"Unveiling the underlying severity of multiple pandemic indicators"

Manuela Alcañiz, Marc Estevez and Miguel Santolino

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública

Abstract

Objective: Multiple interconnected key metrics are frequently available to track the pandemic progression and one of the difficulties health planners face is determining which provides the best description of the status of the health challenge. In this study three COVID-19 indicators broadly used to monitor the evolution of the pandemic are analysed: the numbers of daily hospitalisations, ICU admissions and deaths attributable to the disease. The aim of the paper is to capture the information provided by these magnitudes in a single metric that reveals the underlying severity. Methods: Drawing on official Spanish data, we use one-sided dynamic principal components to convert a multivariate framework in a univariate scheme. The time-varying relationship between underlying severity and the number of positive cases is estimated. Results: A single component adequately explained the variability of the indicators during the analysed period (May 2020–March 2022). The severity indicator was stable up to mid-March 2021, fell sharply until October 2021, before stabilising again. The period marked by a fall coincided with the period of massive vaccination. By age group, the association between underlying severity and positive cases in those aged 80+ was almost 20 times higher than in those aged 20-49. Conclusions: Our methodology can be used in other infectious diseases to provide policy makers with a single metric that describes the severity status of the disease and enabling them to monitor the evolution. The synthetic indicator may be useful for prioritizing the vaccination of high-risk groups and evaluating the severity reduction.

JEL classification: C13, C32, I10, I18.

Keywords: Pandemics, COVID-19, Patient Acuity, Hospitalization, Intensive Care Units, Vaccination.

Manuela Alcañiz: Riskcenter-IREA, Dept. Econometrics, Statistics and Applied Economy, Universitat de Barcelona (UB). Email: malcaniz@ub.edu

Marc Estévez: Riskcenter-IREA, Dept. Econometrics, Statistics and Applied Economy, Universitat de Barcelona (UB). Email: marcestevez18@gmail.com

Miguel Santolino: Riskcenter-IREA, Dept. Econometrics, Statistics and Applied Economy, Universitat de Barcelona (UB). Email: msantolino@ub.edu

INTRODUCTION

The coronavirus (COVID-19) pandemic declared by the World Health Organization in 2020 has been a public health emergency of international concern affecting all areas of health and wellbeing. The unfavourable evolution of this infectious disease, caused by the SARS-CoV-2 virus, represented a major public health challenge including, among others, pressure on the health system, loss of human lives and a decrease in life expectancy [1,2].

Most people infected with the virus experienced mild to moderate respiratory illness and recover without needing hospitalization; however, a significant number fallen seriously ill and required regular hospitalization or admission to an intensive care unit (ICU). Some failed to survive the disease. One of the difficulties encountered in monitoring the evolution of COVID-19 has been selecting an indicator that best describes the pandemic status when multiple intercorrelated metrics are available [3]. Here, we undertake a joint analysis of three COVID-19 severity indicators: namely, the number of hospitalizations, the number of ICU admissions and the number of deaths. These three metrics provide complementary information on the evolution of the pandemic and are closely correlated, and suggest that a dimension reduction approach could result in a synthetic univariate indicator that is capable of capturing the severity of the disease at any given point in time.

Previous research has, in the main, attempted to assess the evolution of disease severity based on single magnitudes, addressing, for example, either the number of hospital admissions [4-6], hospital bed occupancy [7,8], the number of ICU admissions [9-11], ICU bed occupancy [12-14], or the number of deaths [15-17]. Other studies have used multivariate principal component techniques, but again focusing on a single COVID-19 indicator, albeit for multiple countries [18-20]. This approach sheds light on the underlying patterns common to all countries for the specific COVID-19 indicator under consideration and allows countries to be clustered according to their similarities in evolution. Adopting a static approach, the first two principal component coefficients has been used to identify different groups of countries depending on the behaviour of their respective numbers of positive cases and deaths in different periods of the pandemic [20]. Adopting a non-static approach, other studies have used dynamic functional principal components over a single COVID-19 indicator in multiple countries to forecast the cumulative number of diagnosed cases per million individuals [21] or the daily growth rate in the number of positive cases [22].

The aim of our study differs from this preceding line of investigation. We seek to construct a synthetic indicator that, based on the multivariate information provided by the numbers of hospitalizations, ICU admissions and deaths due to COVID-19, captures the evolution of disease severity in a single magnitude. As one-sided dynamic principal components (ODPC) is a methodology that allows the decomposition of multivariate time-series data into a few principal component series [23], we aim to explore whether its application to official Spanish data for the period May 2020 to March 2022 results in a single component capable of explaining the joint evolution of the three disease-severity time series. This would, first, provide us with a component that reveals the underlying severity of COVID-19 and, second, enable us to analyse in a straightforward fashion the relationship between the synthetic indicator proposed and the number of positive cases detected at each point in time. As a result, we would be able to quantify the impact of the COVID-19 vaccination program on this underlying severity, considering the Spanish population as a whole and by age groups.

To the best of our knowledge, this is the first time that this method has been applied to time series of different COVID-19 indicators. Even the static principal component analysis (SPCA) approach, widely used in climatology when dealing with time data [24, 25], has been rarely applied to reduce the dimensionality of a set of COVID-19 indicators. One exception is Swallow et al. [3] who conduct their analysis on the number of daily positive cases, deaths, hospitalizations and hospital occupancy with mechanical ventilation support in the United Kingdom. To account for the temporal dimension, the data were rotated to remove the effect of the temporal correlation and an SPCA was conducted on the weighted data matrix. The temporal weighting matrix in that approach uses the correlation of residuals of a smoothed regression model to introduce dependency on previous time points. Our proposal follows a different approach. Given that we extend the original time series matrix by including their lagged values, original time series can be easily expressed in terms of the components and their lags. We believe that this approach may have advantages in the interpretation of results when, as is the case here, there is sequentiality in the time series data, i.e., deaths occur later than hospitalizations and ICU admissions.

METHODS

Data

In conducting this study, free-access datasets have been used. The daily number of COVID-19 detected cases, hospital admissions, ICU admissions and deaths were obtained from Spain's National Centre of Epidemiology (https://cnecovid.isciii.es). Information for each of these time series is disaggregated in the dataset by province of residence, age and gender for a period that extends from 11 May 2020 to 27 March 2022. Multiplicative weekly seasonality was observed in time-series with lower values being presented during weekends. The seasonal effect was adjusted using the LOESS method for seasonal-trend decomposition (STL) [26]. In addition, the Nadaraya–Watson kernel smoother was applied to remove the noise of the resulting time series [27,28]. Figure 1 shows the original and smoothed COVID-19 time series for the period of observation. Stationarity of the time series was investigated to avoid spurious results when analysing the association between time series [29]. Once weekly seasonality and noise were removed from the original COVID-19 indicators, the stationarity of the resulting time series was confirmed using the augmented Dickey-Fuller (ADF) test.

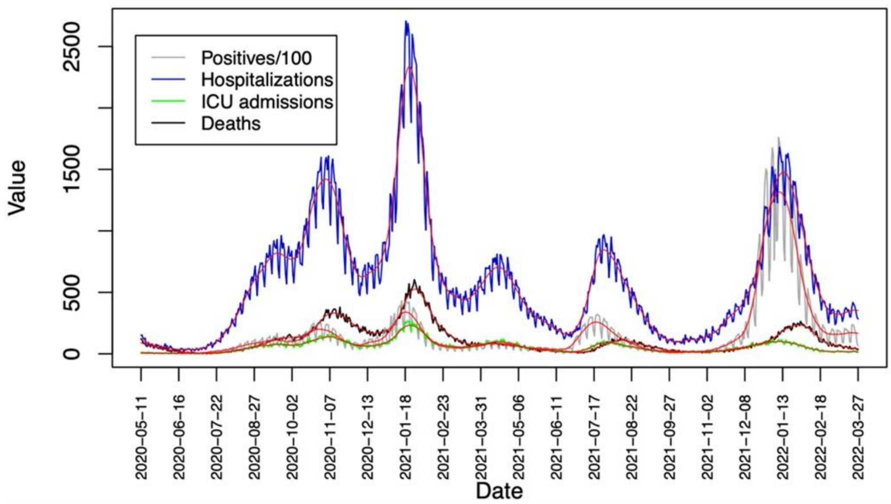

Figure 1. Original COVID-19 time series with smoothed values (red lines) for Spain. Period from 11 May 2020 to 27 March 2022.

One-sided dynamic principal components

Dimension reduction is critical in multivariate vector time series for finding simplifying structures or factors. The application of ODPC is useful when the variability of different time series can be explained with a small number of components [23]. This occurs when time series are highly correlated. Let's consider the vectors of stationary time series $\boldsymbol{z}_1, \dots, \boldsymbol{z}_T$, with $\boldsymbol{z}_t = (z_{t,1}, \dots, z_{t,m})', t = 1, \dots, T$. The ODPC can be defined as linear combinations of present and previous values of the series that minimize the mean square error of the reconstruction. We define the first on-sided dynamic principal component as:

$$f_t(\boldsymbol{a}) = \sum_{h=0}^{k_1} \boldsymbol{z}'_{t-h} \boldsymbol{a}_h, \quad t = k_1 + 1, \dots, T \tag{1}$$

where $\boldsymbol{a}' = (\boldsymbol{a}'_0, \dots, \boldsymbol{a}'_{k_1})$, being $\boldsymbol{a}'_h = (a_{h,1}, \dots, a_{h,m})$, $h = 0, \dots, k_1$, the coefficients associated with the lagged values of the time series, and $k_1 \geq 0$ an integer denoting the number of lags used to compute the dynamic principal component. Only the first component is shown, given that it is the only one computed in this study. More details about how to perform ODPC using more than one component can be found in [23].

Then, defining a matrix $\boldsymbol{B}' = [\boldsymbol{b}_0, \dots, \boldsymbol{b}_{k_2}]$, $\boldsymbol{b}_h \in \mathbb{R}^m$, $h = 0, \dots, k_2$, the lagged values of the dynamic principal component can be used to reconstruct the original time series $\boldsymbol{z}_t$ as

$$\boldsymbol{z}_t^R(\boldsymbol{a}, \boldsymbol{B}) = \sum_{h=0}^{k_2} \boldsymbol{b}_h f_{t-h}(\boldsymbol{a}), \quad t = k_1 + k_2 + 1, \dots, T \tag{2}$$

where $k_2 \geq 0$ is an integer indicating the number of lags of the principal component to be used in the reconstruction.

The optimal values $(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{B}})$ of $\boldsymbol{a}$ and $\boldsymbol{B}$ are defined as those that minimize the mean squared error (MSE) in the reconstruction of the data, which is defined as

$$MSE(\boldsymbol{a}, \boldsymbol{B}) = \frac{1}{T'm} \sum_{t=(k_1+k_2)+1}^{T} \|\boldsymbol{z}_t - \boldsymbol{z}_t^R(\boldsymbol{a}, \boldsymbol{B})\|^2$$

where $T' = T - (k_1 + k_2)$ and $\|\cdot\|$ is the Euclidean norm. Note that if $k_1 = k_2 = 0$, the first ODPC is simply the first ordinary principal component of the data.

Time-varying coefficient linear model

To meet our second objective and study how the underlying severity indicator and the number of COVID-19 positive cases are related, a time-varying coefficient linear model (TVLM) is used. Compared to the classical linear model, the TVLM is distinctive insofar as it allows the coefficients associated with the independent variable to vary over time.

Thus, a TVLM with one regressor of the form $\boldsymbol{y}_t = \boldsymbol{x}_t \beta_t + \boldsymbol{\varepsilon}_t$, $t = 1, \dots, T$, is proposed, where $\boldsymbol{y}_t$ is the response variable, $\boldsymbol{x}_t$ is the independent variable and $\boldsymbol{\varepsilon}_t$ is the error term [30]. Note that the $\beta_t = \beta\left(\frac{t}{T}\right)$ depends on a smoothing real-valued function that varies over time, $\beta: [0,1] \to \mathbb{R}$. The regressor $\boldsymbol{x}_t$ is rescaled by its expected value to be centred at one. This model is best estimated by combining ordinary least squares and the local polynomial kernel estimator [31]. A bandwidth must be selected to indicate the size of the window in which weighted local regressions are estimated. The selection takes at most k previous and k future observations for the estimation of the local regression at each period, where k is an integer number [32].

RESULTS

The analysis was conducted using the R software package. Package *odpc* version 2.0.5. was used to compute the ODPC [33]. R Package *tvReg*, version 0.5.7. was used to carry out to compute the TVLM [30].

Synthetic indicator of underlying severity

High pairwise correlation coefficients were obtained between the number of hospital admissions, ICU admissions and number of deaths, with values between 0.87 and 0.95. Thus, we performed an ODPC analysis to construct a single severity indicator capable of capturing the information from these three COVID-19 series. After rescaling the three time series, the alternating least-squares algorithm and the minimization of the information criteria were applied to select the optimal number of lags [23]. Both approaches led to the conclusion that the optimal number was one. The optimal estimated coefficients of vector $\boldsymbol{a}$ were:

$$\widehat{\boldsymbol{a}}' = (0.03 \quad -0.45 \quad -0.21 \quad 0.22 \quad 0.65 \quad 0.52)$$

This vector $\widehat{\boldsymbol{a}}$ contains the estimated weights in the linear combination to construct the dynamic principal component in (1), which captures the underlying severity of COVID-19. Its first three values are the coefficients associated, respectively, with the number of hospitalizations, ICU admissions and deaths in period t, while the following three coefficients are those associated with the one-lagged values of the same series.

The MSE of the optimal ODPC was 0.031. This value is considerably lower than the MSE associated with one component in SPCA (0.051). Of the variability of the three corrected severity time series, 95.83% was explained by the optimal model. Figure 2 shows the three standardized series used to capture the severity of COVID-19, plus the component obtained from the ODPC analysis, which captures this evolution in a synthetic indicator.

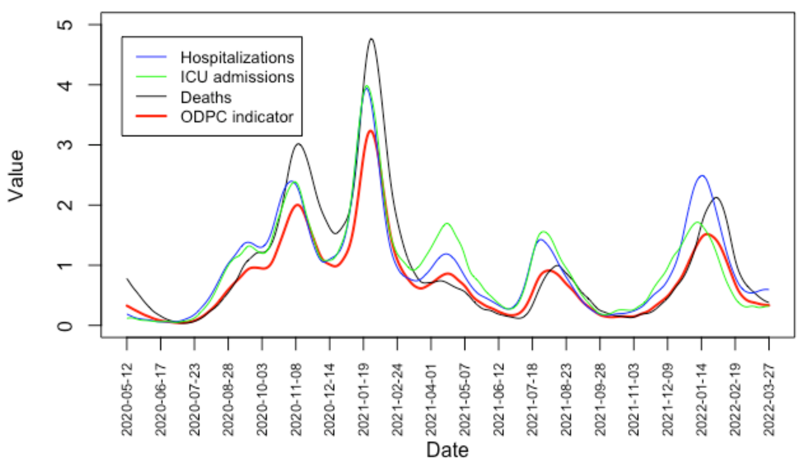

Figure 2. Standardized smoothed COVID-19 time series and ODCP indicator for underlying severity. Period from 11 May 2020 to 27 March 2022.

The number of positive cases diagnosed presented a moderate correlation with the respective numbers of hospital admissions (0.56), ICU admissions (0.36) and deaths (0.30), which seems to suggest that disease severity is not constant over time. This moderate degree of correlation between positive cases and the other indicators can also be observed in Figure 1.

Reconstruction and prediction of severity indicators

The underlying severity indicator can be used with matrix $\widehat{\boldsymbol{B}}$ to reconstruct the standardized corrected COVID-19 time series, as shown in (2). The resulting estimation of matrix $\boldsymbol{B}$ was the following:

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 7.94 & -6.77 \\ 6.80 & -5.66 \\ -1.81 & 3.31 \end{pmatrix}$$

Thus, to reconstruct the hospitalization and ICU admissions time series, the underlying severity indicator has to be multiplied by a positive scalar (7.94 and 6.80, respectively) and the one-lagged indicator by a negative scalar (-6.77 and -5.66, respectively). The series of deaths, on the other hand, is reconstructed by multiplying the underlying severity indicator by a negative scalar (-1.81) and the one-lagged indicator by a positive scalar (3.31). The different signs of the coefficients for reconstructing these original time series indicate that deaths occur after hospital and ICU admissions. In contrast with non-dynamic principal components, the flexibility of ODPC allows us to capture this sequentiality between the time series.

The reconstructed standardized time series of the number of hospitalizations, ICU admissions and deaths are shown in Figure 3 (upper panel). Note that an accurate reconstruction of these three COVID-19 time series is achieved, especially as regards hospitalizations and deaths. The MSE was 0.022 for the number of hospital admissions, 0.038 for the number of ICU admissions and 0.033 for the number of deaths. The underlying severity indicator can then be used to predict the number of hospitalizations, ICU admissions and deaths. First, time series forecasting of the future behaviour of the severity indicator is conducted using a SARIMA model [34]. The predicted severity indicator is multiplied by $\widehat{\boldsymbol{B}}$ to reconstruct the forecasted COVID-19 time series. SARIMA-based forecast residuals are included in the predictions of the COVID-19 time series. Based on the first one hundred observations made in the period of study, one-step-ahead predictions of the number of hospitalizations, ICU admissions and deaths were performed from 19 August 2020 to 27 March 2022. These results are shown in Figure 3 (lower panel). A good forecasting performance is observed. When comparing actual observations with the predicted values, the mean squared prediction error was 0.028 for the number of hospital admissions, 0.019 for the number of ICU admissions and 0.001 for the number of deaths.

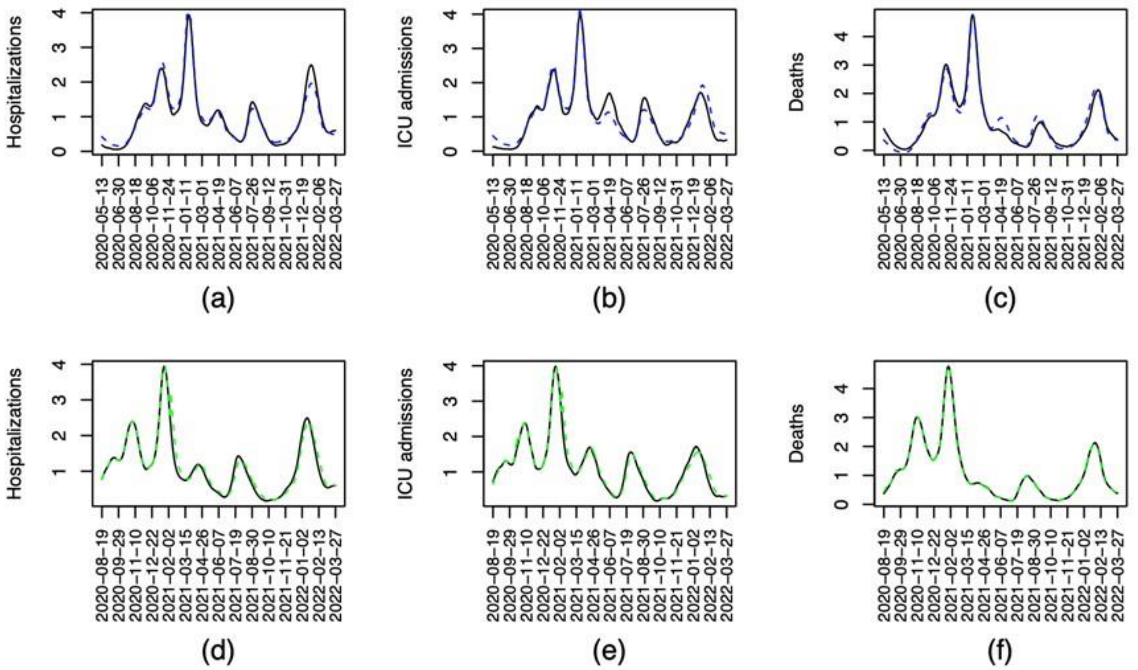

Figure 3. Reconstructed and predicted COVID-19 smoothed time series for Spain. Figures (a), (b) and (c) show the rescaled smooth original time series (black) and reconstructed time series (dashed blue) of COVID-19 for the period 13 May 2020 to 27 March 2022. Figures (d), (e) and (f) show the rescaled smooth original time series (black) and the predicted time series (dashed green) of COVID-19 for the period from 19 August 2020 to 27 March 2022.

Association between positive cases and underlying severity

The relationship between the number of positive cases and the underlying severity of COVID-19 indicator is analysed. We use a TVLM where the dependent variable $\boldsymbol{y}_t$ corresponds to the estimated underlying severity series $f_t(\widehat{\boldsymbol{a}})$, and the regressor $\boldsymbol{x}_t$ is the rescaled number of diagnosed cases at time t. As symptoms of severity usually appear later than the onset of the disease [4], we observe that the underlying severity time series reacts a few days later than the number of positive cases. For this reason, the correlations between underlying severity and the lagged series of positive cases are estimated with lags from 0 to 20. As a result, the highest correlation value is obtained with 8 lags (0.440). Hence, the explanatory variable included in the time-varying coefficient regression model is the eight-lagged number of positives, $\boldsymbol{y}_t = \boldsymbol{x}_{t-8}\beta_{t-8} + \boldsymbol{\varepsilon}_t,\ t = 9, \ldots, T$, for $\beta_{t-8} = \beta\left(\frac{t-8}{T-8}\right)$. The selected bandwidth in the framework of the kernel smoothing estimation of the TVLM is 0.25, so that the integer *k* corresponds to one quarter of the total observations. This window size is selected to prevent the particular phase of the pandemic wave from having an impact on the estimation of the time-varying coefficient.

The estimated model has a good explanatory capacity (pseudo-R^2=0.93). The estimated vector of coefficients β_t contains values from 0.22 (minimum) to 1.67 (maximum), with a mean value of 1.04 and a median of 1.37. Figure 4 shows this evolution over period of study. It can be observed that the relationship between the underlying severity $f_t(\widehat{\boldsymbol{a}})$ and the number of positive cases is quite stable until around mid-March 2021. Up to that point, the estimated coefficient is almost constant around 1.5, before it falls sharply. The drop in value of the coefficient and, therefore, in the expected underlying severity of COVID-19, continues until the beginning of October 2021. After this date, the value of the coefficient stabilizes at around 0.22 until the end of the period.

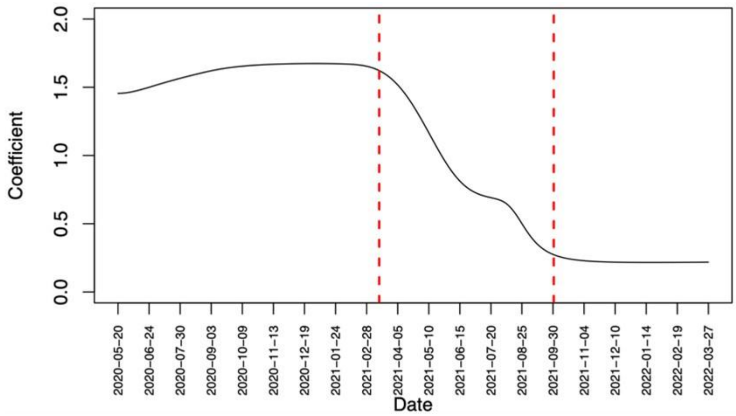

Figure 4. Estimated time-varying coefficient of the regression model in which the underlying COVID-19 severity is regressed by the rescaled series of positive cases. Red dashed lines are set at 15 March 2021 and 1 October 2021.

To conclude, the underlying severity indicator and its relationship with the number of positives is estimated for the following age groups: 20–49, 50–69, 70–79 and 80 years or more (80+). To obtain comparable results, COVID-19 time series by age group were rescaled by the expected value of the time series for the whole population. Figure 5 shows the estimated coefficients of the TVLM between underlying severity and the number of positive cases in each age group. First, time varying coefficients seem clearly associated with age. Note that the estimated coefficients are higher for the older age groups at any point in time, particularly for the 80+. In addition, all age group coefficients present the same 'constant-drop-constant' pattern, albeit at different moments in time: the younger the age group, the later the drop in the coefficient value begins. In Figure 5, the approximate date when the time-varying coefficient associated with each age group starts to decrease is indicated with a dashed vertical line – 15 March 2021 for the 80+ population (vaccination rate of 23.8%, brown dashed line); 7 May 2021 for 70-79 age interval (37.22%, yellow); 13 June 2021 for the 50–69 age interval (32.8%, green); and 17 July 2021 for the 20–49 age group (37.79%; blue).

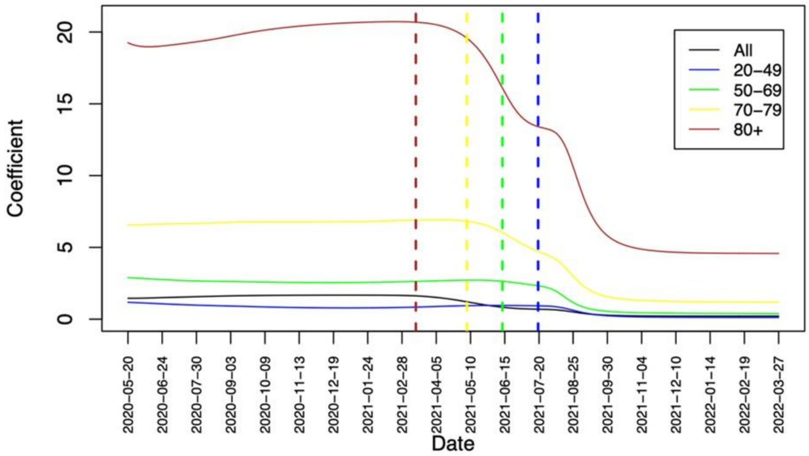

Figure 5. Estimated time-varying coefficients for age-group based regression models in which the underlying COVID-19 severity is regressed by the rescaled series of positive cases. Brown, yellow, green and blue dashed lines are set at March 15, 2021; May 7, 2021; June 13, 2021 and July 20, 2021, respectively.

DISCUSSION

Policy decisions taken by health authorities during the coronavirus pandemic were at least, in part, based on the evolution of COVID-19 indicators. Traditionally, these indicators have been analysed individually [5,10,13,16]; however, here, dynamic principal component techniques have been used to synthesize the information from a set of highly correlated indicators in an effort to monitor the evolution of COVID-19 with a single metric that can capture the underlying severity of the pandemic. We show that the information provided by the respective numbers of hospitalizations, ICU admissions and deaths is adequately represented by the first principal component and its first lag when one-sided dynamic principal components are applied [23].

Static dimensional reduction techniques are unable to capture the serial dependence of the COVID-19 indicators or such features as the pattern of deaths being sequential to hospitalizations and ICU admissions; thus, time series techniques are required for successful dimensionality reduction transform. Previous studies reporting the dimensional reduction of a set of pandemic indicators are scarce [3]. In this study, we show that one-sided dynamic principal components when used to reduce the dimensionality of a time series correctly capture serial dependence and the sequential nature of death. Moreover, because this technique only uses past values to obtain components, it can be used in forecasting. Any statistical technique applied to COVID-19 data should have sufficient explanatory power to understand transmission patterns and disease severity but also be adequate for forecasting, so as to help anticipate future epidemic outcomes [35,36]. Yet, some previous dynamic reduction techniques have been defined as functions of past and future time series values, making them appropriate for series reconstruction but not for forecasting purposes [37,38]. In this study, we have shown that one-sided dynamic dimensionality reduction performs correctly in terms of both its reconstruction accuracy and forecasting the evolution of COVID-19 indicators.

The number of cases diagnosed positive presented a weak linear correlation with the other three metrics of the severity of coronavirus analysed here. As shown, this can be attributed to the fact that their relationship was not constant over the period of observation. For this reason, we opted not to include this number of positives in the dimensional reduction analysis and we analysed separately the time-varying association between this metric and the underlying severity indicator to determine how this relationship changed over time. We found that the relationship between the number of positives and the underlying severity indicator was almost constant until March 2021, presenting, as such, a high linear correlation between these indicators during that period. However, as of March 2021, the relationship steadily decreased until October 2021, reflecting a decline in the consequences for the population with a positive diagnosis. From October 2021 onwards, the relationship between these metrics was once again constant over time, albeit at a much lower intensity. The period marked by a fall in the estimated association broadly coincided with that of the massive vaccination of the Spanish population [39], so massive vaccination seems to have reduced severity of reported cases. Many studies have highlighted the effectiveness of vaccination against the serious health consequences associated with coronavirus [40-41]. However, there are other factors that also contribute to covid disease outcomes, such as change of testing capacity and strategies, change of COVID-19 death criteria, improvement of treatments, natural immunity developed from previous infection, etc. In this study, we have developed a tool to monitor the underlying severity of detected positives which could be useful to assess the impact of vaccination and other factors on the severity evolution of detected cases.

Age is a well-known risk factor of serious illness/death after coronavirus infection. Indeed, older people, as well as those presenting with other medical conditions, were more likely to develop a serious illness or die [42]. In our study, we detected three relevant features associated with age. First, older age groups presented higher values of the underlying severity indicator for the same number of people diagnosed with COVID-19, with this ratio being especially high for the over eighties. Second, the same pattern – i.e. a decrease in the severity of COVID-19 depending on the number of positive cases after a period of stable association – was observed in all age groups; however, the older the group, the earlier the onset of the fall in the coefficients of the relationship between positives and underlying severity. This could be partially attributable to the fact that Spain's vaccination program was initiated among the oldest age groups, with vaccines being made progressively available to younger groups once a high percentage of older people had been vaccinated. Third, the reduction in underlying severity associated with positive diagnoses was more intense (in absolute numbers) with increasing age. However, if we analyse the reduction in relative terms, we find that this association fell by 84% among those aged 20 to 49, by almost 83% among those aged 50 to 69, by almost 80% among those aged 70 to 79 and by 75% in those aged over eighty, indicating that the relative reduction was lower among the older age groups [43,44].

The methodology employed in the current study to create a synthetic metric can be used in other fields of public health. Multiple interconnected metrics are frequently available in relation to public health issues and one of the difficulties health planners face is determining which provides the best description of the status of the health challenge. This study has shown how these alternative metrics can be unified while retaining most of their information, thus, providing policy makers with a single metric that describes the severity status of the health issue and enabling them to monitor the evolution of the disease. Analysis of this synthetic indicator may be useful for prioritizing the vaccination of high risk groups and evaluating the impact of vaccination in terms of reducing disease severity. This information could also be useful for identifying the optimum point in time to initiate vaccination of the next group at greatest risk.

We use publicly available Spanish data from the National Epidemiology Centre from 2020 to 2022 and propose the use of one-sided dynamic principal components to demonstrate that a single metric adequately explain the variability contained in three alternative indicators of the severity of COVID-19 frequently used to monitor the evolution of the pandemic: the numbers of daily hospitalisations, ICU admissions and deaths attributable to the disease. We show that this metric can be used as a composite indicator to monitor the underlying severity of detected positives and to assess the impact of vaccination and other factors on the fall in the severity risk indicator by age groups. The study is not free of limitations. In constructing the composite index of underlying severity, it would be useful for policy-making in pandemic settings if the single metric could include more information. The selection of only three time-series in the construction of the single metric was due to the availability of reliable information of the Spanish surveillance system. In addition, the same importance is attached to all three time-series; yet, hospitalizations, ICU admissions and deaths reflect different degrees of severity of coronavirus. This limitation could be addressed using time-series dimensional reduction techniques that weight the different degrees of severity of the three indicators. However, the underlying severity index is certainly useful for analysing the evolution of the indicators during the period of observation, albeit that its value is not easily and directly interpretable, given that the reconstruction of the coronavirus time series involves the underlying severity indicator and its one period lagged value.

Significance for public health

The methodology employed in the current study to create a synthetic metric can be used in other fields of public health. Multiple interconnected metrics are frequently available in relation to public health issues and one of the difficulties health planners face is determining which provides the best description of the status of the health challenge. This study has shown how these alternative metrics can be unified while retaining most of their information, thus, providing policy makers with a single metric that describes the severity status of the health issue and enabling them to monitor the evolution of the disease. Analysis of this synthetic indicator may be useful for prioritizing the vaccination of high risk groups and evaluating the impact of vaccination in terms of reducing disease severity. This information could also be useful for identifying the optimum point in time to initiate vaccination of the next group at greatest risk.

References

1. Peiffer-Smadja N, Lucet J-C, Bendjelloul G, Bouadma L, Gerard S, Choquet C, et al. Challenges and issues about organizing a hospital to respond to the COVID-19 outbreak: experience from a French reference centre. Clin Microbiol Infect. 2020;26(6):669-672. https://doi.org/10.1016/j.cmi.2020.04.002.

2. Lee A, English P, Pankhania B, Morling JR. Where England's pandemic response to COVID-19 went wrong. Public health. 2021;192:45–48. https://doi.org/10.1016/j.puhe.2020.11.015.

3. Swallow B, Xiang W, Panovska-Griffiths J. Tracking the national and regional COVID-19 epidemic status in the UK using weighted principal component analysis. Phil Trans R Soc A. 2022;380(2233). http://doi.org/10.1098/rsta.2021.0302.

4. Santolino M, Alcañiz M, Bolancé C. Hospitalizations from covid-19: a health planning tool. Rev Saude Publica. 2022;56(51). https://doi.org/10.11606/s1518-8787.2022056004315.

5. Meakin S, Abbott S, Bosse N, Munday J, Gruson H, Hellewell J, et al. Comparative assessment of methods for short-term forecasts of COVID-19 hospital admissions in England at the local level. BMC Med. 2022;20(86). https://doi.org/10.1186/s12916-022-02271-x.

6. Vasileiou E, Simpson CR, Shi T, Kerr S, Agrawal U, Akbari A, et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. Lancet. 2021;397(10285):1646-1657. https://doi.org/10.1016/S0140-6736(21)00677-2.

7. Nguyen HM, Turk PJ, McWilliams AD. Forecasting COVID-19 Hospital Census: A Multivariate Time-Series Model Based on Local Infection Incidence. JMIR Public Health Surveill. 2021;7(8):e28195. https://doi.org/10.2196/28195.

8. Leclerc QJ, Fuller NM, Keogh RH, Diaz-Ordaz K, Sekula R, Semple MG, et al. Importance of patient bed pathways and length of stay differences in predicting COVID-19 hospital bed occupancy in England. BMC Health Serv Res. 2021;21(566). https://doi.org/10.1186/s12913-021-06509-x.

9. Chang R, Elhusseiny KM, Yeh YC, Sun WZ. COVID-19 ICU and mechanical ventilation patient characteristics and outcomes—A systematic review and meta-analysis. PLOS ONE. 2021;16(2): e0246318. https://doi.org/10.1371/journal.pone.0246318.

10. Jain V, Yuan JM. Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. Int J Public Health. 2020;65:533–546. https://doi.org/10.1007/s00038-020-01390-7.

11. Karagiannidis C, Windisch W, McAuley DF, Welte T, Busse R. Major differences in ICU admissions during the first and second COVID-19 wave in Germany. Lancet Respir Med. 2021;9(5):e47-e48. https://doi.org/10.1016/S2213-2600(21)00101-6.

12. Zhao C, Tepekule B, Criscuolo NG, Wendel Garcia PD, Hilty MP, Fumeaux T, et al. icumonitoring.ch: a platform for short-term forecasting of intensive care unit occupancy during the COVID-19 epidemic in Switzerland. Swiss Med Wkly. 2020;150:w20277. https://doi.org/10.4414/smw.2020.20277.

13. Hametner C, Böhler L, Kozek M, Bartlechner J, Ecker O, Du ZP, et al. Intensive care unit occupancy predictions in the COVID-19 pandemic based on age-structured modelling and differential flatness. Nonlinear Dyn. 2022;109:57–75. https://doi.org/10.1007/s11071-022-07267-z.

14. Runge M, Richardson RAK, Clay PA, Bell A, Holden TM, Singam M, et al. Modeling robust COVID-19 intensive care unit occupancy thresholds for imposing mitigation to prevent exceeding capacities. PLOS Glob Public Health. 2022;2(5):e0000308. https://doi.org/10.1371/journal.pgph.0000308.

15. Yanez ND, Weiss NS, Romand JA, Treggiari MM. COVID-19 mortality risk for older men and women. BMC Public Health. 2020;20(1742). https://doi.org/10.1186/s12889-020-09826-8.

16. Meslé MM, Brown J, Mook P, Hagan J, Patore R, Bundle N et al. Estimated number of deaths directly averted in people 60 years and older as a result of COVID-19 vaccination in the WHO

European Region, December 2020 to November 2021. Euro Surveill. 2021;26(47). https://doi.org/10.2807/1560-7917.ES.2021.26.47.2101021.

17. Weinberger DM, Chen J, Cohen T, Crawford FW, Mostashari F, Olson D et al. Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020. JAMA Intern Med. 2020;180(10):1336–1344. https://doi.org/10.1001/jamainternmed.2020.3391.

18. Duarte P, Riveros-Perez E. Understanding the cycles of COVID-19 incidence: principal component analysis and interaction of biological and socio-economic factors. Ann Med Surg. 2021;66(102437). https://doi.org/10.1016/j.amsu.2021.102437.

19. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi A, Band SS. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. Alex Eng J. 2021;60(1):457-464. https://doi.org/10.1016/j.aej.2020.09.013.

20. Nobi A, Tuhin KH, Lee JW. Application of principal component analysis on temporal evolution of COVID-19. PLOS ONE. 2021;16(12):e0260899. https://doi.org/10.1371/journal.pone.0260899.

21. Carroll C, Bhattacharjee S, Chen Y, Dubey P, Fan J, Gajardo A, et al. Time dynamics of COVID-19. Sci Rep. 2020;10(21040). https://doi.org/10.1038/s41598-020-77709-4.

22. Tang C, Wang T, Zhang P. Functional data analysis: An application to COVID-19 data in the United States. arXiv:2009.08363 (Preprint). 2020 (cited 2022 Nov). https://doi.org/10.48550/ARXIV.2009.08363.

23. Peña D, Smucler E, Yohai VJ. Forecasting Multiple Time Series With One-Sided Dynamic Principal Components. J Am Stat Assoc. 2019;114(528):1683-1694. http://dx.doi.org/10.1080/01621459.2018.1520117.

24. Richman, M. B. (1986). Rotation of principal components. Journal of Climatology, 6(3), 293-335

25. Compagnucci RH, Vargas WM. 1986. Patterns of surface pressure field during July 1972–1983 in southern South America and the Antarctic peninsula. In Proceedings of the Third International Conference on Statistical Climatology, Cehak K (ed.). Australian Society of Meteorology: Vienna; 1–14.

26. Cleveland R, Cleveland W, McRae J, Terpenning I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. J Off Stat. 1990;6:3-73.

27. Nadaraya, EA. On Estimating Regression. Teoriya Veroyatnostei i Ee Primeneniya. 1964;9(1): 157–159. https://doi.org/10.1137/1109020.

28. Watson, GS. Smooth Regression Analysis. Sankhyā, Series A. 1964;26(4): 359–372. https://www.jstor.org/stable/25049340.

29. Said SE, Dickey DA. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika. 1984;71(3):599-607. https://doi.org/10.2307/2336570.

30. Casas I, Fernandez-Casal R. tvReg: Time-varying Coefficient Linear Regression for Single and Multi-Equations in R. R J. 2019;14(1): 79-100. https://dx.doi.org/10.2139/ssrn.3363526.

31. Fan J, Gijbels I (1996). Local Polynomial Modeling and Its Applications. Chapman and Hall, London.

32. Fan J, Zhang W. Statistical methods with varying coefficient models. Stat Interface. 2008;1: 179–195. https://doi.org/10.4310/SII.2008.V1.N1.A15.

33. Peña D, Smucler E, Yohai VJ. odpc: One-Sided Dynamic Principal Components. R package version 2.0.5. 2022. Available from: http://cran.r-project.org/web/packages.

34. Peiris M, Perera B. On prediction with fractionally differenced ARIMA models. J Time Ser Anal. 1988;9(3):215-220. https://doi.org/10.1111/j.1467-9892.1988.tb00465.x.

35. James LP, Salomon JA, Buckee CO, Menzies NA. The use and misuse of mathematical modeling for infectious disease policymaking: lessons for the COVID-19 pandemic. Medical Decision Making. 2021; 41: 379-385

36. Nixon K, Jindal S, Parker F, Marshall M, Reich NG, Ghobadi K et al. Real-time COVID-19 forecasting: challenges and opportunities of model performance and translation. 2022, The Lancet, 4(10), E699-E701, October, DOI: https://doi.org/10.1016/S2589-7500(22)00167-4

37. Brillinger DR. Time Series: Data Analysis and Theory. San Francisco: Holden-Day; 1981.

38. Peña D, Yohai VJ. Generalized Dynamic Principal Components. J Am Stat Assoc. 2016; 111: 1121–1131.

39. ECDC. COVID-19 Vaccine Tracker. European Centre for Disease Prevention and Control, https://opendata.ecdc.europa.eu/ (accessed 1 Dec 2022).

40. Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. Lancet. 2021;397(10287): 1819-1829. https://doi.org/10.1016/S0140-6736(21)00947-8.

41. Thompson MG, Stenehjem E, Grannis S, Ball SW, Naleway AL, Ong TC, et al. Effectiveness of Covid-19 vaccines in ambulatory and inpatient care settings. N Engl J Med. 2021;385(15): 1355-1371. https://doi.org/10.1056/NEJMoa2110362.

42. WHO. Overview of coronavirus disease (COVID-19). World Health Organization, 2022. https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed on Nov 11, 2022)

43. Grannis SJ, Rowley EA, Ong TC, Stenehjem E, Klein NP, DeSilva MB, et al. Interim Estimates of COVID-19 Vaccine Effectiveness Against COVID-19-Associated Emergency Department or Urgent Care Clinic Encounters and Hospitalizations Among Adults During SARS-CoV-2 B.1.617.2 (Delta) Variant Predominance - Nine States, June-August 2021. MMWR Morb Mortal Wkly Rep. 2021;70(37):1291-1293. https://doi.org/10.15585/mmwr.mm7037e2.

44. Lin DY, Gu Y, Wheeler B, Young H, Holloway S, Sunny SK, et al. Effectiveness of Covid-19 vaccines over a 9-month period in North Carolina. N Engl J Med. 2022;386(10):933-941.https://doi.org/10.1056/NEJMoa2117128.

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu