incidence), the number of COVID-19 deaths over the previous 14 days (mortality), the number of hospital admissions over the previous 14 days, the number of Intensive care unit (ICU) admissions over the previous 14 days, the number of COVID-19 tests over the previous 14 days (testing), and the number of positive COVID-19 tests over the previous 14 days (positivity).

The respective indicators were calculated per 100,000 inhabitants for the accumulated number of new cases, total of hospitalisations, total of ICU and total of tests along each period. Additionally, the total number of deaths was calculated per one million inhabitants and the positivity of tests as a percentage number.

Missing data and imputation

We have analysed six variables during 43 periods for 22 countries. From the EU 26 countries, we excluded those with less than two million inhabitants (Cyprus, Latvia, Luxembourg, and Malta) since countries with a small dimension have a greater chance of showing abnormal behaviour.

Missing data represents 9.2% of all data values (5676). There is a concentration of the missing data in two of the six indicators (hospital or ICU occupation), representing 86.5% of all the missing data. For these two variables, there is also a concentration of missing values in some countries: from the ECDC source, there is no ICU data for Croatia, Greece, Hungary, Poland and Slovakia, and there is no hospital occupation data for Germany, Greece, and Romania. Also, Lithuania has a high number of missing values for these two variables. In these countries, such missing values represent almost 80% of all missing values.

This data can be considered missing at random since there is no reason for them to be related to other variables or exogenous factors. The missingness situation seems to be related to the data collection process.

For imputation of missing values, we considered three groups of countries: first, those with no data on ICU occupation, second, those without data on Hospital occupation and third, Greece that hasn't either but has data on hospital admissions.

For the first two groups of cases and each country, we used the data of three countries with complete data and similar characteristics (population, GDP, geographical contiguity). We computed the ratio between hospital and ICU occupation for each period. We obtained two sets of ratios: one for the first group of countries and the other for the second and trimmed the outliers in each set. Using the resulting set, we bootstrapped it repeating the procedure one hundred times. The global mean (a ratio between hospital and ICU occupation) was used to estimate the value of hospital occupation for the first group of countries and ICU occupation for the second group proportionally.

In the particular case of Greece, we considered the data from all countries which had values for hospital occupation and hospital admission, and we opted for a similar method to estimate the ratio between occupation and admissions. This ratio was used afterwards on Greece admission values to estimate both hospital and ICU occupation.

After using this process, we reduced missing values to 3.3%. Afterwards, we used a missing value imputation by linear interpolation after visually inspecting each series with missing data using the "imputeTS" R package (Moritz S, Bartz-Beielstein T (2017). "imputeTS: Time Series Missing Value Imputation in R." *The R Journal*, 9(1), 207–218. doi: 10.32614/RJ-2017-009).

Statistical analysis

A preliminary exploratory statistical analysis based on univariate, bivariate and hierarchical cluster analysis[10] was applied to study the marginal empirical distributions of variables, to be used afterwards on the multivariate methodologies, and observe how similar countries may be grouped for each variables' time series.

An exploratory multivariate statistical analysis was then applied, based on a three-way component analysis (double principal component),[11] to obtain a global comparison of the evolution of associations between variables and evolution of countries COVID-19 incidence over the time periods under study, taking as reference the global behaviour of the set of variables under study.

The first step of this multivariate method (interstructure) consists of a statistical analysis of the global evolution of the pandemic over time. Standardised principal components are considered, where "objects" are the centres of gravity of countries' clusters associated with each table $X^{(t)}$, where n is the number of countries, p is the number of variables previously selected and t the time period $(t = 1, 2, \ldots, 43)$. A Euclidean image of 43 tables on a lower dimension space is thus obtained. Generally, the first principal factorial plane explains quite well the evolution of such centres of gravity over time, describing the global evolution of countries' COVID-19 incidence from May 2020 until Mars 2021.

The second step of this multivariate three-way data analysis provides a common space of joint representation of the 43 time periods (intrastructure), based on an optimised criterion (compromise), which maximises the global projected inertia. This makes it possible to characterise the projection space to represent the pandemic trajectory graphically for each country in relation to the global mean rank behaviour in relation to the centre of gravity of the joint table.

$$X_{43n,6} = \begin{bmatrix} X^{(1)} \\ \hline X^{(2)} \\ \hline \vdots \\ \hline X^{(43)} \end{bmatrix} \begin{matrix} \rightarrow \\ \vdots \\ \vdots \\ \rightarrow \end{matrix} \begin{matrix} 14\,days\;until\;May, 4, 2020 \\ \vdots \\ \vdots \\ 14\,days\;until\;February, 22, 2020 \end{matrix}$$

Where $X_{i,j}^{(t)}$ represents the rank of country $i(i=1,\ldots,22)$ on variable $j(j=1,2,\ldots,6)$ in period $t(t=1,\ldots,43)$ and n represents the number of countries included in our study.

On this intrastructure step, a transformation was applied to variables, and PCA was used to the ranked joint table of the 43 periods juxtaposed dataset. This approach is very insensitive to the presence of outliers and analysing a set of ranks is more suitable than examining heterogeneous sets of measurements, which would bias PCA results by its effect on means, variances, covariances and correlations. The ranked PCA thus generates rank trajectories for each country, which is quite adequate given this research goal.

Defining $V^{(t)}$ the variance and covariance matrix associated with the data set represented by table $X^{(t)}$, $\varphi_k^{(t)}$ the K-factor of rank principal component of table $X^{(t)}$ and $\lambda_k^{(t)}$ the inertia explained by K-factor on period t it can be shown that ${}^t\varphi_k^{(t)}V^{(t)}\varphi_k^{(t)} = \lambda_k^{(t)}$.

Thus, for a system of $\text{k}-axes\ (K=1,\ldots,q)$ we define an index value.

$$\Phi(\text{t},\varphi) = \frac{\sum_{k=1}^{q}\lambda_k^{(t)} - \sum_{k=1}^{q}{}^t\varphi_k V^{(t)}\varphi_k}{\sum_{k=1}^{q}\lambda_k^{(t)}}$$

Which measure the relative loss of inertia of cluster $N^{(t)}$ of countries associated to $X^{(t)}$ when such countries for all periods are projected on a common subspace generated by vectors $\varphi_1, \varphi_2, \ldots, \varphi_q$ and not projected on the concrete subspace associated to the principal factors of each table $X^{(t)}$.

So, the global criteria to select an optimum system of axes will minimise the sum of the loss of relative inertia of each cluster $N^{(t)}(t=1,\ldots,43)$:

$$\frac{1}{43}\sum_{t=1}^{43}\Phi(t,\varphi)$$

The statistical approach (intrastructure) will provide a representation of COVID-19 incidence of each European country along the predefined global period. These representations on the first factorial plane are relative rank trajectories of countries pandemic.

The software used for statistical analysis was R version 4.1.0.[12]

Results

Preliminary exploratory study

In this first exploratory approach, country trajectories are analysed for each indicator and similar countries are grouped according to their profile regarding each variable separately.

As stated before, the period under analysis spans between the 19th week of 2020 and the 8th week of 2021, more precisely from May 4, 2020, to February 22, 2021. Each time point corresponds to one week, and thus 43 weeks have been analysed.

For each of the 22 European Union countries under study (Table 1), six variables were considered from ECDC databases: cases rate (per 100 000 inhabitants), deaths rate (per one million inhabitants), Hospital and Intensive Care Unit (ICU) current occupancy (per 100 000 inhabitants), testing rate (per 100 000 inhabitants) and positivity rate (percentage of the number of new confirmed cases from the number of total tests undertaken per week).

Table 1 Countries by variable and group

Country	Cases	Deaths	Hosp	ICU	Tests	Positivity
Austria	1	2	5	4	1	1
Belgium	1	2	5	4	2	1
Bulgaria	3	3	1	4	2	2
Croatia	2	3	1	4	2	2
Czechia	4	3	1	2	2	2
Denmark	3	1	4	6	1	3
Finland	3	1	4	6	2	3
France	1	2	3	3	2	1
Germany	3	1	5	3	2	1
Greece	3	2	5	4	2	3
Hungary	1	3	1	1	2	2
Ireland	3	1	4	5	2	3
Italy	1	3	1	4	2	1
Lithuania	2	4	1	1	2	2
Netherlands	5	1	4	5	2	1
Poland	1	3	1	1	2	2
Portugal	5	4	2	3	2	1
Romania	3	2	3	3	2	2

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **101**

Table Continued...

Country	Cases	Deaths	Hosp	ICU	Tests	Positivity
Slovakia	4	4	2	2	2	4
Slovenia	2	3	1	1	2	2
Spain	5	2	3	3	2	1
Sweden	2	1	4	5	2	1

For each indicator, a rolling mean of two weeks (14 days) was calculated, but ICU occupancy was lagged one week to depict the deferred effect of hospitalisation on ICU occupation.

The rate of cases has priority in the causal chain. The other variables depend partially on their value. Therefore, the study of their evolution in the period, considering all countries, is significant. Figure 1 shows the median, the first and third quartile for each week. The median describes the typical evolution of the new cases rate, and the distance between the two quartiles illustrates the dispersion of the variable between countries.
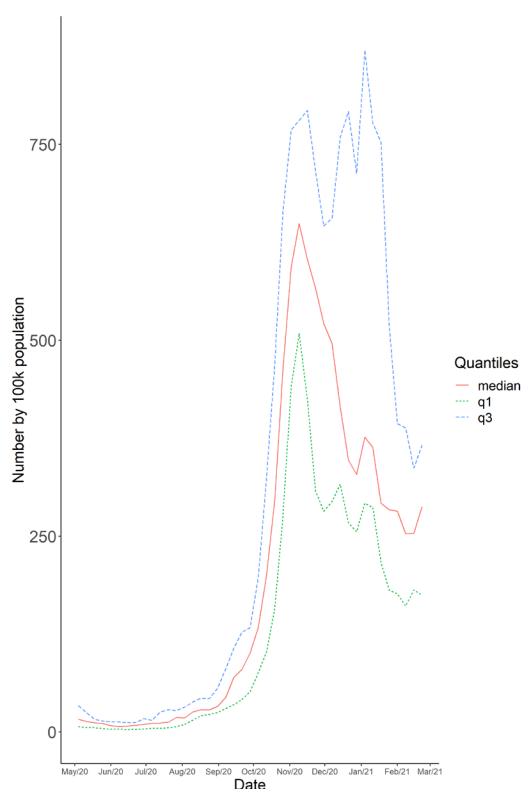
Figure 1 Cases rate 14-day: Evolution of the median and the 10th and 90th percentile (all countries).

This figure allows a rough classification of different sub periods according to the behaviour of the indicator: until September 2020, the median remains almost constant, reflecting the relative stabilisation of the number of new infections, but the distance between the third quartile and the median is greater than the distance between the first quartile and the median. This asymmetry results from the fact that some countries have a disproportionate number of higher case rates. The interquartile range also shows that the period beginning in January 2021 reveals a sharp increase in the dispersion of new cases rate between countries. Figure 2 presents the evolution of the quartile-based variance coefficient (QVC = Interquartile range/median) and indicates a high level of dispersion during August 2020.
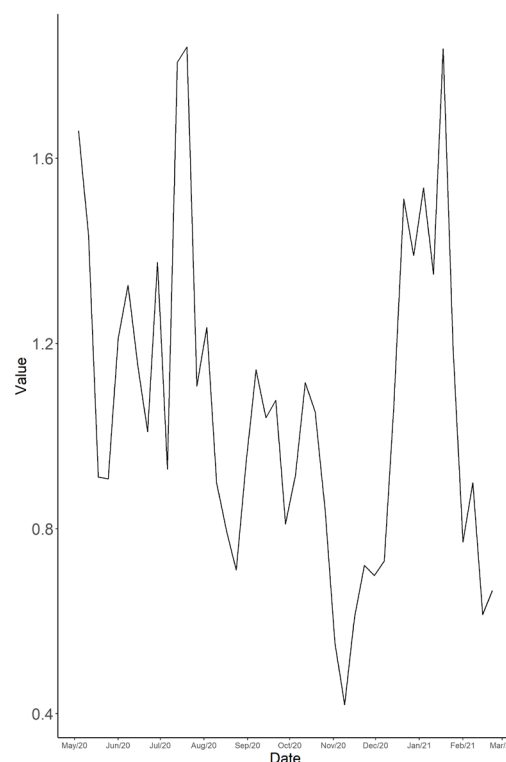

Figure 2 Cases rate 14-day: Evolution of the quartile based variance coefficient (all countries).

The second subperiod spans from September to November 2020, when the median reaches its absolute maximum. From this month onwards, the median of the indicator decreases, with a slight recovery during January 2021. But this last behaviour is not typical for countries with the worst performance on new cases rates. These, represented by the third quartile, reach two high peaks in December 2020 and January 2021 (the last one being the highest) before dropping steeply. Figure 3, representing the median first derivative (more precisely the derivative of a spline of the time series), also illustrates the evolution of new cases rates. When the new cases increase (derivate positive) or decrease (derivate negative), they change at a rate that can be detected visually by the line's steepness representing the derivative. Its value becomes positive even before August 2020 and increases very sharply until late October 2020. Afterwards, it decreases rapidly with fluctuations until late December 2020. After the beginning of 2021, it also increases with instabilities.

The range between quartiles shows that countries had different performances in this indicator, as well as also in the other indicators. But despite those differences, there were also similarities between countries along the period under study. These similarities may result from two aspects: a group of countries presenting similar values or a group of countries with either lower or higher values.
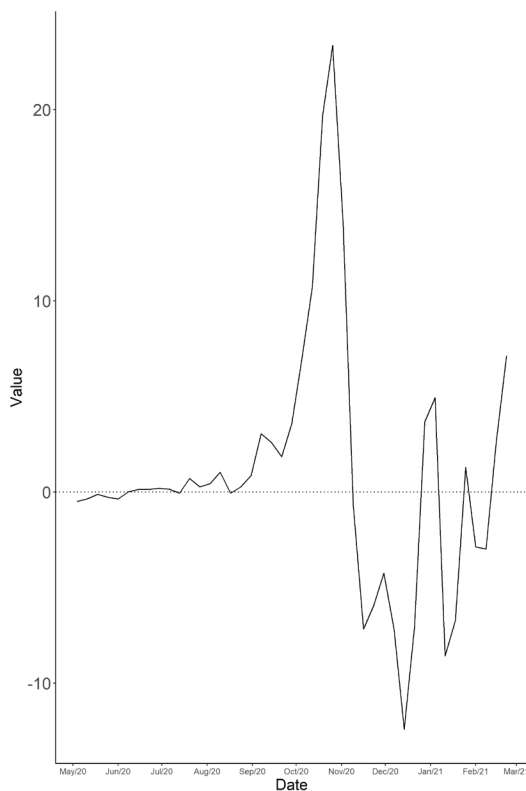
A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **102**

- Group 5 (Spain, Netherlands, Portugal) has two peaks, the first, lower, between the end of October and early November 2020 and the second, higher, between late December and mid-January 2021.


Figure 3 Cases rate 14-day: Evolution of the first derivative of the median (all countries).

To group the countries according to those similarities for each indicator, we calculated the difference between each time series using the Dynamic Time Warping method.[13,14] Distances computed with this method allow for some realignment of series, adjusting time for peaks and lows. Time series are thus classified in the same group if their trajectories present a similar type of profile. A hierarchical cluster analysis was applied over these dissimilarities based on these distances, and a dendrogram was plotted to assess the cluster accuracy visually. Table 1 identifies the cluster partition for each indicator and the respective cluster identification for each country. After clustering countries for each of the six indicators, groups or country clusters were represented using time-series graphs. Countries were then compared graphically for each indicator, ranking their respective boxplots by medians (Figures 5, 7, 9, 11, 13, 15).

For the new cases indicator, five groups of countries were identified (Figure 4):

- Group 1 (Austria, Belgium, France, Hungary, Italy, Poland) shows two peaks: the first is higher, between October and November 2020 and the second considerably smaller, between January and February 2021. Belgium stands out for its highest value at the end of October 2020.

- Group 2 (Croatia, Lithuania, Slovenia, Sweden) has only one peak in mid-December 2020.

- Group 3 (Bulgaria, Denmark, Finland, Germany, Greece, Ireland, Romania) shows only one peak but is characterised by low values (see Figure 5). Ireland peaks in January 2021 due to outliers.

- Group 4 (Czechia and Slovakia) shows two peaks, in late October 2020 and early January 2021.
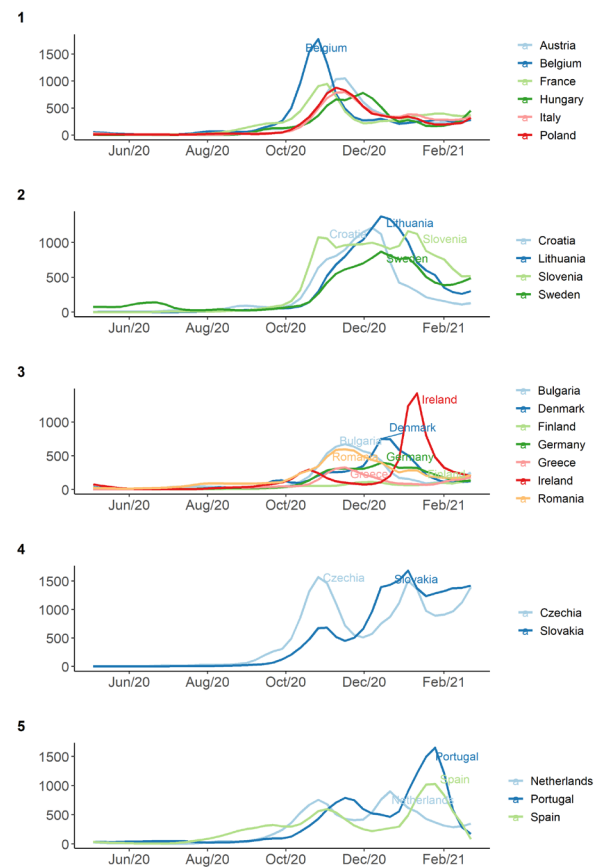
Figure 4 Cases rate (No. by 100k population) by group and country.
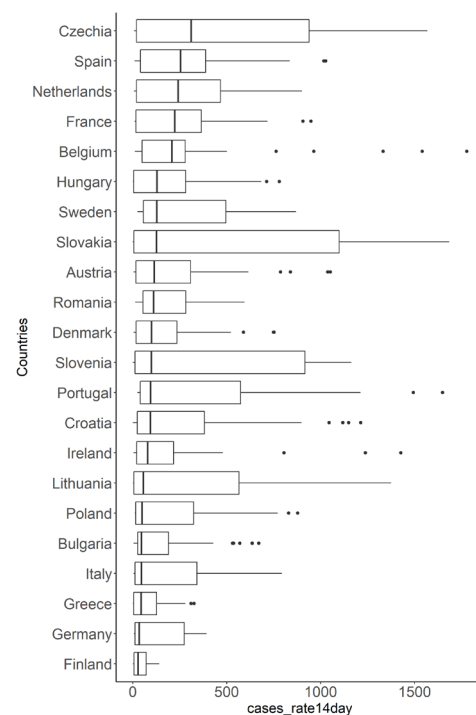
Figure 5 Boxplot of cases rate by country.

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **103**

Most of the distributions of cases rates by country are asymmetric positive because half of the data points have relatively low and similar values for each country. This asymmetry results from the difference between two periods: the first with low and the second with high values.

For the deaths' indicator, four groups were retained (Figure 6).

- Group 1 (Denmark, Finland, Germany, Ireland, Netherlands, Sweden) shows a U-shaped distribution with minima in early September 2020 and a peak in January 2021. This U form may result from the fact that those countries were getting out of the first wave in May 2020. Finland stands out because it has the lowest values for this indicator. Figure 7 shows that Finland has the lowest median and the lowest range.

- Group 2 (Austria, Belgium, France, Greece, Romania, Spain) has an evolution similar to Group 1. Still, the observed minima are observed one month before (August), and the peaks are located generally in November 2020.

- Group 3 (Bulgaria, Croatia, Czechia, Hungary, Italy, Poland, Slovenia) peaked in late November 2020, and after that, death rates stay at a relatively high value. Slovenia stands out with a peak in early December 2020.

- Group 4 (Lithuania, Portugal, and Slovakia) peaked in January 2021. After this peak, observed values drop sharply, except for Slovakia.
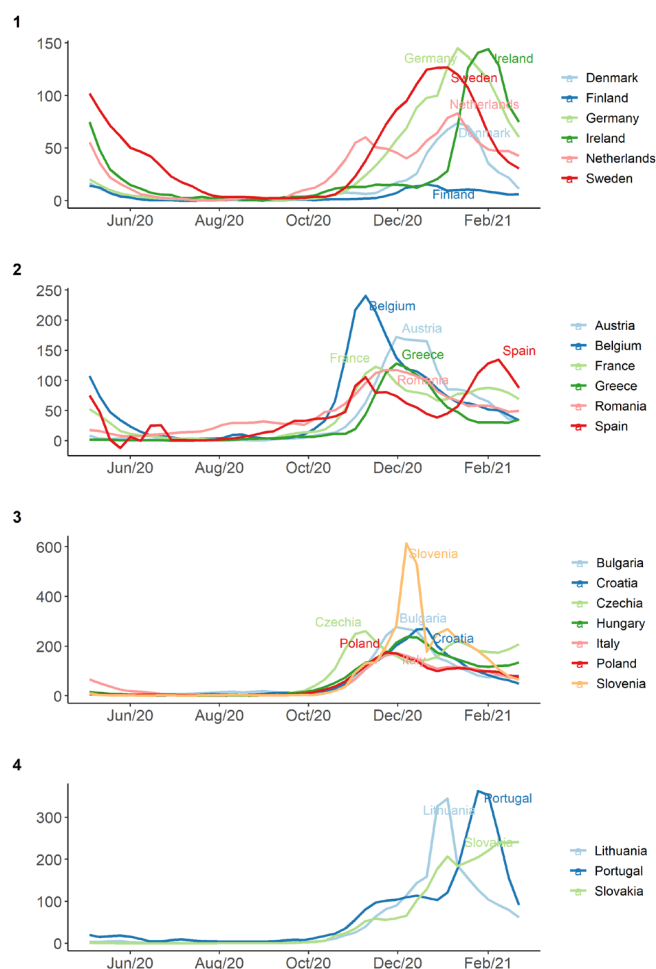

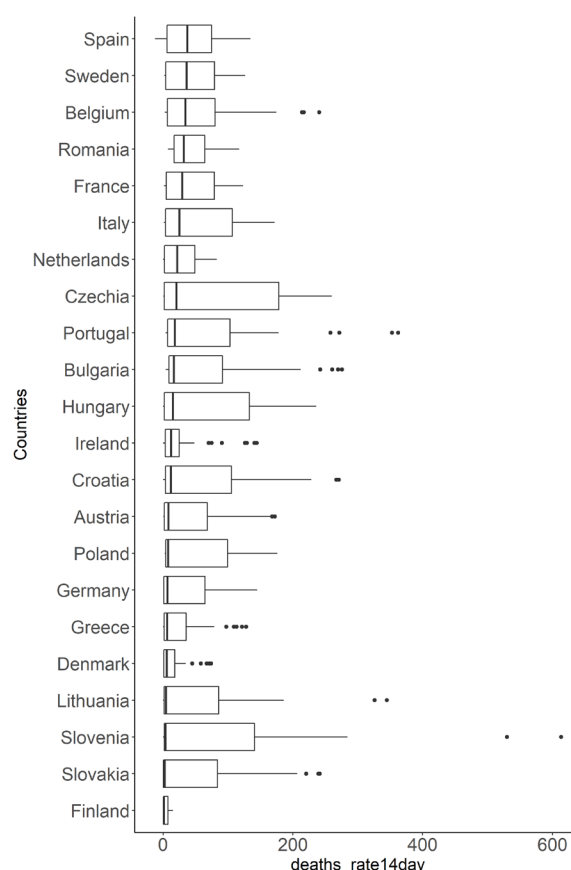**Figure 6** Deaths rate (No. by 1000k population) by group and country.

Figure 7 Boxplot of deaths rate by country.

For the hospital occupation indicator, five groups were retained (Figure 8):

- Group 1 (Bulgaria, Croatia, Czechia, Hungary, Italy, Lithuania, Poland, Slovenia) without a clear generalised trend but peaking in December 2020.

- Group 2 (Portugal, Slovakia) with a very accentuated growing trend. Portugal decreases sharply after the peak in early February 2021.

- Group 3 (France, Romania, Spain) shows a nearly sinusoidal behaviour for France and Spain, with two peaks in November 2020 and February 2021.

- Group 4 (Denmark, Finland, Ireland, Netherlands, Sweden) displays relatively low values, a U-shaped distribution mainly due to the impact of the first wave in Sweden and a peak in early January 2021 (Figure 9).

- Group 5 (Austria, Belgium, Germany, France) also presents a U-shape evolution mainly due to Belgium and peaks during November 2020, except for Germany that peaks during December 2020.

For the Intensive Care Unit occupation indicator, a partition of five groups was selected (Figure 10):

- Group 1 (Hungary, Lithuania, Poland, Slovenia) displays only one peak between early November 2020 and early January 2021.

- Group 2 (Czechia, Slovakia) presents a pronounced rising trend, except for two downturns for Czechia.

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **104**

- Group 3 (France, Germany, Portugal, Romania, Spain) shows a U-shape distribution over the first months, due mainly to the evolution of France. It is possible to distinguish two subgroups of countries: those peaking between November and December 2020 and those peaking later during early February 2021 (Portugal and Spain).

- Group 4 (Austria, Belgium, Bulgaria, Croatia, Greece, Italy) is U-shaped over the first months and peaks between November and December 2020. Afterwards, values first decrease and then increase again in early February 2021.

- Group 5 (Ireland, Netherlands, Sweden) presents a clear U-shape trend but displays shallow ICU values (Figure 11).

For the testing indicator, two clusters were identified (Figure 12 and Figure 13):

- Group 1 (Austria, Denmark) includes two countries that could be considered testing champions, but in two different ways: Austria shows an increase of tests only after January 2021, as Denmark has been displaying an increasing trend since July 2020.

- Group 2 (all other countries) presents an increasing trend. Slovenia stands out after January 2021.
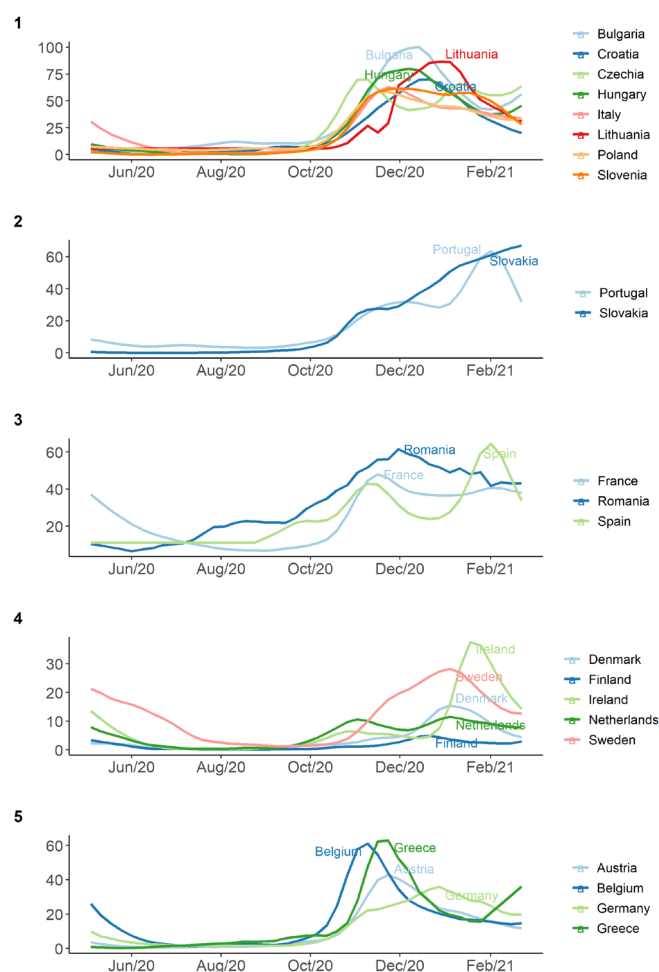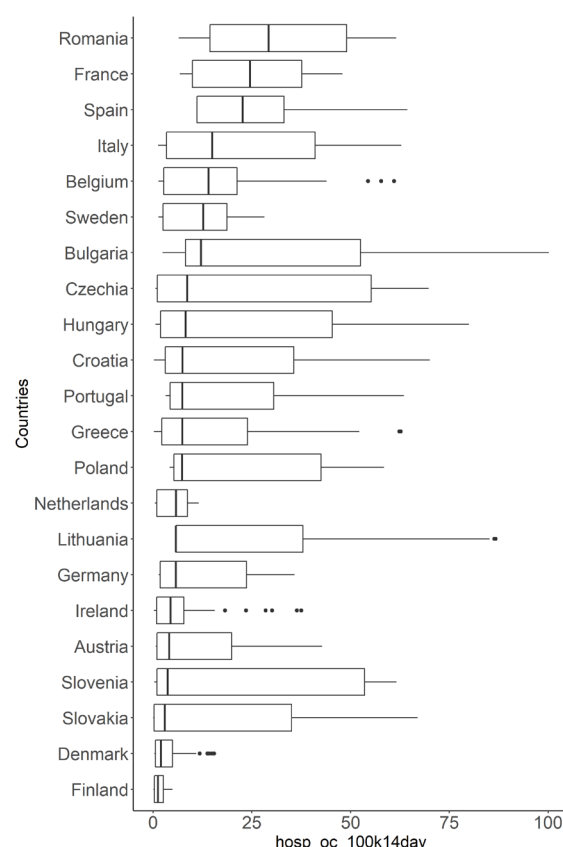
Figure 9 Boxplot of hospital occupation rate by country.

Figure 8 Hospital occupation rate (No. by 100k population) by group and country.
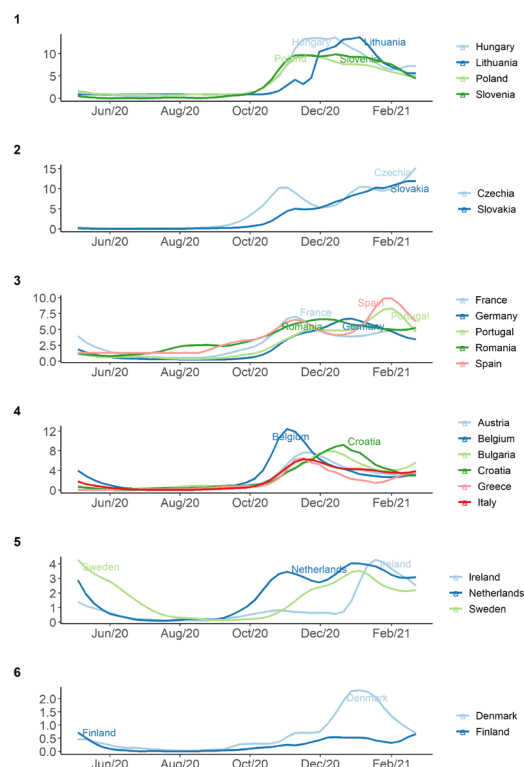
Figure 10 ICU occupation rate (No. by 100k population) by group and country.

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021
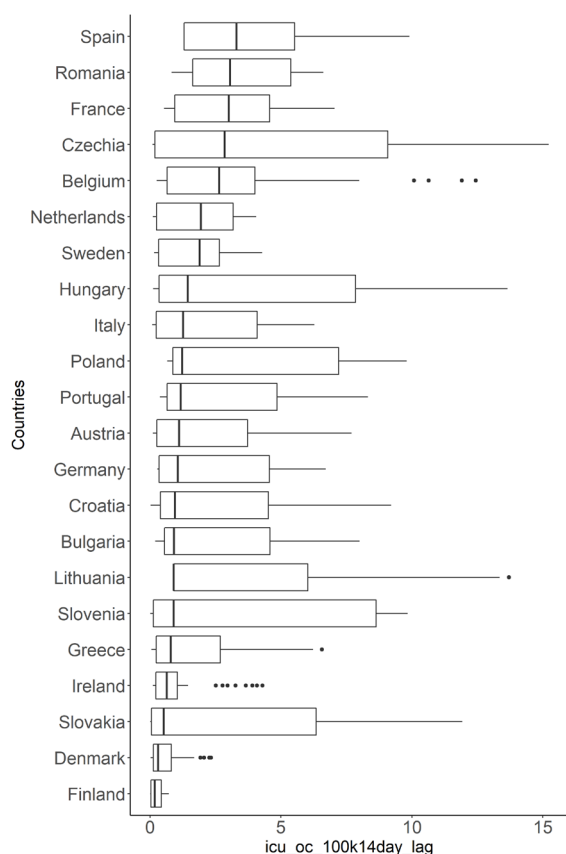
Copyright:
©2021 Tallon et al. **105**

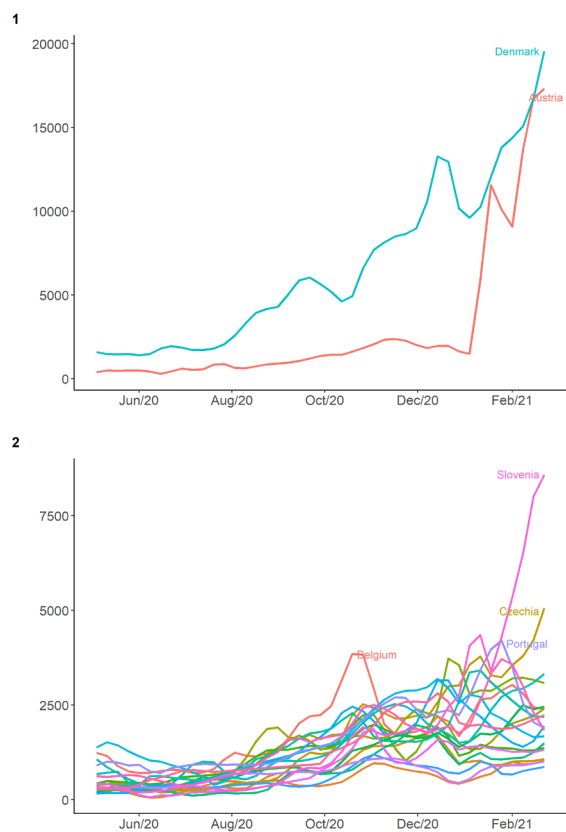**Figure 11** Boxplot of ICU occupation rate by country.

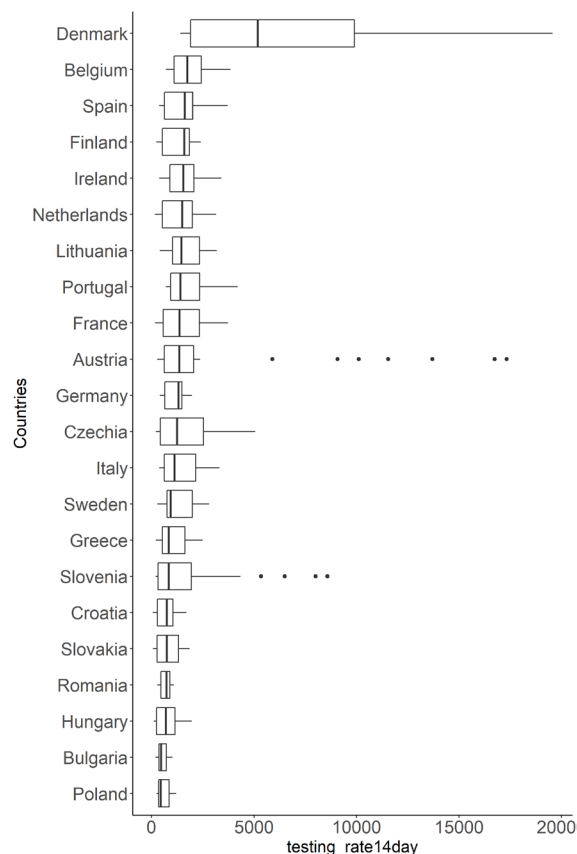**Figure 12** Testing rate (No. by 100k population) by group and country.

Figure 13 Boxplot of testing rate by country.

For the positivity indicator, four groups were selected (Figure 14 and Figure 15):

- Group 1 (Austria, Belgium, France, Germany, Italy, Netherlands, Portugal, Spain, Sweden) presents a not very clear evolution, first showing a U-shape and then two peaks: the first in November 2020 and the second in January 2021.

- Group 2 (Bulgaria, Croatia, Czechia, Hungary, Lithuania, Poland, Romania, Slovenia) peaks in November 2020 and displays a downturn in January 2021, increasing afterwards.

- Group 3 (Denmark, Finland, Greece, Ireland) shows two peaks: the first one, smaller, between October and November 2020, and the second one, higher, only for Ireland in January 2021.

- Group 4 (Slovakia) displays a steady increasing trend since June 2020.

After analysing country trajectories for each indicator separately, a multivariate approach is undertaken, where relative country trajectories will be compared, considering all indicators simultaneously. A three-way component analysis is thus applied with two complementary steps, presented over the following sections: the interstructure stage and the intrastructure stage.

Multivariate analysis: interstructure stage

The interstructure study is the first step of the three-way component analysis (double principal component) here applied. It consists of the standardised principal components analysis of the centres of gravity of clouds associated with X_t (t=1, 2, …, 43) descriptors of the incidence of COVID-19 and tests carried out in the 22 European countries studied over the forty-three 14 days' time-periods.
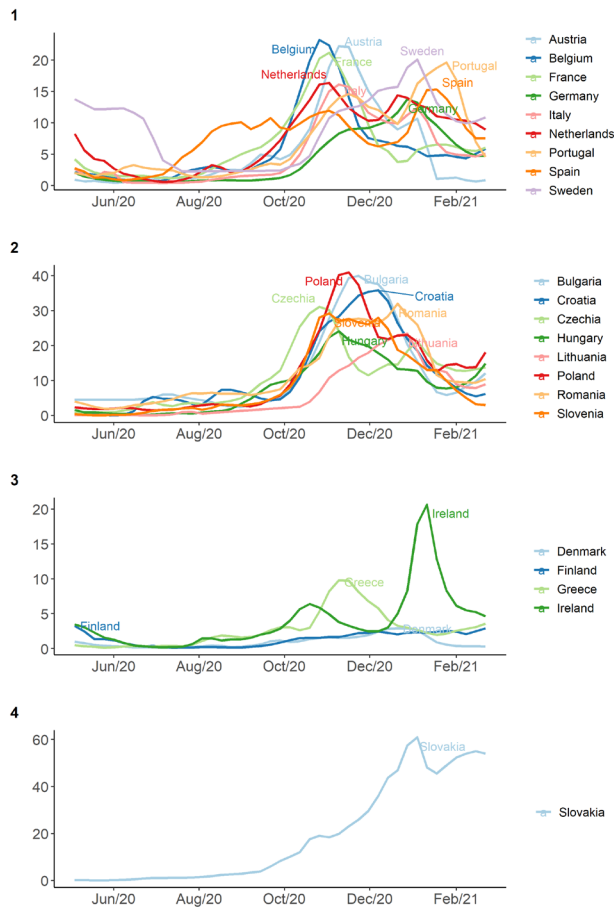
A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   106

Figure 14 Positivity rate (No. by 100 tests) by group and country.
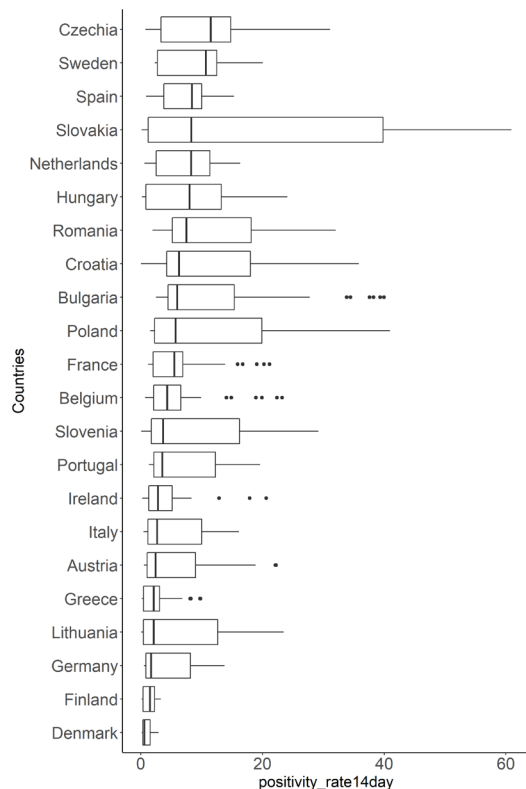

Figure 15 Boxplot of positivity rate by country.

The first two axes explain about 97% of the total inertia of the multivariate data (equal to 6, number of variables). Thus, a representation of the variables and time periods on the first factorial plane was considered (Table 2).

Table 2 Eigenvalues and inertia of interstruture

Axes	Eigenvalues	% Inertia	% Cumulative of inertia
1	5.53	92.15	92.15
2	0.29	4.82	96.97
3	0.17	2.76	99.73

All variables under study are strongly and positively correlated with the first factor (Table 3), which is then a "size factor", expressing the fact that the main variability between the time periods is quite related to the COVID-19 incidence.

Table 3 Correlation between variables and two first principal factors, interstructure (correlation circle)

	Factor 1	Factor 2
Case rate	0.973	-0.149
Deaths rate	0.965	-0.002
Hospital occupancy	0.988	-0.012
ICU occupancy	0.996	0.008
Test rate	0.881	0.448
Positivity rate	0.949	-0.256

Figure 16 represents the variables on the correlations circle of the first factorial plane, and Figure 17 shows the 43 periods considered in the present study over the first factorial plane (Table 11).
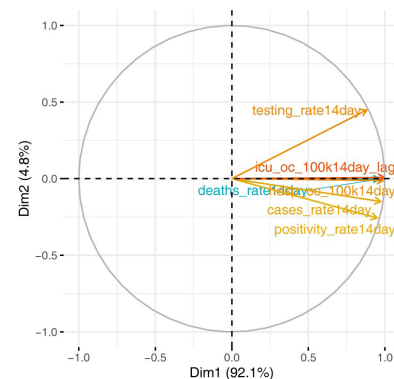

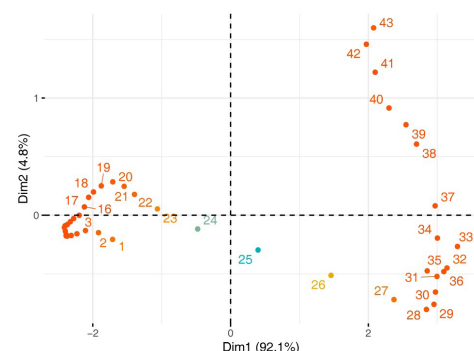**Figure 16** Representation of variables on first factorial plane (correlation circle).

Figure 17 Representation of 43 periods on first factorial plane.

Table 11 Period, year_week and dates

Period	year_week	date
1	2020-19	2020-05-04
2	2020-20	2020-05-11
3	2020-21	2020-05-18
4	2020-22	2020-05-25
5	2020-23	2020-06-01
6	2020-24	2020-06-08
7	2020-25	2020-06-15
8	2020-26	2020-06-22
9	2020-27	2020-06-29
10	2020-28	2020-07-06
11	2020-29	2020-07-13
12	2020-30	2020-07-20
13	2020-31	2020-07-27
14	2020-32	2020-08-03
15	2020-33	2020-08-10
16	2020-34	2020-08-17
17	2020-35	2020-08-24
18	2020-36	2020-08-31
19	2020-37	2020-09-07
20	2020-38	2020-09-14
21	2020-39	2020-09-21
22	2020-40	2020-09-28
23	2020-41	2020-10-05
24	2020-42	2020-10-12
25	2020-43	2020-10-19
26	2020-44	2020-10-26
27	2020-45	2020-11-02
28	2020-46	2020-11-09
29	2020-47	2020-11-16
30	2020-48	2020-11-23
31	2020-49	2020-11-30
32	2020-50	2020-12-07
33	2020-51	2020-12-14
34	2020-52	2020-12-21
35	2020-53	2020-12-28
36	2021-01	2021-01-04
37	2021-02	2021-01-11
38	2021-03	2021-01-18
39	2021-04	2021-01-25
40	2021-05	2021-02-01
41	2021-06	2021-02-08
42	2021-07	2021-02-15
43	2021-08	2021-02-22

Therefore, the first factor represents a "time factor", which highlights the contrast on the first 23 time periods (between the registered mean values at May 4 and 5 Oct 2020), when the incidence of the epidemic was quite heterogeneous within the European region and particularly reduced for several eastern and southern countries – explaining relatively limited average values for the epidemiological variables under study; and the period related to 14 days until October 19, 2020, when a steep evolution of these incidences was recorded for five consecutive weeks, then reaching the majority of European countries (beginning of the second wave).

Finally, over the last eight time periods (January 11 to Feb 2021), the COVID-19 indicators remained globally relatively high (3rd wave), with some countries experiencing a counter-cycle of divergent trends in global behaviour. However, a slight decrease in COVID-19 incidence was observed by the end of January 2021.

This may mean that this pandemic experienced a significant worsening over the European region during the second and third waves, especially deadly in the countries that had begun to ease containment measures, as it will be addressed at a later phase of this study.

Despite the residual contribution of the second axis to the inertia explained by the first two axes (4.8%), the second main component presents a positive and significant linear correlation with variable "number of tests" and a non-negligible linear correlation with variables "new cases" and "test positivity". The second axis particularly opposes periods 27-30 (November 2 until November 23, 2020), where Europe reached the peak of positivity in tests and new cases (2nd wave), to the last period of study, when this percentage decreases about 50%, with a concomitant reduction in the global number of new cases.

Complementarily, the second axis illustrates the evolution of the number of tests, especially during the third wave, registering a growth of about 70% from the beginning of the peak of that wave until the last weeks of the studied time period.

Therefore, from the second wave on, the second axis works as a "sentinel axis", perhaps signalling the effect of the alpha variant and its more accelerated contamination process, with a growing number of positive tests and new cases, as well as the progressive increase in the testing process in most countries included in this study. The percentage of positive tests thus generally decreased along the last six periods with an intensification of the testing strategy allied with a slight attenuation of the pandemic.

Multivariate analysis: intrastructure stage

On the second step of the Double Principal Components Analysis, a Ranked PCA was applied to the cloud of nT "individuals" (n=22, T=43), centred in relation to its rank centre of gravity defined by the six global epidemiological variables under study.

The trajectory of the COVID-19 incidence is represented in a system of axes generated by the normalised main principal components (Table 4).

The first two axes were selected, explaining about 92.8% of total inertia, and a representation of the variables on the first principal plane was obtained (Figure 18).

The interpretation of these first two principal factors is related to their correlation with the "compromise-position" of the variables. These coordinates are just the average correlation between the variables and the principal components in the present study (Table 5).

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **108**

Table 4 Eigenvalues and inertia of intrastruture

Axes	Eigenvalues	% Explained Inertia	% Cumulative of inertia
I	4.83	80.48	80.48
2	0.74	12.27	92.75
3	0.31	5.12	97.87

Table 5 Correlation between variables and two first factors of intrastruture

	Factor I	Factor 2	% Of the variance explained
Case rate	0.949	-0.198	93.67
Deaths rate	0.945	0.128	90.94
Hospital occupancy	0.936	0.234	93.09
ICU occupancy	0.961	0.161	94.94
Test rate	0.659	-0.746	99.08
Positivity rate	0.896	0.208	84.61
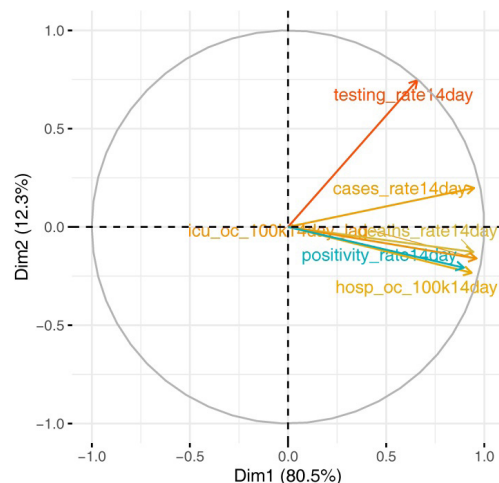
Figure 18 Representation of variables on first factorial plane, intrastruture (correlation circle).

All the variables under study are positively correlated with the first factor ("size factor"), which means that the first axis will allow positioning the countries over the 43 time periods according to the intensity of the COVID-19 incidence. Therefore, the first factor is a linear combination of the variables under study (four epidemiological variables and two variables related to testing), explaining about 81% of the global variability of the data and allowing to evaluate the evolution of COVID-19 relative incidence.

Complementarily, the second factor opposes variable "Test rate" to variables "Positivity rate" and hospital admissions, explained by the fact that more intense testing strategies tend to identify more positive cases, albeit in a smaller percentage. Therefore, the second factor essentially assesses the relative "intensity of testing" carried out by countries over the period under study.

The projection of the nT cases on the first factorial plane highlights specific relative positions of EU countries over the 43 time periods, as well as the relative pandemic peaks over successive waves, the extreme values in terms of test positivity, the degree of relative testing intensity, and, finally, the greater or lesser stability of the countries' trajectories (compared with the origin of the plane, that stands for the global average of ranks over the entire period studied, that is, the centre of gravity within a 6th-dimensional space of the global cloud of (22×43) points, associated with the juxtaposition of 22 datasets, each described over 43 time periods).

The first note to be stressed when analysing these trajectories is the relative heterogeneity and specificity of the evolution profiles of relative COVID-19 incidence and population testing. Generally, some common denominators and some contrasts stand out:

1 - The COVID-19 incidence in the EU region worsened significantly between the first and the following waves, although it burdened the National Health Systems differently in each country. A more pronounced oscillation of this incidence was registered in most EU countries, after a phase of relative deconfinement, throughout Summer 2020 and the beginning of Autumn 2020.

2 - More precisely, from late September to early November 2020, most relative trajectories revealed a sudden worsening of the pandemic situation, followed by a plateau evolution until mid-December 2020.

3 - However, the third wave took on very different intensities among the EU countries over the whole period under study. Some countries reached a peak in the last weeks of December 2020, and others, with greater severity, during the last month of January 2021, are also under the increasing influence of the alpha variant from the United Kingdom.

4 - Over the last four periods (40-43 – 14 days until February 1, 2021, to 14 days until February, 22nd, 2021), the relative incidence seemed to weaken in 10 EU countries, namely Austria, Belgium, Croatia, Denmark, Finland, Germany, Portugal, Slovenia, Spain, and Sweden. On the contrary, this is not observed for the remaining 12 countries.

Figures 19, 20, 21 and 22 illustrate the trajectories of four countries that revealed particular trends in incidence/testing and whose trajectories will be analysed in more detail, namely Austria, Denmark, Poland, and Portugal. The respective trajectories of the other countries under study may be observed in Figures 23 – 40.
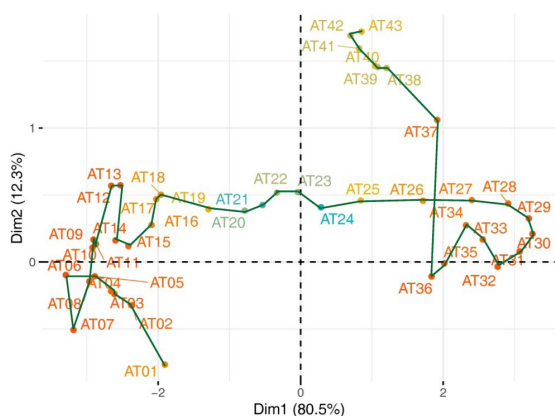
A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al. **109**

Figure 19 Austria relative incidence trajectory on first factorial plane.
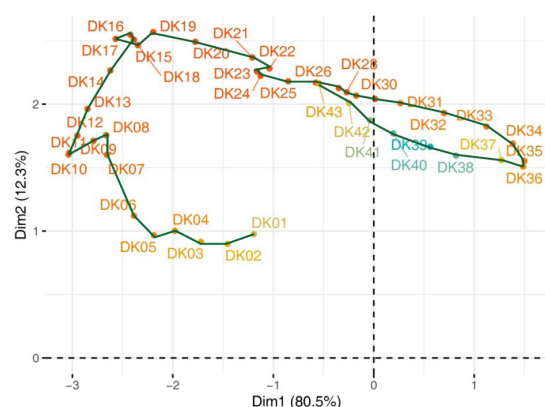

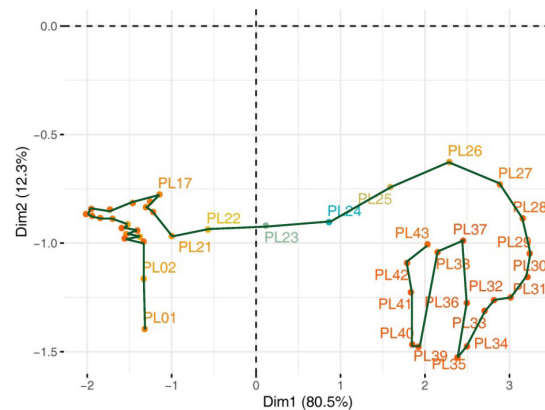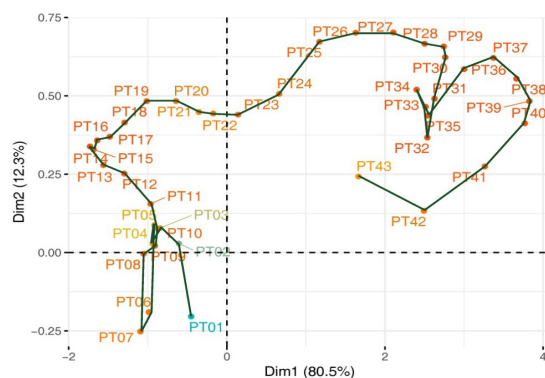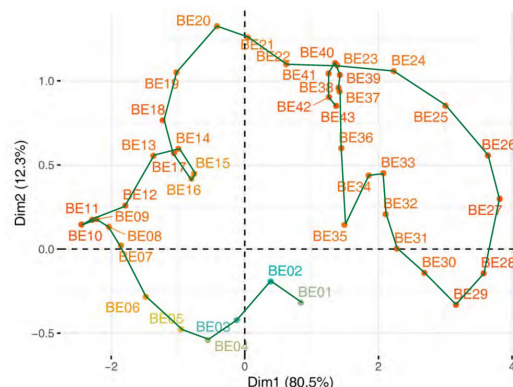**Figure 23** Belgium relative incidence trajectory on first factorial plane.

Figure 20 Denmark relative incidence trajectory on first factorial plane.

Figure 24 Bulgary relative incidence trajectory on first factorial plane.
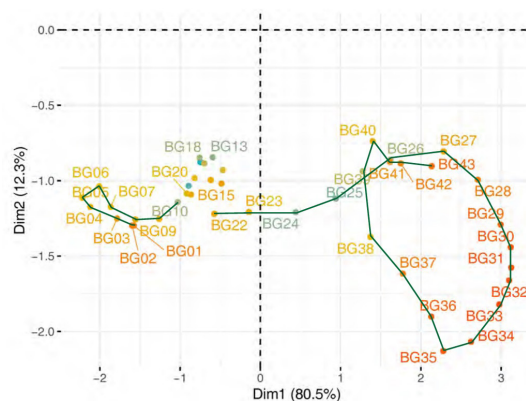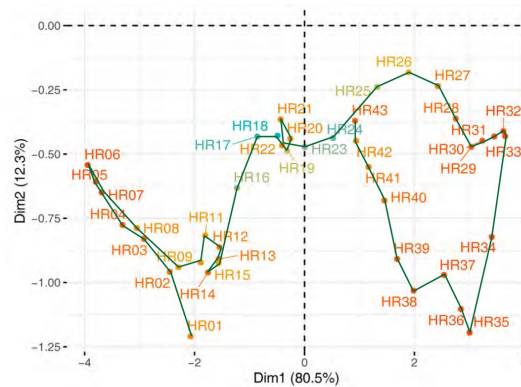

Figure 21 Poland relative incidence trajectory on first factorial plane.

Figure 25 Croatia relative incidence trajectory on first factorial plane.
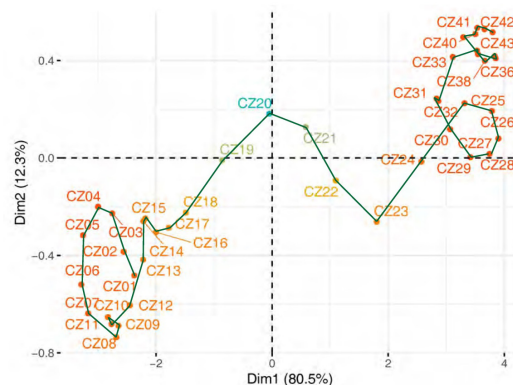

Figure 22 Portugal relative incidence trajectory on first factorial plane.

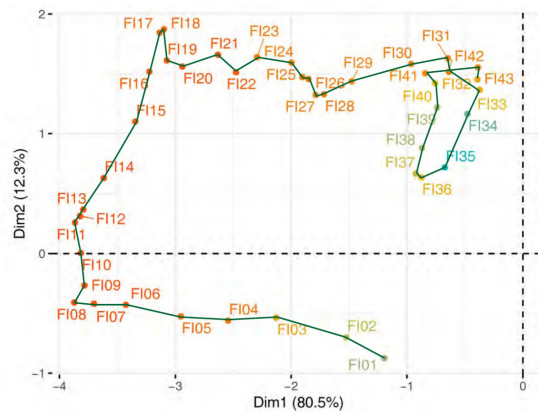**Figure 26** Czechia relative incidence trajectory on first factorial plane.

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

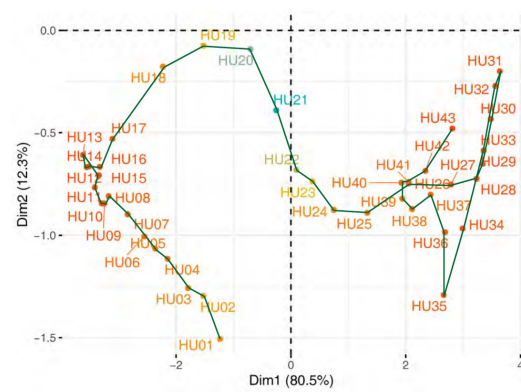Copyright:
©2021 Tallon et al.   **110**

Figure 27 Finland relative incidence trajectory on first factorial plane.

Figure 28 France relative incidence trajectory on first factorial plane.

Figure 29 Germany relative incidence trajectory on first factorial plane.

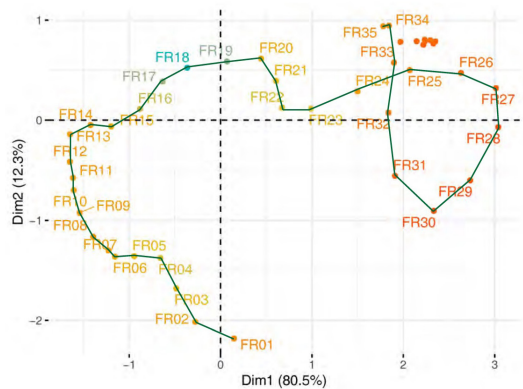**Figure 30** Greece relative incidence trajectory on first factorial plane.

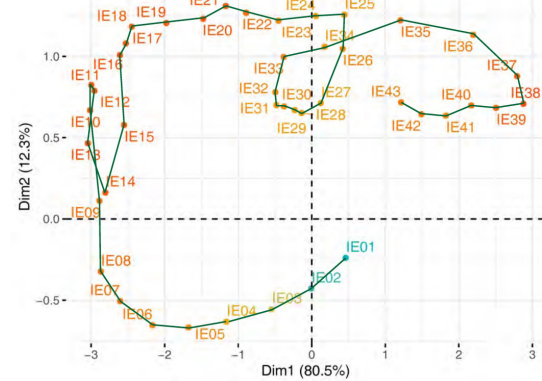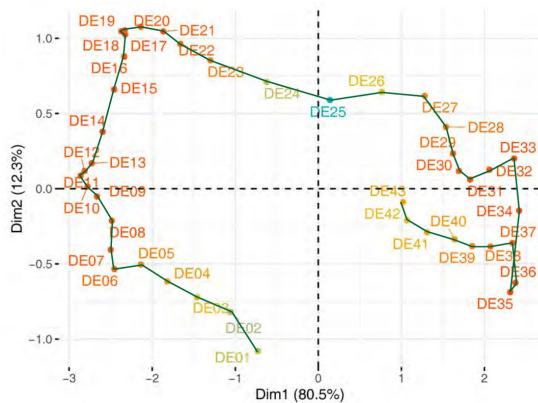**Figure 31** Hungary relative incidence trajectory on first factorial plane.

Figure 32 Ireland relative incidence trajectory on first factorial plane.

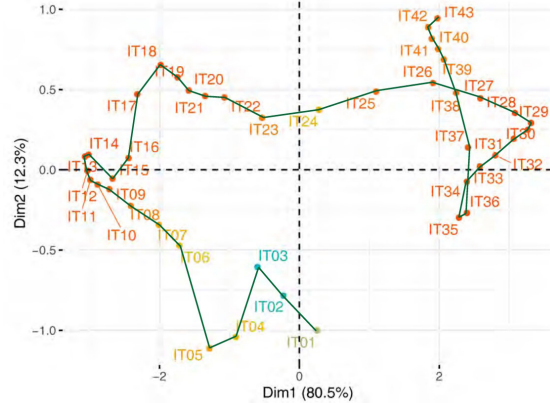**Figure 33** Italy relative incidence trajectory on first factorial plane.

Figure 34 Lithuania relative incidence trajectory on first factorial plane.

Citation: Tallon JM, Gomes P, Bacelar-Nicolau L, et al. A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021. *Biom Biostat Int J.* 2021;10(3):98–114. DOI: 10.15406/bbij.2021.10.00336

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from
May 2020 to February 2021

Copyright:
©2021 Tallon et al.   111

Figure 35 Netherlands relative incidence trajectory on first factorial plane.

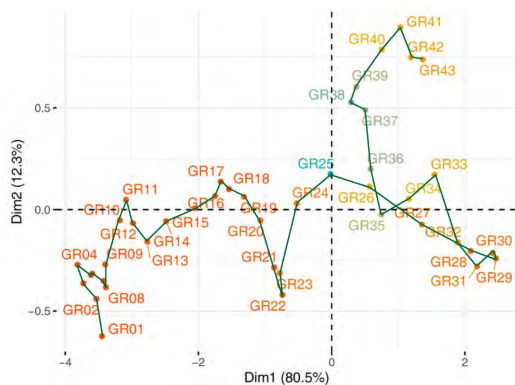**Figure 36** Romania relative incidence trajectory on first factorial plane.

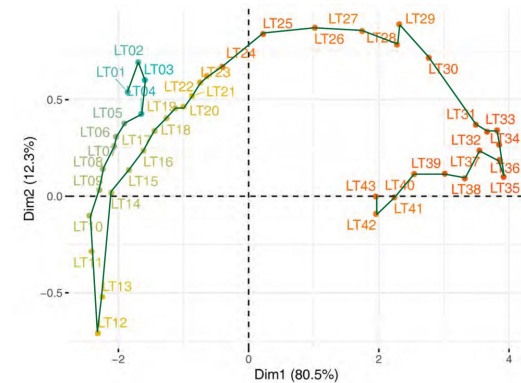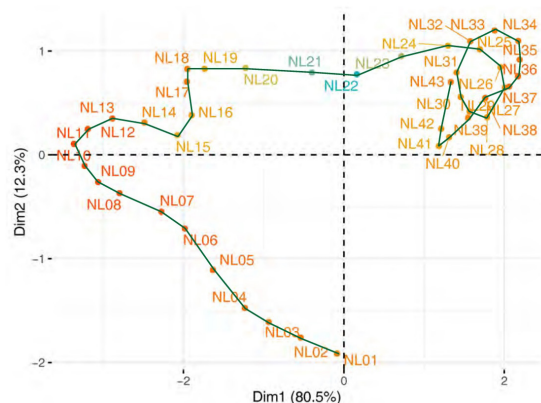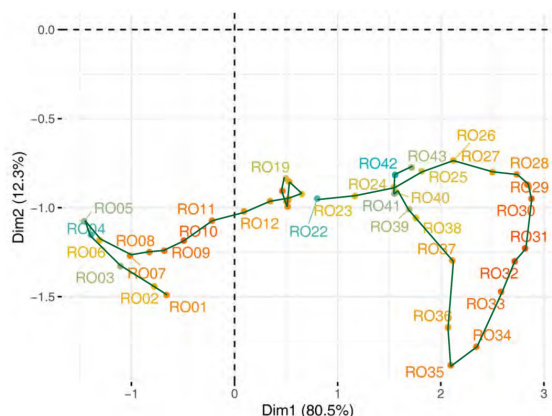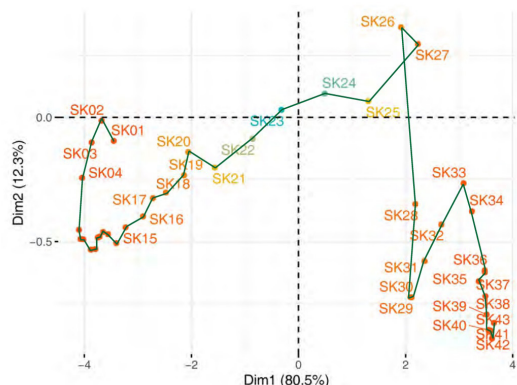**Figure 37** Slovakia relative incidence trajectory on first factorial plane.
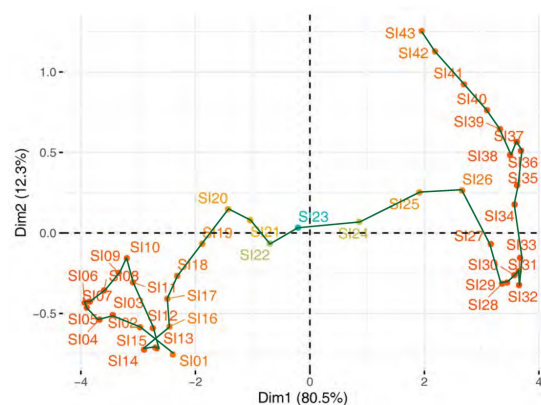

Figure 38 Slovenia relative incidence trajectory on first factorial plane.
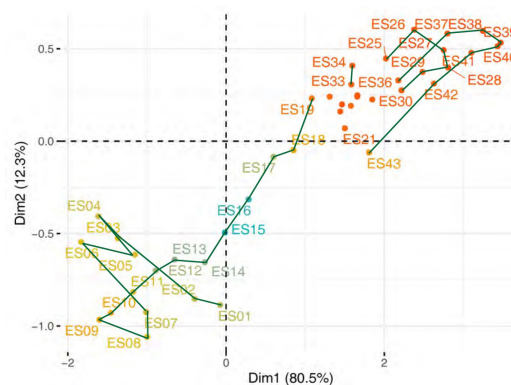

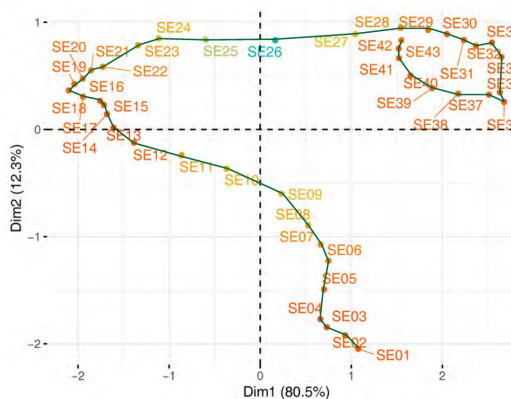**Figure 39** Spain relative incidence trajectory on first factorial plane.

Figure 40 Sweden relative incidence trajectory on first factorial plane.

Austria showed a unique profile of its COVID-19 trajectory of incidence and testing strategy: a post-first wave relatively stable and relatively unaffected by the pandemic, followed by a marked worsening after the 24th period (14 days until October 12, 2020), until a relative peak was reached during the 14th days period evaluated on the 30th period. Over the subsequent period, a pronounced reduction was detected after the 38th week (14 days until November 23, 2020), with incidence values near the robust estimate of the global average over the time period under study (Table 6).

Table 6 Austria Covid incidence in 38-43 period versus Austria Covid incidence along all the studied period

	Cases in 14 days per 100 000 inhabit.	Deaths in 14 days per million inhabit.	Hospitality occupancy per 100 000 inhabit.	ICU Occupancy per 100 000 inhabit.
Mean value (38-43) period	237.9	56.7	15.2	3.2
Trimean value (01-43) period	137.9	21.8	7.3	1.3

The location of the last periods on the main factorial plane highlighted the Austrian strategic reorientation towards large-scale testing on the post-peak pandemic period, with average values much higher than the average of testing on previous periods, as well as higher than the robust estimate of the global average of testing itself (Table 7).

In Denmark, the relative COVID-19 incidence trajectory evolves through the 1st and 2nd quadrants of the first factorial plane (Figure 20), which means that this country tested much more over all periods than the global ranks average of EU countries on the total period under study. On the other hand, the COVID-19 incidence suffered a progressive

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   112

constant worsening until period 30, but with values always below the global average. On period 31-35, Denmark registered a sudden relative worsening, reaching the peak of the second wave at the end of that time period (periods 34-35, from 14 days until December 21, 2020, to 14 days until December 28, 2020) – (Table 8). Afterwards, its trajectory goes downward, with an accompanying reinforcement of the population testing efforts. Therefore, the ranked COVID-19 incidence in Denmark was higher, in relative terms, than the global average, only during ten of the 43 periods under study.

Table 7 Tests in Austria after Covid-19 pandemic peak versus the robust mean value along all the studied period

	Tests in 14 days per 100 000 inhabit.
Mean value (38-43) period	13081.2
Trimean value (01-43) period	1344.5

Table 8 Denmark Covid incidence in 31-35 period versus Denmark incidence along all the studied period

	Cases in 14 days per 100 000 inhabit.	Deaths in 14 days per million inhabit.	Hospitality occupancy per 100 000 inhabit.	ICU Ocupancy per 100 000 inhabit.
Mean value (31-35) period	591.6	33.2	8.5	1.5
Trimean value (01-43) period	112.5	7.7	2.4	0.4

In Poland, the COVID-19 trajectory evolves through the 3rd and 4th quadrants (Figure 21), which shows that the testing process in this country has been far below the EU's global average performance. Over the first 21 weeks, the evolution of the pandemic incidence was relatively stable, with an average of new cases in 14 days significantly lower than in Central and Southwestern European countries. However, after the 22nd week (beginning of October 2020), a severe relative deterioration was observed, worsening until the peak was reached on periods 29-30 (3rd-4th week of November 2020) (Table 9), followed by a small downward trajectory until period 35 with a relative "plateau" behaviour " over the three subsequent weeks. Finally, it experienced a slightly more favourable evolution over the last four time periods under study. Therefore, since the beginning of October 2020, Poland has experienced a relative pandemic incidence significantly higher than the EU global average, registering some of the highest EU ranks.

Table 9 Poland Covid incidence in 23-30 period versus Poland incidence along all the studied period

	Cases in 14 days per 100 000 inhabit.	Deaths in 14 days per million inhabit.	Hospitality occupancy per 100 000 inhabit.	ICU Occupancy per 100 000 inhabit.
Mean value (23-30) period	547.0	84.3	36.5	6.0
Trimean value (01-43) period	109.4	30.0	15.6	2.6

Portugal, after an initial period of relative stability of COVID-19 incidence and relative improved testing indicators, registered a sudden worsening of incidence indicators in the middle of the second wave, during October 2020, until reaching a peak in period 30 (November 10 to 24, 2020), with the new cases in 14 days mean reaching 752.8 cases per 100,000 inhabitants, corresponding to about 5,400 cases per day. A calmer period followed, with a slightly downward trajectory and stability regarding its relative position, until the fortnight ending on December 22, 2020. Nevertheless, at the end of December 2020, a slightly worsening, followed by a downward trajectory in the COVID incidence, was observed on the period 35 to 39 (3rd wave) (Table 10), when the average number of new cases in14 days increased from 558.5 to 1649.4 per 100 000 inhabitants. Following severe containment measures set up by the Portuguese authorities, a sharp downward trajectory was seen during about four weeks, reaching at the end of February 2021 an average number of new cases of about 174 per 100 000 inhabitants. Concomitantly, Portugal reached, at the peak of the 3rd wave, an average of about 364 deaths in 14 days per million inhabitants. Finally, although testing may have increased over the last weeks of the period under study, it weakened relative to the also increasing efforts of other EU countries analysed. Globally, though, results indicate that testing in Portugal was always higher than the global average rank value achieved in the EU from September 2020 onward (Figure 22).

Table 10 Portugal Covid incidence in 35-39 period versus Portugal incidence along all the studied period

	Cases in 14 days per 100 000 inhabit.	Deaths in 14 days per million inhabit.	Hospitality occupancy per 100 000 inhabit.	ICU Occupancy per 100 000 inhabit.
Mean value (35-39) period	1162.9	207.2	40.9	6.5
Trimean value (01-43) period	200.2	36.7	12.4	2.0

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **113**

Discussion and conclusions

The preliminary exploratory statistical analysis undertaken allowed us to study the marginal empirical distributions of variables and the similarities and dissimilarities between country trajectories for each variable separately. This more detailed approach may be useful for decision makers to pinpoint each country's position regarding each indicator, which may help decisions to implement more or less strict specific measures that may impact these health and health systems indicators.[15-17]

However, a more detailed approach may also make it more challenging to understand which indicators are globally more associated with variability across the various periods under study within the EU. It may also become harder to compare relative country trajectories regarding all these indicators simultaneously.

Therefore, the global trend of the COVID-19 pandemic across the EU region regarding all these indicators simultaneously and comparing relative country trajectories becomes much clearer when a three-way component analysis is applied, respectively, on the interstucture on the intrastructure stage.

The first objective of this cross-sectional study was to investigate the EU countries' evolution of COVID-19 incidence over ten months from May 2020 until February 2021, and, therefore, the successive pandemic waves that occurred within this period, preceding the start of the general vaccination process.

The epidemiological indicators under study were registered weekly for 43 periods from May 4, 2020, until February 22, 2021, evaluating, for each period, the average number of new cases, new deaths, hospital admissions, ICU admissions, tests, and percentage of positive tests, covering the fourteen days preceding the period date.

A dataset $X^{(t)}$ was associated with each period, describing the COVID-19 incidence for a group of 22 countries, covering about 99% of the total EU population. In line with the proposed Double Principal Components method, an exploratory multivariate analysis was proposed, taking simultaneously into account countries' epidemiologic information and time trajectories. Firstly, each $X^{(t)}$ data set was represented by a six-dimensional vector, the centre of gravity of the respective cluster $N^{(t)}(i=1, ..., 43)$. The Principal Components Analysis of the dataset, taking as rows the 43 centres of gravity, offered a picture on the first principal plane of the global pandemic evolution along the ten months under analysis (interstructure).

This representation clearly illustrates the existence of three phases in the evolution of the pandemic over the EU territories:

- The first phase covers periods 1 to 23 (between May and October 2020) and displays a relative heterogeneity, with a group of 8 countries facing a downward pandemic trajectory and experiencing a process of deconfinement in multiple forms, while the remaining countries suffer with different intensity a gradual increase of new COVID-19 cases, anticipating that Europe would sooner or later be overwhelmed by new pandemic waves over the next phase. Therefore, the graphical multivariate representation of these initial periods still showed that, in global terms, the relative values of COVID-19 EU incidence were quite stable. However, the univariate exploratory analysis carried out in this paper identified different country profiles with specificities incidence or testing strategies.

- Over the following periods (24-33, between October 12 and December 14, 2020), the pandemic has successively, and sometimes intensely, worsen throughout the EU territory, with this 2nd wave reaching its peak within the 14 days period ending on December 15, 2020.

- The COVID-19 incidence would remain high in most countries under study, although showing some counter-cycle behaviours, partly explained by the more or less severe impact of the alpha variant (affecting the UK from October 2020) on this 3rd wave. At the same time, the first factorial plane also revealed a growing focus from national authorities to promote large-scale COVID testing campaigns, when the certainty of the 3rd wave reaching the European region became indisputable.

The second aim of our analysis was to show the benefits of applying a three-way statistical analysis of the data to study the combined behaviour of the 12 countries over the 43 time periods. The existence of several outliers and the relative heterogeneity of each variable under study when covering all countries throughout all periods recommended using a non-parametric approach, transforming the values of the variables into rank statistics. This transformation into ranks, under the condition that the variables under study present a more or less continuous distribution and with few ties, increased the homogeneity of the matrix to be analysed.

The principal components analysis of this joint dataset essentially showed that all variables appear again as positively correlated with the 1st factor (size factor), which allows assessing the relative pandemic incidence in any country in direct comparison with the global average, represented by the origin of the plane. Similarly, the 2nd axis also explains the relative degree of testing within each country, distinguishing in the 1st and 2nd quadrants the countries and periods with a higher level of testing than the global average rank.

Consequently, it was possible to achieve the central objective of our research of evaluating in the first factorial plane the relative pandemic and level of testing trajectories in each country under study:

- The 1st quadrant includes the countries/periods with an above rank average incidence and an above rank average testing.

- The 2nd quadrant contains countries/periods with a below rank average incidence and an above rank average testing.

- The 3rd quadrant covers countries/periods with a below rank average incidence and a below rank average testing.

- The 4th quadrant comprises countries/periods with an above rank average incidence and a below rank average testing.

The very high percentage of inertia restored by the first two axes gives this representation an accurate description of the relative position of each country in comparison with the global rank average. Additionally, our approach emphasises the several specificities of countries trajectories.

Future research paths will deepen this study by promoting a classificatory approach using different distance metrics to compare the 22 trajectories. An effort will also be undertaken to complement this approach linking the relative country trajectories to other characteristics of countries that may contribute to a more or less severe COVID-19 impact and evolution between countries and within each country: such as socio-economic indicators; health indicators like health comorbidities;[18] health systems indicators like the existence of linked electronic records;[19,20] non-pharmaceutical measures implemented, including communication effectiveness.[21,22]

This three-way approach thus may make it possible to identify different country trajectories profiles throughout successive pandemic

A three-way multivariate data analysis: comparison of EU countries' COVID-19 incidence trajectories from May 2020 to February 2021

Copyright:
©2021 Tallon et al.   **114**

waves and counter-cyclical behaviours, which might contribute to harmonising public policies throughout the EU, thus improving equity and effectively tackling current issues and future pandemics within this region.

Acknowledgments

None.

Conflicts of interest

The authors report no conflicts of interest in this work.

References

1. Kagan D, Moran-Gilad J, Fire M. Scientometric trends for coronaviruses and other emerging viral infections. *bioRxiv. Published online*. 2020:1–24.

2. Aronson JK. Coronaviruses – a general introduction. *CEBM - Centre for Evidence-Based Medicine*. 2020.

3. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med*. 2020;172(9):577–582.

4. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019- nCoV) infections among travellers from Wuhan, China, 20 28 January 2020. *Eurosurveillance.* 2020;25(5):2000062.

5. Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med*. 2020;382:1177–1179.

6. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26(5):672–675.

7. Tallon JM, Gomes P, Bacelar–Nicolau L. Profiling European countries on COVID–19 prevalence and association with non–pharmaceutical interventions. *Biom Biostat Int J*. 2020;9(4):118–130.

8. Tallon JM, Gomes P, Bacelar–Nicolau L. Comparative prevalence of COVID–19 in european countries: a time window at second wave. *Biom Biostat Int J*. 2020;9(6):196–207.

9. European Centre for Disease Prevention and Control. COVID–19 Databases, Stockholm: ECDC; 2021.

10. Lebart L, Piron M, Morineau A. *Statistique Exploratoire Multidimensionnelle 4ème edition*. *Dunod* 2006.

11. Dazy F, LeBarzic JF. *L'analyse des données évolutives. Edts Technip*. 1996.

12. R Core Team. *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna*, Austria. 2021.

13. Handhika T, Lestari DP, Sari I. Multivariate time series classification analysis: State-of-the-art and future challenges. *InIOP Conference Series: Materials Science and Engineering.* 2019;536(1):012003.

14. Ruiz AP, Flynn M, Large J,et al. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*. 2021;35:401–449.

15. European Centre for Disease Prevention and Control. Rapid Risk Assessment: *Coronavirus disease*. 2019.

16. European Centre for Disease Prevention and Control. Coronavirus disease 2019 (COVID–19) in the EU/EEA and the UK – tenth update, 11 June 2020, ECDC. 2020.

17. Ferguson N. Report 9: Impact of non–pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand – Working paper, *Imperial College COVID–19 Response Team*, London. 2020.

18. Nogueira PJ, de Araújo Nobre M, Costa A, et al. The role of health preconditions on COVID-19 deaths in Portugal: Evidence from surveillance data of the first 20293 infection cases. *J Clin Med*. 2020;9(8):2368.

19. Bacelar-Nicolau L, Rodrigues T, Fernandes E, et al. Picturing inequities for health impact assessment: linked electronic records, mortality and regional disparities in Portugal. *Impact Assessment and Project Appraisal*. 2018;36(1):90–104.

20. Bacelar-Nicolau L, Rodrigues T, Fernandes E, et al. Helping decision-makers visualize inequities in health impact assessment: linked electronic records, mortality and regional disparities in Portugal. *Value in Health*. 2016;19(7):623.

21. Bernardino M, Bacelar-Nicolau L. The importance of reliable social media information during the COVID-19 pandemic", *European Journal of Public Health*. 2020;30(5):165.067.

22. Bacelar-Nicolau L. The still untapped potential of social media for health promotion: The WHO example. *DPH*. 2019;125.