

Modeling the COVID-19 transmission dynamics and identifying meteorological and sociodemographic predictors

Rodic A.^{1*}, Salom I.², Milicevic O.³, Markovic S.¹, Zigic D.², Ilic B.², Djordjevic M.², Djordjevic M.¹

¹Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia

²Institute of Physics, University of Belgrade, Belgrade, Serbia

³Department for Medical Statistics and Informatics, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

*andjela.rodic@bio.bg.ac.rs

Key words: COVID-19, epidemiological model, basic reproduction number, SARS-CoV-2 transmissibility predictors, machine learning, feature selection and regularization, Principal Component Analysis, mitigation measures

Motivation and Aim: The spread of the SARS-CoV-2 virus during the first wave of the COVID-19 epidemic depended on its inherent transmissibility in a given population and the effectiveness of practicing the social distancing measures. To analyze the conditions favoring the fast infection spread in a population, one first needs to choose an appropriate measure of that spread. The Basic Reproduction Number (R_0) measures the inherent virus transmissibility in a fully susceptible population in absence of any interventions. In addition, we argue that it is highly independent from the testing capacity. Then, one needs to identify the direct predictors of R_0 among potentially relevant meteorological and sociodemographic factors. However, this is challenging due to high mutual correlations between some of the factors, which can therefore be obtained as indirect predictors. The social distancing measures can significantly impact the rate of the infection spread, but they have been introduced in different populations with different efficiencies. In China, the numbers of cases and deaths reached in Wuhan were orders of magnitude higher than the total numbers reported for the rest of the country. Therefore, one may ask if this prominent difference can be explained by an interplay of the inherent virus transmissibility and the effects of the applied control measures.

Methods and Algorithms: We constructed a dynamic compartmental model describing the infection propagation upon the virus introduction in a population. For 118 countries, we estimated the values of R_0 from the fit of the model to the initial exponential phase of the case growth curves, characterized by the lack of intervention measures. First, we determined bivariate Pearson correlations of R_0 with different meteorological and sociodemographic variables [1]. Then, we performed an advanced statistical analysis to infer the direct R_0 predictors [2]. Namely, we applied the Principal Component Analysis to the meteorological and demographic variable sets to exchange them for the corresponding sets of non-correlated Principal Components (PCs). Next, we searched for robust R_0 predictors among the PCs using independently four different methods: the custom multiple regression analysis, the Stepwise regression, the Lasso, and the Elastic net. The last two methods utilize both regularization and variable selection. To interpret the resulting predictor PCs, we analyzed their correlations with the starting variables. Regarding the epidemic growth phase influenced by social distancing measures, we enabled the model to describe it by incorporating the gradual flow of susceptible, healthy individuals practicing the measures to a “protected” compartment unavailable to the infection [3]. We used a constrained numerical

procedure to fit the full model to the cumulative case counts data for each of the China provinces and estimated the corresponding virus transmission (R_0) and social distancing effectiveness parameters.

Results: The results of the advanced statistical analysis indicate that many of the strong correlations with R_0 , obtained by the bivariate analysis [1] are probably the indirect consequences of the mutual variable correlations [2]. All four methods (Fig. 1, A–D) singled out the demographic PCs 1, 7 and 9 as the most direct predictors of R_0 , which are dominantly correlated with the country development indicators, the built-up area per capita, and the time from the global epidemic onset respectively. Three methods (see Fig. 1, A–C) indicate the importance of an unhealthy lifestyle and environment through the PC4, while two methods (see Fig. 1, C–D) acknowledge the independent influence of the weather seasonality through the meteo PC1, despite its high correlation with the country development level. Among the China provinces, Wuhan appears as an outlier, with relatively high inherent virus transmission combined with less efficient epidemic control measures [3]. While it shows the highest Case Fatality Ratio, its Infection Fatality Ratio is average. The results also indicate that provinces tightened the measures with the increase of R_0 .

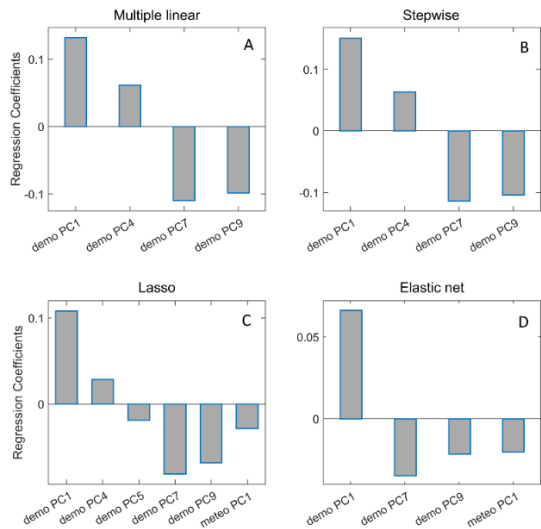


Fig. 1. Inferring direct R_0 predictors using four independent methods (A–D). Adapted from [2]

Conclusion: Among different meteorological and sociodemographic factors, which can be correlated both mutually and with R_0 , the prosperity level, indoor crowdedness, the epidemic onset, and unhealthy lifestyle appear as the most robust, direct predictors of the SARS-CoV-2 transmissibility. The interplay of the inherent virus transmissibility and the effectiveness of the introduced intervention measures can reasonably explain very different epidemic outcomes in the provinces of China.

Acknowledgements: This study is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

1. Salom I. et al. Effects of Demographic and Weather Parameters on COVID-19 Basic Reproduction Number. *Front Ecol Evol.* 2021;8:617841.
2. Djordjevic M. et al. Inferring the Main Drivers of SARS-CoV-2 Global Transmissibility by Feature Selection Methods. *GeoHealth* 2021;5(9):e2021GH000432.
3. Djordjevic M. et al. A systems biology approach to COVID-19 progression in population. *Adv Protein Chem Struct Biol.* 2021;127:291-314.