**////Title: Using Functional Data Analysis to Better Understand COVID-19 Mortality**

**////Stand-first**:

Italy was the first European country to experience a major outbreak of COVID-19. Across the country, intriguingly different patterns of mortality were seen across regions. Dr. Marzia Cremona from Université Laval, in collaboration with a global team of researchers from the Pennsylvania State University and the Sant’Anna School of Advanced Studies, used novel statistical tools from the field of functional data analysis (FDA) to identify several important factors contributing to these differences. This work has critical implications for policymaking and also demonstrates the useful application of FDA techniques to epidemiological research.

**////Body text:**

A novel coronavirus (now known as COVID-19) was identified from a cluster of atypical pneumonia cases in Wuhan in late 2019. This virus has gone on to sweep the globe, leaving almost no country untouched.

COVID-19 infection and death rates have varied massively from country to country, depending on factors such as border controls, testing strategies, and, later on, vaccine availability. As well as differences between countries, there have been marked differences between regions of the same country. This was the case in Italy, which was the first European country to have a major outbreak of COVID-19. The first epidemic wave of COVID-19 unfolded very differently in different parts of the country between February and May 2020.

Dr. Marzia Cremona from Université Laval in Canada, a specialist in the development of statistical techniques for the analysis of complex data, in collaboration with a global team of researchers from the Pennsylvania State University in the USA and the Sant’Anna School of Advanced Studies in Italy, used new statistical methods to understand the extreme differences observed among the twenty different regions in Italy during the first wave of the epidemic, when vaccines were not yet available. In addition to Dr. Cremona, the team included Prof. Francesca Chiaromonte, Dr. Tobia Boschi, Dr. Jacopo Di Iorio and Mr. Lorenzo Testa.

…

Mathematical and epidemiological models have been used throughout the pandemic to understand what has happened and why, as well as to try to anticipate what might happen next. Dr. Cremona and her collaborators used a different approach, exploiting tools from functional data analysis (FDA), a field of statistics that studies observations consisting of curves – such as mortality curves plotting death rates over time – to extract patterns and associations from the data. This approach can complement traditional epidemiological analyses and also provide new answers to critical questions. The techniques used by the team included previously tested methods along with brand new methods specifically developed by the group.

The researchers used FDA to compare mortality curves of the first epidemic wave across the twenty regions of Italy and to understand whether mortality could be partially explained by mobility and positivity (that is, the proportion of performed COVID-19 tests returning positive results). They also investigated the role of socio-demographic, infrastructural, and environmental factors on mortality curves.

…

Due to significant underreporting of official COVID-19 deaths in Italy, the researchers had to determine more accurate death counts. They did this by computing, for each region, the difference between the mortality in 2020 and the average mortality over the last five years.

FDA techniques were used to smooth, align and cluster mortality curves, to account for differences in the timing of outbreaks, and to find groups of regions behaving similarly. Within the country there appeared to be two different epidemics; a relatively mild one in the majority of the regions and a much harsher one detected in a few specific areas.

For example, Veneto and Lombardia are geographically close regions with similar population structures and were two of the first regions in Italy to report confirmed COVID-19 cases. However, the outbreaks took strikingly different paths with Lombardia being hit much harder and Veneto undergoing a milder epidemic, probably thanks to the implementation of an aggressive testing approach early on.

…

Local mobility refers to how much people move around their local area, for example, going to the shops. For mobility information, the team used Google’s ‘grocery and pharmacy’ data which shows the percentage change in mobility – compared to a baseline chosen before the epidemic outbreak – of people leaving their homes for essentials such as food and medicine.

The data showed that mobility dropped by 30% in the first week after the lockdown was announced, then continued to fall by as much as 60% on weekdays and 100% on weekends. The team found that mobility was strongly associated with COVID-19 mortality and that local mobility levels early on in the pandemic could help predict peak mortality, hence suggesting that staying at home really does help to reduce the number of COVID-19 deaths.

…

To build positivity curves relating to COVID-19, the team used FDA to smooth out noise caused by testing and reporting delays from publicly available test and case number data. These curves showed that 10% of tests in Lombardia were positive by 22nd February 2020, with positivity increasing until it peaked in mid-March. Although not as strongly as mobility, early positivity was also associated with mortality. Interestingly, the data also showed that high positivity later on in the pandemic was also associated with peak mortality, probably due to the hard-hit areas taking longer to recover.

…

Dr. Cremona and her collaborators then investigated several other socio-demographic, infrastructural, and environmental factors which may have affected mortality rates. These were: the ageing of the population, the prevalence of pre-existing conditions, the quantity of primary and hospital-based health care, and the potential for places such as homes, schools, public transport, and hospitals to provide a place for transmission.

A technique known as ‘Cheng and Church’s biclustering algorithm’ was employed to find subsets of regions with similar socio-demographic, infrastructural and environmental variables, and to study their interdependencies. Among the resulting biclusters, one included central and southern regions where the first wave of the epidemic was milder. Those regions were characterised by limited concentrations in hospitals, nursing homes, and workplaces, as well as many people with pre-existing conditions and a small number of Intensive Care Unit places. Another group consisted of some of the northern regions which were worst hit by the pandemic such as Lombardia, as well as some regions with milder epidemics such as Veneto. This group was characterised by larger hospitals and more crowded workplaces and schools, as well as an older population.

The researchers found that the strongest predictor for mortality was having a low number of primary care doctors per patient. The number of intensive care unit beds (a marker for large hospitals) and the size of hospital wards, classrooms, and workplaces also had a positive association with COVID-19 deaths. All these variables capture the effects of contagion hubs, i.e. of places where contacts among people can favour COVID-19 spread. Importantly, even accounting for these factors, mobility, and positivity still acted as strong predictors of mortality.

…

Using FDA techniques, Dr. Cremona and her collaborators demonstrated that local mobility and positivity rates are strongly associated with COVID-19 mortality, and that well-distributed primary care can mitigate mortality.

Whilst these findings have important implications for public health and policymaking, they don’t completely explain why some regions were much harder hit than others. To improve the model, more accurate data on cases, deaths and hospitalisations are needed as well as data being available at a more local scale than just that of the twenty Italian regions. Dr. Cremona notes that this highlights the importance of collecting and sharing high-quality data in determining the response to key public health questions.

Dr. Cremona wishes to thank the Office of the Vice Dean of Research and Innovation at the Faculty of Business Administration of Université Laval for providing funding support for this SciPod.

This SciPod is a summary of the paper ‘Functional data analysis characterizes the shapes of the first COVID-19 epidemic wave in Italy’, published in Scientific Reports. DOI: <https://doi.org/10.1038/s41598-021-95866-y>

For further information, you can connect with Marzia Cremona at marzia.cremona@fsa.ulaval.ca