Letter to the Editor

A new stratification model for a population health risk assessment, based on a large cohort of patients infected by COVID-19

Dear Editor,

The coronavirus disease 2019 (COVID-19) pandemic dramatically challenged health care systems worldwide in many ways, from the need for accurate and accessible diagnostics for the disease surveillance, to the necessity of managing the growing demand for hospital care [1]. In the USA of an estimated 20% of individuals that gets infected a median of 11.2 million hospitalizations, 2.7 million Intensive Care Unit admissions, 1.6 million patients requiring a ventilator, 62.3 million hospital bed days, and $163.4 billion in direct medical costs have been calculated [2].

Despite the recent start of the vaccination programs across several countries COVID-19 will likely continue to be a global health challenge with a growing economic impact if no effective strategic control measures are implemented.

In recent years, there has been a growing interest in using a huge variety of real-world data (RWD) obtained during routine clinical care, with the aim to provide valuable rapid information to better understand the natural history and course of disease and answer urgent questions on effectiveness and safety of treatments and health economics [3]. COVID-19 pandemic has expanded the potential of RWD to guide clinical trial development, find answers to urgent clinical questions, and support regulatory decisions.

RWD-based studies focus on several aspects, including the feasibility of rapid diagnostic tests in the management of the COVID-19 outbreak, the mortality following the infection, and clinical efficacy of pharmacological interventions [4]. However, only a few studies provided a comprehensive risk score on large enough populations to predict the development of critical illness or death among COVID-19-infected patients; this strategy is of crucial importance to properly drive the diagnostic therapeutic assistance pathway of interventions in clinical practice, also during other epidemics.

The epidemic spread to and increased exponentially in Italy earlier than in any other western countries, with over 3.08 million confirmed cases of SARS-CoV-2 infections as of March 8. In order to closely monitor the health impact of pandemic diffusion and to implement targeted preventive, therapeutic or rehabilitative interventions, it has become necessary for the Regional Health Authority of Lombardy to identify a statistical model based on RWD. The model has been designed and implemented by the Regional Epidemiologic Observatory of the Welfare General Directorate, in collaboration with the Epidemiologic Observatories of the eight Local Health Authorities that insist in the territory of Lombardy (each of which is a separate geographically based public company delivering public health services).

The model (called "MIP" i.e., Predictive Indexing Model) has been designed to allow a population stratification based on the risk to experience severe adverse outcomes during COVID-19 infection, including severe illness (requiring hospital or ICU admission) and death, by taking into account demographic, clinical, chronic e comorbid individual conditions.

The model included all symptomatic patients with confirmed SARS-CoV-2 infection, extracted from the regional COVID-19 integrated surveillance database (DB COVID-19) between 21 February 2020 and 30 November 2020. Patients who died within one week of ascertaining positivity (in which the adverse outcome could be related to the severity of pre-existing clinical conditions) were excluded.

The model was integrated with the Lombardy Region's administrative databases, containing information on 10 million residents (16% of the Italian population) who receive National Health System (NHS) assistance: drug prescriptions, hospital discharge records, payment exemptions for pathology or income, outpatient visits.

In order to identify chronic patients, the information thus obtained was further merged in a Regional unique database able to provide, through appropriate algorithms, the chronic disease and its severity level.

The population was divided into the following three age groups: (1) young people: 19–44 years old, (2) middle-aged people: 45–64 years old, and (3) elderly people: 65–79 years old.

On the basis of current Lombardy surveillance data on COVID-19 patients, different outcomes for each age group were assessed trough: (i) a composite outcome made up of admission to any department (including ICU) or death for group 1; (ii) a composite outcome made up of admission to ICU or death for group 2, and (iii) death, for group 3.0.

We included 307711 patients (50.4% women) with a mean (±SD) age of 48.7 ± 15.6 years. The distribution into different age groups was: 118150 patients in group 1 (38.4%), 135925 in group 2 (44.2%), 53636 in group 3 (17.4%). Adverse events occurring in younger group were 3334 (in 2.82% of these patients), mostly hospitalizations, in the middle group were 2935 (2.16%), mostly ICU admissions, and in the elderly group were 6333 deaths (11.81%).

The statistical approach involved the study of three different logistic regression models with specific variable selection methods: Backward [5], Stepwise, LASSO [6]. The variables included were gender, age (continuous), presence or absence of a chronic disease (selected among a list of 37 diseases), and number of chronic comorbidities.

Using the same predictors identified in all the three analyses, a final multivariate logistic regression model has been performed, with an excellent predictive performance (area under the curve – AUC – of 0.71 for group 3, of 0.76 for group 2, 0.64 for group 1). Significant differences among age groups were demonstrated for adverse outcomes and the variables identified as independent predictors of different outcomes were (Odds Ratio; [95% Confidence Interval]):

19–44 years: for the risk of hospitalization, the predictors (in

descending order) that contribute to defining risk levels were *Chronic kidney disease, Dialysis* (3.265; [1.773–6.012]), *Cancer* (2.127; [1.568–2.885]), *Number of comorbidities* (1.506; [1.429–1.586]), *Gender* (1.124; [1.049–1.205]), *Age* (1.057; [1.052–1.062]);.

45–64 years: for the risk of hospitalization and ICU admission, the predictors were: *Cancer* (3.386; [2.903–3.950]), *Respiratory disease* (3.01; [1.838–4.931]), *Chronic kidney disease* (2.033; [1.616–2.558]), *Epilepsy* (1.815; [1.351–2.438]), *Chronic kidney disease, Dialysis* (1.781; [1.251–2.535]), *Diabetes* (1.652; [1.467–1.861]), *Heart failure* (1.428; [1.146–1.780]), *Gender* (3.411; [3.123–3.726]), *Number of comorbidities* (1.253; [1.202–1.307]), *Age* (1.098; [1.090–1.106]);.

65–79 years: for the risk of death, the predictors were: *Cancer* (1.932; [1.762–2.118]), *Heart failure* (1.696; [1.544–1.863]), *Parkinson's disease and parkinsonism* (1.637; [1.344–1.995]), *HIV* (1.617; [1.138–2.297]), Chronic *kidney disease* (1.474; [1.305–1.665]), *Chronic kidney disease, Dialysis* (1.392; [1.114–1.738]), *Cardio-cerebrovascular diseases* (1.268; [1.179–1.364]), *Diabetes* (1.223; [1.14–1.311]), *Gender* (2.049; [1.93–2.176]), *Number of comorbidities* (1.113; [1.081–1.145]), *Age* (1.104; [1.096–1.111]).

Using the logistic regression equation, the clinical risk has been calculated for each Lombardy resident aged between 19 and 79 years (7,633,829 people). For each age group, a separate cluster analysis has been performed, obtaining 7 population clusters with an increasing level of risk.

The higher risk (level 9) and the lowest risk (level 1) were automatically attributed to people aged ≥80 years and to people aged ≤18 years, respectively, without applying any specific model but considering the different and specific pandemic diffusion and related outcomes in these age groups.

Descriptive analysis was used to compare the ability of the model to risk stratify patient with the rate of adverse outcomes (with a progressive severity level: hospitalization, ICU use, mortality) occurred in the total Lombardy COVID-19 infected patients.

Results are coherent with the risk order given by the stratification model and the rate of more severe outcomes increases within the clusters at major risk level: e.g., in the lowest risk class there are 0 death, 1 ICU admission, 23 total hospitalizations per 1000 COVID-19 positives; on the contrary, in the highest risk class are distributed 247 deaths, 34 ICU admissions and 424 total hospitalizations per 1000 COVID-19 positives. The tool is able to triage 95% of mortality rate to the fourth higher-risk clusters. Based on these results, we were able to confirm our algorithmic approach.

The risk level related to COVID-19 infection (from 1 to 9) obtained by the model is specific for each age group and for the combination of demographic, clinical and chronic characteristics emerging as predictors in the statistical analysis. Our study confirms that chronic conditions tend to form risk clusters, in addition to increasing age.

Our model, resulting from the intersection of three different logistic regression methods, had an excellent predictive performance (in terms of AUC values) indicating that this risk-stratification system can be used for clinical predictivity purposes. The random split into the linked datasets (administrative and chronic disease databases, DB COVID-19) of a whole regional population of approximately 10 million people allowed a reliable methodology for the validation.

During its use a monitoring process will be performed by the Regional Epidemiologic Observatory with a periodical analysis of the model used at a population level. Furthermore, our findings supporting the role of some chronic and coexisting diseases as risk factors for clinical severity and prognostic evolution of COVID-19 infection, is consistent with current available literature [7].

In general, this risk-stratification model can be a useful tool for policy-makers (e.g., regional and territorial decision-makers for public health program planning) and health care providers (e.g., general practitioners) to identify groups of patients with severe conditions at high-risk for adverse outcomes and with similar care needs requiring target preventive, therapeutic or rehabilitative interventions.

This clusters approach could also be useful to prioritize the introduction of new care pathways associated to the different levels of risk and the revision of multimorbidity guidelines for the most common diseases and their combinations.

References

[1] World Health Organization, Coronavirus disease 2019 (COVID-19): situation report – 51. World HealthOrganization, Geneva, 11 March, 2020. www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57 (Accessed 8 March 2021).

[2] S.M. Bartsch, M.C. Ferguson, J.A. McKinnell, K.J. O'Shea, P.T. Wedlock, S. S. Siegmund, B.Y. Lee, The Potential health care costs and resource use associated with COVID-19 in The United States, Health Aff. 39 (6) (2020) 927–935.

[3] US Food and Drug Administration, Coronavirus (COVID-19) Update: FDA Collaborations Promote Rigorous Analyses of Real-World Data to Inform Pandemic Response. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-collaborations-promote-rigorous-analyses-real-world-data-inform (Accessed 8 March 2021).

[4] X. Li, Y. Yang, L. Liu, X. Yang, X. Zhao, Y. Li, Y. Ge, Y. Shi, P. Lv, J. Zhang, T. Bai, H. Zhou, P. Luo, S. Huang, Effect of combination antiviral therapy on hematological profiles in 151 adults hospitalized with severe coronavirus disease 2019, Pharmacol. Res. 160 (2020), 105036.

[5] D. Dunkler, M. Plischke, K. Leffondre, G. Heinze, Augmented backward elimination: a pragmatic and purposeful way to develop statistical models, PLoS One 9 (11) (2014), e113677.

[6] S. Walter, H. Tiemeier, Variable selection: current practice in epidemiological studies, Eur. J. Epidemiol. 24 (12) (2009) 733–736.

[7] X. Fang, S. Li, H. Yu, P. Wang, Y. Zhang, Z. Chen, Y. Li, L. Cheng, W. Li, H. Jia, X. Ma, Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis, Aging 12 (13) (2020) 12493–12503.

Roberto Blaco[a], Olivia Leoni[a], Gjiliola Cukaj[a,b], Michele Ercolanoni[a,b], Carla Carnovale[c,*]

[a] *Regional Epidemiologic Observatory, Welfare General Directorate, Lombardy Region, Milan, Italy*

[b] *Epidemiologic Observatory, Digital Information HUB, ARIA S.p.A., Milan, Italy*

[c] *Unit of Clinical Pharmacology, Department of Biomedical and Clinical Sciences L. Sacco, "Luigi Sacco" University Hospital, Università di Milano, Milan, Italy*

[*] Corresponding author.
E-mail address: carla.carnovale@unimi.it (C. Carnovale).