Stratification of the risk of developing severe or lethal Covid-19 using a new score from a large Italian population: a population-based cohort study

Giovanni Corrao 1,2, Federico Rea 1,2, Flavia Carle 2,3, Salvatore Scondotto 2,4, Alessandra Allotta 4, Vito Lepore 5, Antonio D’Ettorre 5, Cinzia Tanzarella 5, Patrizia Vittori 6, Sabrina Abena 6, Marica Iommi 3, Liana Spazzafumo 2,7, Michele Ercolanoni 5, Roberto Blaco 8, Simona Carbone 5, Cristina Giordani 9, Dario Mantellotto 10, Massimo Galli 11, Giuseppe Mancia 12,13

On behalf of the 'Monitoring and Assessing care Pathways (MAP)’ working group of the Italian Ministry of Health

ABSTRACT

Objectives To develop a population-based risk stratification model (COVID-19 Vulnerability Score) for predicting severe/fatal clinical manifestations of SARS-CoV-2 infection, using the multiple source information provided by the healthcare utilisation databases of the Italian National Health Service.

Design Retrospective observational cohort study.

Setting Population-based study using the healthcare utilisation database from five Italian regions.

Participants Beneficiaries of the National Health Service, aged 18–79 years, who had the residentship in the five participating regions. Residents in a nursing home were not included. The model was built from the 7 655 502 residents of Lombardy region.

Main outcome measure The score included gender, age and 29 conditions/diseases selected from a list of 61 conditions which independently predicted the primary outcome, that is, severe (intensive care unit admission) or fatal manifestation of COVID-19 experienced during the first epidemic wave (until June 2020). The score performance was validated by applying the model to several validation sets, that is, Lombardy population (second epidemic wave), and the other four Italian regions (entire 2020) for a total of about 15.4 million individuals and 7031 outcomes. Predictive performance was assessed by discrimination (areas under the receiver operating characteristic curve) and calibration (plot of observed vs predicted outcomes).

Results We observed a clear positive trend towards increasing outcome incidence as the score increased. The areas under the receiver operating characteristic curve of the COVID-19 Vulnerability Score ranged from 0.85 to 0.88, which compared favourably with the areas of generic scores such as the Charlson Comorbidity Score (0.60). A remarkable performance of the score on the calibration of observed and predicted outcome probability was also observed.

Conclusions A score based on data used for public health management accurately predicted the occurrence of severe/fatal manifestations of COVID-19. Use of this score may help health decision-makers to more accurately identify high-risk citizens who need early preventive or treatment interventions.

STRENGTHS AND LIMITATIONS OF THIS STUDY

The COVID-19 Vulnerability Score (CVS), based on demographic (age and gender) and clinical (29 conditions and diseases) predictors of the COVID-19 severity, may be easily obtained from electronic health databases covering beneficiaries of the National Health Service.

The CVS was developed and validated on a large (more than 15 million Italian individuals) and unselected population.

The CVS was validated across different temporal (first and second epidemic wave) and geographical (five Italian regions) conditions.

Predictors were restricted to those routinely collected and available in the Italian administrative databases. Thus, education, functional status and socioeconomic information were not included.

INTRODUCTION

The pandemic spread of the SARS-CoV-2 has dramatically exceeded the diagnostic and treatment capabilities of virtually all countries around the world. This has fuelled a debate on the need to establish priority criteria that might identify patients with COVID-19 at greater risk of progressing to hospitalisation or a fatal event, in order to make them the preferential recipients of currently available effective treatment strategies, the goal being to reduce the number of deaths and prevent collapse of hospital facilities. The problem
involves who should receive early diagnostic testing, who can be treated outside hospital among infected people, who should be given new, sometimes expensive and necessarily rationed drugs (eg, monoclonal antibodies) and who should be selected for early vaccination. The case of vaccination is particularly delicate because demand will outstrip supply for many months ahead in low/middle-income countries.

Associations between certain chronic diseases and conditions and serious/critical/fatal clinical manifestations of the SARS-CoV-2 infection have been reported from several studies, which potentially helps to identify the multiple prognostic factors that are involved in COVID-19. However, although some factors have been accepted as ‘established’ by the scientific community, their overall predictive value has not been robustly evaluated. It should also be considered that basing predictions on a list of individual conditions or diseases does not take into account that comorbidities can make the global risk different from that predictable by individual contributions. Finally, some predictive scores have been developed and validated in hospital care settings, their use requiring specialised image acquisition or sophisticated laboratory examinations, which may not be readily applicable in a population context. A valuable goal would therefore be to develop a score that could reliably predict the risk of progression of COVID-19 to severe or lethal forms, using simple and easily collectable information.

Our population-based study was performed under the auspices of the Italian Health Ministry. We aimed to develop and validate a novel score predictive of severe/fatal clinical manifestations of the SARS-CoV-2 infection using the multiple source information provided by the healthcare utilisation databases of the Italian National Health Service (NHS).

**METHODS**

**Setting**

This study was based on the NHS beneficiaries of five Italian regions that voluntarily joined the protocol and contributed to the data collection. The regions are located in Northern (Valle d’Aosta and Lombardy), Central (Marche), Southern (Puglia) Italy and in the Italian islands (Sicily). Overall, the data covered nearly 20.5 million people (34% of the entire Italian population) who, during 2020, experienced 712,408 cases of COVID-19, with a total of 31,957 deaths. Selected features of the participating regions are reported in online supplemental table S1.

**Data sources**

All Italian citizens have equal access to healthcare services provided by the NHS. Computerised information systems on the provided services have been created within each of the 21 Italian regions and autonomous provinces, the related regional healthcare databases including (1) demographic and administrative data of residents who receive NHS assistance (the NHS beneficiaries, practically coinciding with the entire resident population); (2) hospital discharge records reporting information on the primary diagnosis, as well as on up to five coexisting conditions and procedures, coded according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) classification system (http://icd9.chrisendres.com/); and (3) drug prescriptions reimbursed by the NHS, coded according to the Anatomical Therapeutic Chemical (ATC) classification system (https://www.whocc.no/atc-ddd_index/). Since the start of the COVID-19 pandemic, almost all regions established, with the coordination of the National Health Institute, a population-based registry of patients with a confirmed diagnosis of infection with SARS-CoV-2, and, among these, those who were admitted to intensive care units or died. In the present study, these various types of data were interconnected by using for each citizen a single identification code in all databases. To preserve privacy, each identification code was automatically deidentified. Analyses of the regional databases were performed under the rule that the inverse process, that is, patient identification, was allowed only to the Regional Health Authority upon request from the judicial authority.

**Predictors of COVID-19 severity**

Taking into consideration the morbidity and mortality predictors reported in epidemiological studies, as well as comorbidity scores widely used worldwide or tuned to the Italian population (the Charlson Comorbidity Index and the Multisource Comorbidity Score (MCS), developed for the general Italian population), we identified 61 candidate predictors. Twenty-seven candidate predictors were traced from inpatients diagnostic codes, 5 from outpatients who were prescribed drugs, and the remaining 29 from both diagnostic and therapeutic codes, depending on the availability of specific diagnostic codes and drug therapies. Four of us (FR, DM, MG and GM) independently attributed the ICD-9 and ATC codes to the individuals in whom 1 or more of the 61 candidate predictors were detectable. Discrepancies were resolved in conference. The list of candidate predictors, and the corresponding codes, are reported in online supplemental table S2.

**Score development**

Since among the five participating regions, Lombardy has the largest resident population (16% of the entire Italian population) and had been hit by the pandemic more than any other region during the months between March and June 2020 (in that period, 48% of the COVID-19 deaths registered in Italy occurred in Lombardy), we used the data from the first epidemic wave that hit Lombardy to develop the score.

We included all the NHS beneficiaries who on 21 February 2020 were residents in Lombardy for at least 2 years, were aged 18–79 years and did not reside in a nursing home. Multivariate logistic regression was fitted for investigating the association between gender, four age classes (18–45, 46–59, 60–69 and 70–79 years) plus the above-mentioned 61 candidate predictors,
and the odds of experiencing the outcome of interest, which was the composite of hospitalisation in an intensive care unit or death with a COVID-19 diagnosis, up to 30 June 2020. Candidate predictors entered as dichotomous variables in the model, with value 1 or 0 according to whether the specific condition was or was not recorded at least once within the 781 days prior to the baseline period, that is, from 1 January 2018 until 20 February 2020. The least absolute shrinkage and selection operator (LASSO) method was applied for selecting the conditions able to predict the outcome. Finally, a score was assigned to each condition selected with the LASSO method by using the coefficient estimated from the model. The coefficient was converted into a score by multiplication by 10 and rounding to the nearest whole number. Scores were sequentially summed to produce a total aggregate score. The index so obtained was termed COVID-19 Vulnerability Score (CVS). To verify the extension of the association between the increasing value of the score and the increasing occurrence of severe/fatal forms of COVID-19, CVS categories of width 10 was plotted against the outcome incidence. The prevalence of the Lombardy cohort members according to CVS categories was also calculated. Restricted cubic spline with 3 df was used to represent the corresponding smoothed trends.

Score validation and performance
To validate the model across different temporal and geographical conditions (ie, to assess the performance of CVS for different treatment options, climatic characteristics, intensity of the epidemic spread, etc), the score developed from the Lombardy cohort was applied to several validation sets selected by using the same inclusion/exclusion criteria of the original (Lombardy) one. One validation set consisted of the cohort of Lombardy NHS beneficiaries who were free from COVID-19 up to 1 July 2020, after which date a new observation period was started and continued until censorship at the outcome occurrence (intensive care admissions or deaths) or at 31 December 2020, whatever happened first. Other validation sets consisted of NHS beneficiaries from each of the other regions included in the study. For these other regional cohorts, observations started on 1 March 2020 and were censored at the outcome occurrence or at 31 December 2020, whatever happened first.

The performance of CVS was assessed through discrimination and calibration. Discrimination was evaluated by the receiver operating characteristic (ROC) curves and the corresponding underlying areas (area under the ROC curves (AUCs)). Calibration plots displayed observed versus predicted outcome probabilities. The Hosmer-Lemeshow goodness-of-fit test modified by Yu et al was used for testing the null hypothesis of agreement between observed and predicted outcome probabilities.

Patient and public involvement
No patient was involved in setting the research question or the outcome measures, nor were patients involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results.

RESULTS
COVID-19 Vulnerability Score
The 31 demographic and clinical conditions that significantly contributed to CVS are reported in table 1. As expected, older age was the major contributor to the outcome of interest, but also male gender gave a relevant contribution. Nearly 40% of NHS beneficiaries had at least one clinical condition contributing to CVS. Diabetes (especially if under insulin therapy), psychosis, coronary and peripheral vascular disease, gout, use of corticosteroids, HIV infection, malignancies and anaemias were the most relevant contributors to the outcome. However, other 19 clinical conditions (ranging across all major nosologic macrocategories) contributed to CVS.

Figure 1A shows that the probability of experiencing the outcome of interest had a clear positive trend as CVS increased, the risk being lower than 0.05% for CVS value ≤29, progressing to 2% for a CVS value between 60 and 69, and reaching a much higher value (around 4%) for CVS values ≥70. Sixty-nine per cent of NHS beneficiaries had a CVS value ≤29, almost 30% ranged from 30 and 69, and less than 1% (0.16%) exhibited a CVS value ≥70 (figure 1B).

CVS performance
Figure 2A shows that the AUC of CVS was 0.89. This area compared favourably with the AUC of the models based on scores not specifically addressing COVID-19, the AUC values being 0.60 for the Charlson Comorbidity Index and 0.77 for MCS. The 95% CIs are not indicated in the figure because, due to the very large sample size, they practically coincided with the AUC values. As shown in figure 2B, the CVS AUC values were almost superimposable between the different regions participating in the study, that is, 0.88, 0.86, 0.86, 0.85 and 0.86 for Lombardy, Valle d’Aosta, Marche, Puglia and Sicily cohorts, respectively.

Figure 3 shows that there was a good agreement between the observed and the predicted outcome probabilities, with the calibration intercept close to the ideal value of 0 and the recalibration slope close to the ideal value of 1 (0.93). The null hypothesis of agreement between observed and predicted frequencies could not be rejected according to the modified Hosmer-Lemeshow test.

DISCUSSION
Our study shows that a score based on demographic and clinical information derived from healthcare utilisation data currently used throughout Italy for the management of NHS is able to stratify NHS beneficiaries aged 18–79...
<table>
<thead>
<tr>
<th>Condition</th>
<th>Number (%)</th>
<th>Number of outcome events</th>
<th>Incidence every 10 000</th>
<th>OR*</th>
<th>90% CI*</th>
<th>Weight†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male gender</td>
<td>3797636 (49.6)</td>
<td>6849</td>
<td>18.0</td>
<td>3.07</td>
<td>2.95 to 3.19</td>
<td>11</td>
</tr>
<tr>
<td>Age ≤45</td>
<td>3111426 (40.6)</td>
<td>271</td>
<td>0.9</td>
<td>1.00</td>
<td>Reference</td>
<td>0</td>
</tr>
<tr>
<td>Age 46–59</td>
<td>2305062 (30.1)</td>
<td>1435</td>
<td>6.2</td>
<td>5.95</td>
<td>5.36 to 6.62</td>
<td>18</td>
</tr>
<tr>
<td>Age 60–69</td>
<td>1222310 (16.0)</td>
<td>2506</td>
<td>20.5</td>
<td>15.62</td>
<td>14.09 to 17.32</td>
<td>27</td>
</tr>
<tr>
<td>Age 70–79</td>
<td>1016704 (13.3)</td>
<td>4948</td>
<td>48.7</td>
<td>27.64</td>
<td>24.96 to 30.61</td>
<td>33</td>
</tr>
<tr>
<td>HIV infection</td>
<td>31300 (0.4)</td>
<td>154</td>
<td>49.2</td>
<td>1.52</td>
<td>1.33 to 1.74</td>
<td>4</td>
</tr>
<tr>
<td>Other infectious and parasitic diseases</td>
<td>42422 (0.6)</td>
<td>443</td>
<td>104.4</td>
<td>1.37</td>
<td>1.26 to 1.49</td>
<td>3</td>
</tr>
<tr>
<td>Malignancies</td>
<td>177024 (2.3)</td>
<td>1073</td>
<td>60.6</td>
<td>1.42</td>
<td>1.35 to 1.50</td>
<td>4</td>
</tr>
<tr>
<td>Diabetes without insulin therapy</td>
<td>278785 (3.6)</td>
<td>1419</td>
<td>50.9</td>
<td>1.60</td>
<td>1.53 to 1.68</td>
<td>5</td>
</tr>
<tr>
<td>Insulin therapy</td>
<td>101996 (1.3)</td>
<td>973</td>
<td>95.4</td>
<td>2.35</td>
<td>2.21 to 2.49</td>
<td>9</td>
</tr>
<tr>
<td>Obesity</td>
<td>16571 (0.2)</td>
<td>103</td>
<td>62.2</td>
<td>1.34</td>
<td>1.13 to 1.58</td>
<td>3</td>
</tr>
<tr>
<td>Disorders of fluid, electrolyte and acid-base balance</td>
<td>8576 (0.1)</td>
<td>135</td>
<td>157.4</td>
<td>1.29</td>
<td>1.11 to 1.49</td>
<td>3</td>
</tr>
<tr>
<td>Gout</td>
<td>164428 (2.2)</td>
<td>1518</td>
<td>92.3</td>
<td>1.57</td>
<td>1.50 to 1.66</td>
<td>5</td>
</tr>
<tr>
<td>Coagulation defects</td>
<td>3603 (0.1)</td>
<td>36</td>
<td>99.9</td>
<td>1.41</td>
<td>1.07 to 1.85</td>
<td>3</td>
</tr>
<tr>
<td>Anaemias</td>
<td>613430 (8.0)</td>
<td>2228</td>
<td>36.3</td>
<td>1.51</td>
<td>1.45 to 1.58</td>
<td>4</td>
</tr>
<tr>
<td>Dementia/Alzheimer</td>
<td>12671 (0.2)</td>
<td>145</td>
<td>114.4</td>
<td>1.26</td>
<td>1.09 to 1.46</td>
<td>2</td>
</tr>
<tr>
<td>Psychosis</td>
<td>138034 (1.8)</td>
<td>684</td>
<td>49.6</td>
<td>1.94</td>
<td>1.80 to 2.08</td>
<td>7</td>
</tr>
<tr>
<td>Depression</td>
<td>588688 (7.7)</td>
<td>1729</td>
<td>29.4</td>
<td>1.35</td>
<td>1.29 to 1.42</td>
<td>3</td>
</tr>
<tr>
<td>Parkinson’s disease and parkinsonism</td>
<td>40885 (0.5)</td>
<td>274</td>
<td>67.0</td>
<td>1.21</td>
<td>1.09 to 1.34</td>
<td>2</td>
</tr>
<tr>
<td>Epilepsy and recurrent seizures</td>
<td>122171 (1.6)</td>
<td>510</td>
<td>41.7</td>
<td>1.37</td>
<td>1.26 to 1.48</td>
<td>3</td>
</tr>
<tr>
<td>Other diseases of the nervous system and sense organs</td>
<td>35495 (0.5)</td>
<td>253</td>
<td>71.3</td>
<td>1.26</td>
<td>1.13 to 1.40</td>
<td>2</td>
</tr>
<tr>
<td>Ischaemic heart disease/angina</td>
<td>91539 (1.2)</td>
<td>845</td>
<td>92.3</td>
<td>1.18</td>
<td>1.11 to 1.26</td>
<td>2</td>
</tr>
<tr>
<td>Heart failure</td>
<td>21840 (0.3)</td>
<td>428</td>
<td>196.0</td>
<td>1.30</td>
<td>1.18 to 1.43</td>
<td>3</td>
</tr>
<tr>
<td>Vascular diseases</td>
<td>14936 (0.2)</td>
<td>217</td>
<td>145.3</td>
<td>1.17</td>
<td>1.04 to 1.32</td>
<td>2</td>
</tr>
<tr>
<td>Cerebrovascular diseases</td>
<td>35205 (0.5)</td>
<td>333</td>
<td>94.6</td>
<td>1.12</td>
<td>1.02 to 1.23</td>
<td>1</td>
</tr>
<tr>
<td>Hypertension</td>
<td>796044 (10.4)</td>
<td>3136</td>
<td>39.4</td>
<td>1.20</td>
<td>1.15 to 1.25</td>
<td>2</td>
</tr>
<tr>
<td>Coronary and peripheral vascular disease</td>
<td>658737 (8.6)</td>
<td>2668</td>
<td>40.5</td>
<td>1.75</td>
<td>1.68 to 1.82</td>
<td>6</td>
</tr>
<tr>
<td>Oral anticoagulant agents</td>
<td>144713 (1.9)</td>
<td>1221</td>
<td>84.4</td>
<td>1.39</td>
<td>1.32 to 1.47</td>
<td>3</td>
</tr>
<tr>
<td>COPD/asthma</td>
<td>20034 (0.3)</td>
<td>268</td>
<td>133.8</td>
<td>1.15</td>
<td>1.03 to 1.28</td>
<td>1</td>
</tr>
<tr>
<td>Liver cirrhosis and other liver chronic diseases</td>
<td>29484 (0.4)</td>
<td>177</td>
<td>60.0</td>
<td>1.31</td>
<td>1.16 to 1.49</td>
<td>3</td>
</tr>
</tbody>
</table>

Continued
Table 1 Continued

<table>
<thead>
<tr>
<th>Incidence every 10000</th>
<th>OR*</th>
<th>90% CI*</th>
<th>Weight†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic kidney disease</td>
<td>216.8</td>
<td>1.32</td>
<td>1.20 to 1.46</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissues</td>
<td>33.1</td>
<td>1.10</td>
<td>1.00 to 1.20</td>
</tr>
<tr>
<td>Chronic pain</td>
<td>52.6</td>
<td>1.28</td>
<td>1.21 to 1.36</td>
</tr>
<tr>
<td>Corticosteroids</td>
<td>27.7</td>
<td>1.62</td>
<td>1.55 to 1.68</td>
</tr>
<tr>
<td>Individuals without any of the 29 conditions above listed</td>
<td>2.9</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

The analysis was based on the cohort of 765502 beneficiaries of the Lombardy Region Health Service for at least 2 years, who on 21 February 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9160 severe (intensive care unit admitted and mechanically ventilated via intubation) and/or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10000 people at risk.

*OR, and 90% CI, estimated by multivariable logistic regression. ORs measured the strength of the association between the presence/absence of each of the listed contributors and the outcome odds.
†Weights were obtained from the coefficients of the logistic model; the latter were converted into scores by multiplying them by 10 and rounding them to the nearest whole number.

COPD, chronic obstructive pulmonary disease.
that the association between the severity of COVID-19 and the dispensed drugs we found in our study is not in contrast with the use of some of these drugs for the treatment of COVID-19, because in our analysis, previous drug therapies were searched for to track background comorbidities and not to investigate their possible direct effect on the disease. In this context, it is likely that use of corticosteroids and other immunosuppressive agents reflected the existence of autoimmune diseases, while use of anticoagulants reflected the existence of atrial fibrillation, thromboembolic states or other cardiovascular disorders, which have been shown to reduce patients’ defence against the virus.50

Our study has implications for several aspects of the public health policy against COVID-19, the most important of which is the priority criteria to adopt for the third dose of vaccine to be delivered to the Italian population by the Italian Ministry of Health. As done in the first vaccination campaign, the plan is to offer an early cost-free priority third dose to people residing in a nursing home and aged 80 years or older. This has a strong rationale because of the 24,575 severe/fatal cases of COVID-19...
registered in Lombardy during 2020, 12,593 (51%) occurred in people aged 80 years and older. Furthermore, in Italy, the average age of COVID-19 fatalities during the entire pandemic period has been reported to be 82 years, which means that in octogenarians and nonagenarians, search for and use of a risk score more complex than age alone may carry a limited practical advantage. However, this is not the case for the vaccination programme to be implemented in people aged 79 years or less, in which administration of the third dose vaccine is planned after completion of the third dose vaccination in older individuals. In these people, use of CVS may offer the possibility of identifying more accurately those at a high risk of development of a severe or lethal form of COVID-19 and thus to predispose their vaccination reinforcement at an earlier time. The same advantage can be foreseen for

**Figure 2** Receiver operating characteristic (ROC) curves comparing discriminant power (A) of COVID-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (B) of CVS from several validation sets. Derivation set (left box) was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least 2 years, who on 21 February 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (intensive care unit admitted and mechanically ventilated via intubation) and/or fatal outcomes. Validation sets (right box) were based on: (1) 7,575,924 resident in Lombardy whom observation started on 1 July 2020 and who experienced 2,822 severe/fatal outcomes within 31 December 2020; (2) 92,267, 1,110,570, 3,012,754 and 3,649,518 beneficiaries of Valle d’Aosta, Marche, Puglia and Sicily regional health services, whom observation started on 1 March 2020 and who, respectively, experienced 173, 542, 1,953 and 1,541 severe/fatal outcomes within 31 December 2020.

**Figure 3** Calibration plot of observed (X-axis) versus predicted (Y-axis) risk of severe/fatal outcomes. The analysis was based on the pooled validation sets of 15,441,033 residents from Lombardy, Valle d’Aosta, Marche, Puglia and Sicily who experienced 7,031 severe/fatal outcomes from starting (1 July 2020 in Lombardy, or 1 March 2020 in the other regions) until 31 December 2020.
the criteria to adopt for the delivery of future treatment strategies such as new antiviral drugs or monoclonal antibodies, if current research will prove their life-saving role. In this case, the high cost of these treatments will make priority criteria for their use absolutely necessary.

The present study has several strengths and some limitations. An important strength is that our sample of NHS beneficiaries was not only extremely large but it also reflected an unselected population. Another strength is that the Italian healthcare utilisation database allows to track services provided by the NHS with considerable accuracy because providers must document services to claim reimbursement, and incorrect reports carry legal consequences. Finally, a remarkable finding of our study is that, although built from the Lombardy data collected during the first epidemic wave (ie, before the summer 2020), CVS performed similarly well during the second epidemic wave (ie, after the summer 2020), despite differences in treatment options for inpatients and outpatients as well as hospitalisation criteria compared with the first epidemic wave. It is also remarkable that the CVS performance was virtually superimposable in all regions of Italy, despite their different social features, climatic characteristics and intensity of the epidemic spread. This suggests that the advantages of the CVS score for stratification of the risk of COVID-19 complications extend across different temporal and geographical conditions.

The limitations are that the predictors of COVID-19 complications we searched for are restricted to those routinely collected and available in the administrative databases (the same for all regions of Italy), that is, hospital admissions and drug dispensed. Thus, educational factors, functional status, socioeconomic characteristics and other extraclinical variables that can affect the prognosis of patients with COVID-19 were not included. Our scoring system also did not capture the severity of associated comorbidities, health services and treatments supplied by private providers, and misdiagnosis (due to poor accuracy in reporting diagnoses and comorbidities) and upcoding of hospital records.

Finally, our approach may have failed to identify comorbidities that, although increasing the risk of severe/fatal clinical manifestations of COVID-19 limited social contacts, thereby favouring an escape from the SARS-CoV-2 virus infection of the individuals affected. However, because the purpose of our study was to identify individuals to which offer earlier protection, patients with a disease that makes them unexposed to the infection should receive later preventive interventions (ie, treatments or vaccination). Of course, exclusion from the scoring system of diseases so debilitating or incapacitating to limit social contacts but requiring a caregiver is a major limitation of our study.

**CONCLUSION**

In summary, we developed and validated a score derived from data used for public health management, which predicts severe/fatal outcomes of COVID-19 in a large number of beneficiaries of the Italian NHS more accurately than other available scores. Our findings show that this can be achieved by combined use of demographic (age and gender) and clinical (29 conditions/diseases) predictors of the COVID-19 outcome. Because of its performance, use of this score may help health decision-makers to achieve a more accurate identification of high-risk citizens who need early preventive interventions.

**Author affiliations**

1Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy
2National Centre for Healthcare Research and Pharmacoepidemiology, University of Milan-Bicocca, Milan, Italy
3Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
4Department of Health Services and Epidemiological Observatory, Regional Health Authority of Sicily, Palermo, Italy
5Regional Health Agency of Puglia, Bari, Italy
6Regional Health Authority, Aosta, Italy
7Regional Health Agency of Marche, Ancona, Italy
8Regional Welfare Service, Milan, Italy
9Department of Health Planning, Italian Health Ministry, Rome, Italy
10Department of Internal Medicine, Hospital Fatebenefratelli, Rome, Italy
11Institute of Tropical and Infectious Diseases, University of Milan L Sacco Hospital, Milan, Italy
12University of Milano-Bicocca, Milan, Italy
13Policlinico di Monza, Monza, Italy


**Contributors** GC conceived the idea for this manuscript. GC, FR and FC designed the study. GC and GM drafted the manuscript. FR, AA, AD, SA, MI and ME performed the data analysis. SS, VL, CT, PV, LS and RB extracted the data and authorised their utilisation. All authors assisted in the results interpretation and manuscript revision. All authors read and approved the final manuscript. Giovanni Corrao is the guarantor of the overall content of the work.

**Funding** This study was funded by a research grant from the Italian Health Ministry: ‘Modelli per il monitoraggio e la valutazione delle cure integrate (CI) nell’ambito del Nuovo Sistema di Garanzia dell’assistenza sanitaria’ project (grant number J59H06000160001).

**Disclaimer** The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Competing interests** GC received research support from the European Community (EC), the Italian Agency of Drug (AIFA), the Italian Ministry of Education, University...
and Research (MIUR), and the Italian Health Ministry. He took part in a variety of projects that were funded by pharmaceutical companies (ie, Novartis, GSK, Roche, AMGEN and BMS). He also received honoraria as member of Advisory Board from Roche. GM received honoraria for participation as speaker/chairman in national/international meetings from Boehringer Ingelheim, Ferrer, Medtronic, Menarini, Merck Serono, Recordati and Servier.

Patient consent for publication Not required.


Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data may be obtained from a third party and are not publicly available. The data that support the findings of this study are available from the Italian Regions, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the Italian Regions upon reasonable request.

Supplemental material This content has been supplied by the author(s).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs Giovanni Corrao http://orcid.org/0000-0002-1034-8444
Federico Rea http://orcid.org/0000-0001-2798-5101
Vito Lepore http://orcid.org/0000-0002-5466-780X

REFERENCES
