BMJ Open

Addressing vaccination priority by stratifying general population according with frailty: the new Covid-19 vulnerability score (CVS)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-053281
Article Type:	Original research
Date Submitted by the Author:	10-May-2021
Complete List of Authors:	Corrao, Giovanni; Università degli Studi di Milano-Bicocca Rea, Federico; University of Milan–Bicocca Carle, Flavia; Polytechnic University of Marche Scondotto, Salvatore; Department of Epidemiologic Observatory, Health Department of Sicily Allotta, Alessandra; Regional Health Authority, Department of Health Services and Epidemiological Observatory Lepore, Vito; Regional Health Agency of Puglia D'Ettorre, Antonio; Regional Health Agency of Puglia Tanzarella, Cinzia; Regional Health Agency of Puglia Vittori, Patrizia; Regional Health Authority Abena, Sabrina; Regional Health Authority Iommi, Marica; Polytechnic University of Marche Spazzafumo, Liana; Regional Health Agency of Marche Ercolanoni, Michele; Regional Welfare Service Blaco, Roberto; Regional Welfare Service Carbone, Simona; Italian Health Ministry, Department of Health Planning Giordani, Cristina; Italian Health Ministry, Department of Health Planning Manfellotto, Dario; Hospital Fatebenefratelli - AFaR, Department of Internal Medicine Galli, Massimo; University of Milan L. Sacco Hospital, Institute of Tropical and Infectious Diseases Mancia, Giuseppe; University of Milano-Bicocca, Clinical Medicine and Prevention; Ospedale San Gerardo, Clinia Medica
Keywords:	COVID-19, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Addressing vaccination priority by stratifying general population according with frailty: the new Covid-19 vulnerability score (CVS)

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

Word count (manuscript): 3,041

Tables: 1

Figures: 3

Address for correspondence: Prof. Giovanni Corrao, Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi, 8, Edificio U7, 20126 Milano, Italy. Tel.: +39.02.64485854; E-mail: giovanni.corrao@ unimib.it

"Monitoring and Assessing care Pathways" MAP working group (Italian Health Ministry, Health Planning Dept):

- Italian Ministry of Health, Dept of Health Planning: Donata Bellentani, Simona Carbone (coordinator), Carla Ceccolini, Angela De Feo, Cristina Giordani, Rosanna Mariniello, Modesta Visca; Dept of health prevention: Natalia Magliocchetti, Giovanna Romano; External Expert: Antonio Lora, Paola Pisanti, Rinaldo Zanini.
- Polytechnic University of Marche (coordinator): Flavia Carle, Marica Iommi, Edlira Skrami.
- University of Milano-Bicocca, Laboratory of Healthcare Research & Pharmacoepidemiology: Anna Cantarutti, Giovanni Corrao, Matteo Monzio Compagnoni, Pietro Pugni, Federico Rea.
- Department of Epidemiology Lazio Region: Marina Davoli, Mirko Di Martino, Adele Lallo.
- Aosta Valley Region: Patrizia Vittori, Giuliana Vuillermin
- Campania Region: Alfonso Bernardo, Anna Frusciante
- Emilia Romagna Region: Laura Belotti, Rossana De Palma.
- Friuli Venezia Giulia Region: Andrea Di Lenarda, Marisa Prezza
- Lazio Region: Danilo Fusco, Chiara Marinacci
- Lombardy Region: Roberto Blaco, Olivia Leoni
- Marche Region: Liana Spazzafumo, Simone Pizzi
- Molise Region: Lolita Gallo
- Puglia Region: Ettore Attolini, Vito Lepore
- Sicily Region: Salvatore Scondotto, Giovanni De Luca
- Tuscany Region: Paolo Francesconi, Carla Rizzuti
- Veneto Region: Francesco Avossa, Silvia Vigna
- Research and Health Foundation (Fondazione ReS -Ricerca e Salute-): Letizia Dondi, Nello Martini, Antonella Pedrini, Carlo Piccinni
- National Agency for Regional Health Services: Mimma Cosentino, Maria Grazia Marvulli
- ANMCO (National Association of Hospital Cardiologists) Study Center: Aldo Maggioni

Abstract

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

Design. Retrospective observational cohort study.

Setting. Population-based study using the healthcare utilization database of five Italian Regions.

Participants. Beneficiaries of the National Health Service, resident in one of the five participating Regions, aged 18-79 years, and not reside in a nursing home. The model was built from the 7,655,502 beneficiaries of the Lombardy Region Health Service.

Main outcome measure. The score included gender, age and 29 conditions selected from a list of 61 candidates for independently predicting severe/fatal clinical manifestations of infection. The outcome was the severe (ICU admitted)/fatal manifestations of Covid-19 experienced during the first epidemic wave (until June 2020). CVS performance was validated by applying the model to several validation sets (populations from Lombardy, second epidemic wave, and other four Italian regions during 2020) for a total of about 15.4 million individuals and 7,031 outcomes. Predictive performance was assessed by discrimination (areas under the ROC curve) and calibration (plot of observed vs. predicted outcomes).

Results. A clear positive trend towards increasing outcome incidence as CVS increases was observed. Areas under the ROC curve of CVS ranged from 0.85 to 0.88, which were better than those of generic scores such as the Charlson Comorbidity Score (0.60). A remarkable calibration of observed and predicted outcome probability was observed.

Conclusions. A simple score from data used for public health management accurately predicted the occurrence of severe/fatal manifestations of Covid-19. Because of its performance, the use of CVS may help health decision-makers to achieve more accurate identification of high-risk citizens who need early preventive interventions, mainly the vaccine coverage.

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, is able to predict severe/fatal clinical manifestations of SARS-CoV-2 infection among citizen aged 18-79 years.
- The score may be easily drawn from electronic health databases covering beneficiaries of health systems (e.g., National Health Service, health insurance companies).
- The CVS showed a better discriminating power than other comorbidities scores, and it performed similarly well across Italian Regions as well as across time (first and second epidemic waves).
- 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 virus 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 Covid-19 patients at greater risk to progress to hospitalization 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 (e.g., monoclonal antibodies [1]) and who should be selected for early vaccination. The case of vaccination is particularly delicate because demand will outstrip supply for many months ahead.

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 [2-4], so opening the gate towards identifying multiple prognostic factors for the Covid-19 disease. However, although some factors have been accepted as "established" by the scientific community, their overall predictive value has not been robustly evaluated [5]. It should be in addition considered that basing the prediction 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 proposed and validated in hospital care settings [6,7], their use requiring specialized image acquisition or sophisticated laboratory examinations, which might be hardly applicable to the population at large. This implies that developing a score able to reliably predict the risk of progression of the Covid-19 disease to its most severe or lethal forms in the general population via simple and easily collectable information would be a valuable goal.

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 utilization 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) and Islands (Sicily) of Italy. 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 supplementary **Table S1**.

Data sources

All Italian citizens have equal access to healthcare services provided by the NHS. Computerized information systems on the provided services have been created within each of the 21 Italian Regions and autonomous Provinces, the related regional health care databases including 1) demographic and administrative data of residents who receive NHS assistance (the NHS beneficiaries, practically coincident 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 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 starting of the Covid-19 pandemic, almost all Regions established, under the coordination of the National Health Institute, a population-based

registry of patients with a confirmed diagnosis of infection with the SARS-CoV-2 virus, 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, i.e. patient identification, was allowed only to the Regional Health Authority upon request from the judicial authority.

Candidate predictors

Taking into consideration the morbidity and mortality predictors reported in epidemiological studies [5,7-9], as well as comorbidity scores widely used worldwide or tuned to the Italian population (the Charlson index (CCI) [10] and the Multisource Comorbidity Score (MCS), developed for the general Italian population [11]), we identified 61 candidate predictors. Twenty-seven candidate predictors were traced from inpatients diagnostic codes, five from outpatients who were prescribed drugs, and the remaining twenty-nine 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 one 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 supplementary **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 elected to use the data from the first epidemic wave that hit Lombardy to develop the score.

We included all the NHS beneficiaries who at February 21, 2020 were resident in Lombardy since at least two years, were aged 18 to 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 hospitalization in an Intensive Care Unit or death with a Covid-19 diagnosis, up to June 30, 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, i.e., from January 1st, 2018 until February 20, 2020. The least absolute shrinkage and selection operator (LASSO) method was applied for selecting the conditions able to predict the outcome [12]. 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 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 degrees of freedom was used to represent the corresponding smoothed trends [13].

Score validation and performance

Validity of CVS was investigated by applying the score developed from the Lombardy cohort hit by the first pandemic wave (derivation set) to several validation sets selected by using the same inclusion/exclusion criteria of the derivation one. One validation set consisted of the cohort of Lombardy NHS beneficiaries who were free of Covid-19 up to July 1, 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 December 31, 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 March 1, 2020 and was censored at the outcome occurrence or at December 31, 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 (AUCs) [14]. Calibration plots displayed observed vs. predicted outcome probabilities. The Hosmer-Lemeshow goodness-of-fit test modified by Yu et al [15] was used for testing the null hypothesis of agreement between observed and predicted outcome probabilities.

Patient and Public involvement

No patients were involved in setting the research question or the outcome measures, nor were they 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 even 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, human immunodeficiency virus (HIV) infection, malignancies and anaemias were the most relevant contributors of the outcome. However, other 19 clinical conditions (ranging across all major nosologic macrocategories) contributed to CVS.

Figure 1, upper box, 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 ≥80. Sixty-nine percent 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 1, lower box**).

Covid-19 Vulnerability Score performance

Figure 2, left box, shows that the area under the ROC curve of CVS was 0.89. This compared favourably with the AUC of the models based on scores not specifically addressing Covid-19 whose AUC values were 0.60 and 0.77 for the CCI and MCS, respectively. The 95% confidence intervals 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 2, right box**, the CVS AUC values were 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 calibration intercept close to the ideal value of 0 and 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 simple score based on demographic and clinical information derived from healthcare utilization data currently used by all Italian Regions for the management of NHS is able to stratify NHS beneficiaries aged 18 to 79 years according to their risk to develop severe/fatal clinical manifestations of Covid-19. In a very large number of individuals from several Italian Regions the score we developed (termed Covid Vulnerability Score or CVS) exhibited a

significantly better discriminating power than the CCI, i.e. the most worldwide used comorbidity score [10] which has been recently validated also for predicting mortality in Covid-19 patients hospitalised for pneumonia [16]. Furthermore, CVS outperformed a comorbidity score recently validated by our group for the general Italian population and found to outperform the CCI. This allows concluding that our CVS can reliably identify people in whom age, gender and a variety of comorbidities interact to make them frailer to the development of the severe and fatal clinical manifestations of the Covid-19 infection. This may provide a useful tool for establishing priority in the vaccination programs for the general Italian population up to 79 years of age, with the exception of the priority given to individuals involved in specific job categories, is currently based in a descending fashion on age alone as well as on individually listed conditions or diseases that have shown a greater prevalence of severe or lethal Covid-19 infections in clinical studies. This tool may find a useful application also for the establishment of priority access to future treatment options, such as monoclonal antibodies, if their cost will be too high to allow NHS to plan an extended use.

Our study identified several prognostic factors that, in addition to age and gender, predict the severity of Covid-19 and are included among the medical illnesses and dispensed drugs retrievable in the healthcare utilization database. Consistently with a recent meta-analysis [4], diabetes (mainly when under insulin therapy), cardiovascular disease (mainly coronary and peripheral vascular disease), hypertension, malignancies, chronic respiratory and kidney diseases, dementia and obesity were all associated with the Covid-19 outcome. People with HIV [17], and those who had a recent history of severe clinical manifestations of an infectious disease, including tuberculosis [18], also showed a significant association. Additionally, and according with other studies, we found that diseases of the neurological system (e.g., epilepsy, recurrent seizures [19] and Parkinson disease and parkinsonism [20]), of the gastrointestinal tract (e.g., liver cirrhosis and other liver chronic diseases [21]), of the metabolism (e.g., gout [22]), of the skin (e.g. psoriasis

[23]), and of the blood and blood-forming organs (e.g. coagulation defects [24], anaemias [25]) contributed to the Covid-19 related clinical frailty. We also confirmed the involvement of frailty conditions related to mental disorders, such as psychosis and depression [26] as well as of recent dispensations of drugs with immunosuppressive properties (such as corticosteroids [27]) agents against chronic pain, e.g., narcotic analgesics [28]), and with an anticoagulant [29] action. This confirms the now established notion that alterations of the structure and function of virtually all organs and systems of the body may adversely affect resistance to the Covid-19 disease. It should be emphasized that the association between the severity of Covid-19 and the dispensed drugs we found in our study is not in contrast with the frequent use of some of these drugs for the treatment of Covid-19, because in our analysis previous drug therapies were searched for to further 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 and resistance to the virus [30].

Our study has implications for several aspects of the public health policy against Covid-19 to be considered in the future, in particular, as mentioned above, for the vaccination program planned by the Italian Ministry of Health. This program offers cost-free priority immunization to individuals carrying on jobs of fundamental social importance as well as to those in whom exposure to the infection is high. This is followed by people aged 80 years and older while in people aged 79 years or below vaccination is planned in the later stages of the vaccination program possibly (but this is not yet entirely clear) giving priority to people with serious comorbidities. The rationale for offering early vaccination to the oldest fraction of the population is strong because of the 24,575 severe/fatal cases of Covid-19 registered in Lombardy during 2020, almost 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 have been recently reported to be 82 years, which means that search for and use of a risk score more complex than age alone in octogenarians and nonagenarians may carry a limited practical advantage. On the contrary, use of our score may offer the possibility of identifying accurately younger high-risk people to whom offer vaccination at an earlier time.

The present study has several strengths. Because in Italy a public funded healthcare system involves virtually all citizens, our sample of NHS beneficiaries had not only an extremely large size but also reflected an unselected population. It is also a strength that the Italian healthcare utilization 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. It should also be emphasized the excellent ability of CVS for discriminating individuals who will experience and those who will not experience Covid-19 related serious complications. Finally, it is also remarkable that, although built according to Lombardy data collected during the first epidemic wave (i.e. before the summer 2020), CVS performed similarly well during the second epidemic wave (i.e. after the summer 2020) as well as in areas of Italy different from Lombardy for social features, climatic characteristics, and intensity of the epidemic spread. This suggests that the advantages of the CVS score for stratification of the risk the Covid-19 complications extends across different temporal and geographic conditions.

Limitations are that the predictors of Covid-19 we searched for are restricted to those routinely collected and available in the administrative databases, which are the same for all Regions of Italy, i.e., hospital admissions and drug dispensed. In addition, our scoring system did not capture the severity of associated comorbidities. Furthermore, health services and treatments supplied by private providers were not captured by our analysis. Moreover, misdiagnosis (due to poor accuracy in reporting diagnoses and comorbidities) and upcoding in hospital records (sometimes in pursuit of higher reimbursements) might have generated too conservative estimates of the CVS

performance. Finally, our approach may have failed to identify comorbidities that were as severe, debilitating or incapacitating as to substantially limit social contacts, thereby escaping infection by the SARS-CoV-2 virus. However, because the purpose of our study was to identify individuals at greater risk of severe/fatal clinical manifestations in order to offer them earlier protection, patients with a disease so debilitating as to make them unexposed to the infection would belong to a low risk category anyway.

Conclusion

In summary, we developed and validated a simple 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 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, mainly the vaccine coverage.

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). 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.

Conflicts of Interest

Giovanni Corrao received research support from the European Community (EC), the Italian Agency of Drug (AIFA), and the Italian Ministry of Education, University and Research (MIUR). He took part to a variety of projects that were funded by pharmaceutical companies (i.e., Novartis, GSK, Roche, AMGEN and BMS). He also received honoraria as member of Advisory Board from Roche.

Giuseppe Mancia received honoraria for participation as speaker/chairman in national/international meetings from Boehringer Ingelheim, Ferrer, Medtronic, Menarini Int, Merck Serono, Recordati, and Servier.

For the remaining authors, nothing was declared.

Contributors

GC conceived the idea for this manuscript. GC, FR and FC designed the study. GC and GM drafted the manuscript. FR, AA, ADE, 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 the results interpretation and manuscript revision. All authors read and approved the final manuscript.

Ethical approval

Under the rules of the Italian Drugs Agency (available at: http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf), retrospective studies using administrative databases do not require Ethics Committee protocol approval.

Data sharing

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.

Dissemination to participants and related patient and public communities

There are no plans to disseminate the results of the research to study participants or the relevant patient community.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained

References

[1] Lloyd EC, Gandhi TN, Petty LA. Monoclonal Antibodies for COVID-19. JAMA. 2021 Feb 5. doi: 10.1001/jama.2021.1225. Epub ahead of print. PMID: 33544136

[2] Földi M, Farkas N, Kiss S, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. Obes Rev. 2020 Oct;21(10):e13095

[3] Mantovani A, Byrne CD, Zheng MH, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2020;30:1236-48

[4] Lippi G., Wong J., Henry B. M. Hypertension and its severity or mortality in Coronavirus Disease 2019 (COVID-19): a pooled analysis. Polish Archives of Internal Medicine 2020 doi: 10.20452/pamw.15272

[5] Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. PLoS One 2020;15:e0241955

[6] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients with COVID-19. JAMA Intern Med. 2020; 10.1001/jamainternmed.2020.2033

[7] Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328

[8] Gupta RK, Marks M, Samuels THA, et al; UCLH COVID-19 Reporting Group. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J 2020;56:2003498

[9] Ebrahimi M, Malehi AS, Rahim F. COVID-19 Patients: A Systematic Review and Meta-Analysis of Laboratory Findings, Comorbidities, and Clinical Outcomes Comparing Medical Staff versus the General Population. Osong Public Health Res Perspect 2020;11:269-79

[10] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83

[11] Corrao G, Rea F, Di Martino M, et al. Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. BMJ Open 2017;7:e019503

[12] Tibshirani R. The LASSO method for variable selection in the Cox model. Stat Med 1997;16:385-95

[13] Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. Bone Marrow Transplantation 2020;55:675-80

[14] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72

[15] Yu W, Xu W, Zhu L. A modified Hosmer–Lemeshow test for large data sets. Communications in Statistics - Theory and Methods 2017;46:11813-25

[16] Christensen DM, Strange JE, Gislason G, et al. Charlson Comorbidity Index Score and Risk of Severe Outcome and Death in Danish COVID-19 Patients. J Gen Intern Med 2020;35:2801-3

[17] Shareef MA, Bashaiwth HM, AlAkbari AO, et al. A systematic review of contemporary evidence on SARS-CoV-2 and HIV coinfection: What does it look like up to date? Avicenna J Med 2020;10:189-97

[18] Liu Y., Bi L., Chen Y. Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity. medRxiv. March 2020 doi: 10.1101/2020.03.10.20033795. 2020.03.10.20033795

[19] Cabezudo-Garcia P, Ciano-Petersen NL, Mena-Vazquez N, et al. Incidence and case fatality rate of COVID-19 in patients with active epilepsy. Neurology 2020;95:e1417-e1425

[20] Vignatelli L, Zenesini C, Belotti LMB, et al. Risk of hospitalization and death for COVID-19 in people with Parkinson's disease or parkinsonism. Mov Disord. 2020 Nov 16:10.1002/mds.28408

[21] Bajaj JS, Garcia-Tsao G, Wong F, et al. Cirrhosis is associated with high mortality and readmissions over 90 days regardless of COVID-19: A multi-center cohort. Liver Transpl 2021 Jan 11. doi: 10.1002/lt.25981

[22] Safdarian AR, Momenzadeh K, Kahe F, Farhangnia P, Rezaei N. Death due to COVID-19 in a patient with diabetes, epilepsy, and gout comorbidities. Clin Case Rep 2020 Nov 25:10.1002/ccr3.3557

[23] Mahil SK, Dand N, Mason KJ, et al. Factors associated with adverse COVID-19 outcomes in patients with psoriasis-insights from a global registry-based study. J Allergy Clin Immunol 2021;147:60-71

[24] Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020;18:844-7

[25] Roy NBA, Telfer P, Eleftheriou P, et al. Protecting vulnerable patients with inherited anaemias from unnecessary death during the COVID-19 pandemic. Br J Haematol 2020;189:635-9

[26] Druss BG. Addressing the COVID-19 pandemic in populations with serious mental illness. JAMA Psychiatry 2020;77:891-2

[27] Suissa S, Patenaude V, Lapi F, Erns P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. Thorax 2013;68:1029-36

[28] Wiese AD, Griffin MR, Schaffner W, et al. Long-acting Opioid Use and the Risk of Serious Infections: A Retrospective Cohort Study. Clin Infect Dis 2019;68:1862-9

[29] Kollias A, Kyriakoulis KG, Dimakakos E, et al. Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action. Br J Haematol 2020;189:846-7

[30] Rentsch CT, Beckman JA, Tomlinson L, et al. Early initiation of prophylactic anticoagulation for prevention of COVID-19 mortality: a nationwide cohort study of hospitalized patients in the United States. medRxiv. 2020 Dec 11:2020.12.09.20246579

Table 1. Prevalence of male gender, age categories and 29 conditions/diseases contributing to the Covid Vulnerability Score (CVS). For each listed contributor, the outcome incidence among the exposed people, the odds ratio (and 90% confidence interval), and the corresponding weight of the contribution to CVS, are reported

	No. (%)	No. outcome events	Incidence every 10,000	Odds ratio†	(90% confidence interval†)	Weight‡
Male gender	3,797,636 (49.6%)	6,849	18.0	3.07	2.95 to 3.19	11
Age ≤ 45	3,111,426 (40.6%)	271	0.9	1.00	(reference)	0
Age 46-59	2,305,062 (30.1%)	1,435	6.2	5.95	(5.36 to 6.62)	18
Age 60-69	1,222,310 (16.0%)	2,506	20.5	15.62	(14.09 to 17.32)	27
Age 70-79	1,016,704 (13.3%)	4,948	48.7	27.64	(24.96 to 30.61)	33
HIV infection	31,300 (0.4%)	154	49.2	1.52	(1.33 to 1.74)	4
Other infectious and parasitic diseases	42,422 (0.6%)	443	104.4	1.37	(1.26 to 1.49)	3
Malignancies	177,024 (2.3%)	1,073	60.6	1.42	(1.35 to 1.50)	4
Diabetes without insulin therapy	278,785 (3.6%)	1,419	50.9	1.60	(1.53 to 1.68)	5
Insulin therapy	101,996 (1.3%)	973	95.4	2.35	(2.21 to 2.49)	9
Obesity	16,571 (0.2%)	103	62.2	1.34	(1.13 to 1.58)	3
Disorders of fluid, electrolyte, and acid-base balance	8,576 (0.1%)	135	157.4	1.29	(1.11 to 1.49)	3
Gout	164,428 (2.2%)	1,518	92.3	1.57	(1.50 to 1.66)	5
Coagulation defects	3,603 (0.1%)	36	99.9	1.41	(1.07 to 1.85)	3
Anaemias	613,430 (8.0%)	2,228	36.3	1.51	(1.45 to 1.58)	4
Dementia / Alzheimer	12,671 (0.2%)	145	114.4	1.26	(1.09 to 1.46)	2
Psychosis	138,034 (1.8%)	684	49.6	1.94	(1.80 to 2.08)	7
Depression	588,688 (7.7%)	1,729	29.4	1.35	(1.29 to 1.42)	3
Parkinson's disease and parkinsonism	40,885 (0.5%)	274	67.0	1.21	(1.09 to 1.34)	2
Epilepsy and recurrent seizures	122,171 (1.6%)	510	41.7	1.37	(1.26 to 1.48)	3
Other diseases of the nervous system and sense organs	35,495 (0.5%)	253	71.3	1.26	(1.13 to 1.40)	2
Ischaemic Heart Disease/Angina	91,539 (1.2%)	845	92.3	1.18	(1.11 to 1.26)	2
Heart failure	21,840 (0.3%)	428	196.0	1.30	(1.18 to 1.43)	3
Vascular diseases	14,936 (0.2%)	217	145.3	1.17	(1.04 to 1.32)	2
Cerebrovascular diseases	35,205 (0.5%)	333	94.6	1.12	(1.02 to 1.23)	1
Hypertension	796,044 (10.4%)	3,136	39.4	1.20	(1.15 to 1.25)	2

Coronary and peripheral vascular disease	658,737 (8.6%)	2,668	40.5	1.75	(1.68 to 1.82)	6
Oral anticoagulant agents	144,713 (1.9%)	1,221	84.4	1.39	(1.32 to 1.47)	3
COPD/Asthma	20,034 (0.3%)	268	133.8	1.15	(1.03 to 1.28)	1
Liver cirrhosis and other liver chronic diseases	29,484 (0.4%)	177	60.0	1.31	(1.16 to 1.49)	3
Chronic kidney disease	17,109 (0.2%)	371	216.8	1.32	(1.20 to 1.46)	3
Diseases of the skin and subcutaneous tissues	106,747 (1.4%)	353	33.1	1.10	(1.00 to 1.20)	1
Chronic pain	191,442 (2.5%)	1,007	52.6	1.28	(1.21 to 1.36)	2
Corticosteroids	935,246 (12.2%)	2,588	27.7	1.62	(1.55 to 1.68)	5
Individuals without any of the 29 conditions above listed	4,600,012 (60.1%)	1,350	2.9	-	-	-

HIV, Human immunodeficiency virus; COPD, Chronic obstructive pulmonary disease

The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 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 (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk

† Odds ratio, and 90% confidence interval, estimated by multivariable logistic regression. Odds ratios 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

Legend of Figures

Figure 1. Relationship between categories of CoViD-19 Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles

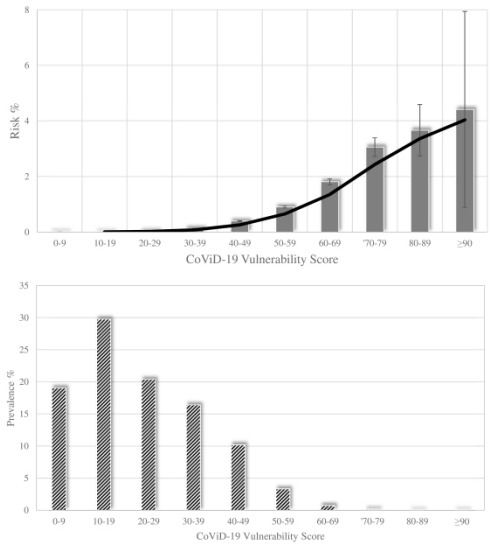
Footnote. The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk.

Figure 2. Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of CoViD-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of CoViD-19 Vulnerability Score (CVS) from several validation sets (right box)

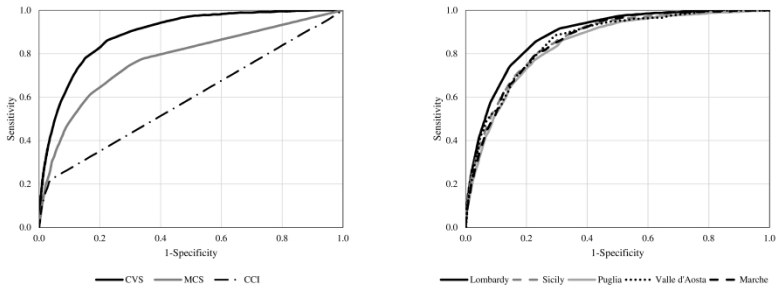
Footnote. Derivation set (left box) was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. Validation sets (right box) were based on: (i) 7,575,924 resident in Lombardy whom observation started on July 1, 2020 and who experienced 2,822 severe/fatal outcomes within December 31, 2020; (ii) 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 March 1, 2020 and who respectively experienced 173, 542, 1,953 and 1,541 severe/fatal outcomes within December 31, 2020

Figure 3. Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

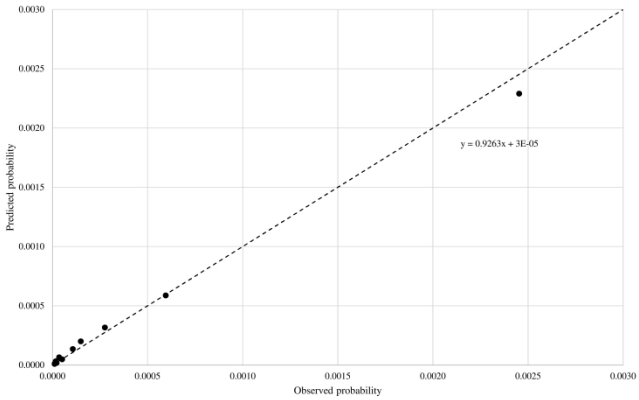
Footnote. The analysis was based on the pooled validation sets of 15,441,033 from Lombardy, Valle d'Aosta, Marche, Puglia and Sicily who experienced 7,031 severe/fatal outcomes rom starting (July 1, 2020 in Lombardy, or March 1, 2020 in the other regions) until December 31, 2020



Relationship between categories of CoViD-19 Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles



Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of CoViD-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of CoViD-19 Vulnerability Score (CVS) from several validation sets (right box)



Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

Addressing vaccination priority by stratifying general population according with frailty: the new Covid-19 vulnerability score (CVS)

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Selected features of regional populations included into the validation study in comparison with entire Italian population

		Italian 2020 population census†		Indicators of Covid-19 Epidemic Spread (March-December 2020)‡	
Region	Location	Whole population	Population aged 18 – 79 years	Ascertained cases	Deaths
Lombardy	Norther Italy	10,027,602	7,663,864	478,903	25,123
Valle d'Aoste	North Italy	125,034	95,914	7,273	379
Marche	Central Italy	1,512,672	1,150,809	41,624	1,571
Puglia	Southern Italy	3,953,305	3,055,720	90,964	2,472
Sicily	Island	4,875,290	3,744,848	93,644	2,412
	Total	20,493,903	15,711,155	712,408	31,957
	Italy	59,641,488	45,788,626	2,107,166	74,159

† source: http://demo.istat.it/popres/index.php?anno=2020&lingua=ita

‡ source: Protezione Civile. Dati COVID-19 Italia (available at https://github.com/pcm-dpc/COVID-19)

Supplementary Table S2. List of candidate conditions for predicting the risk of experiencing severe fatal forms of COVID-19 infection

Diagnostic categories	#	Disease / condition	ICD-9 CM	ATC
Infectious and parasitic diseases	1	HIV infection	042.x, V08	J05AB14, J05AE, J05AF01, J05AF02, J05AF04, J05AF05, J05AF06, J05AF09. J05AG, J05AR, J05AX07, J05AX08, J05AX09, J05AX12
	2	Tuberculosis and Other infectious and parasitic diseases	010.x - 018.x, 001.x-009.x, 020.x-027.x, 030.x-0.41.x, 045.x–057.x, 060.x-066.x, 070.x-088.x, 090.x-104.x, 110.x-118.x, 120.x-139.x	J04AB
Neoplasms	3	Solid malignancies and Neoplasm of lymphatic and haematopoietic tissue	140.x-165.x, 170.x-176.x, 179.x-199.x, V58.0, 92.2, 200.x-208-x	L01, L03AC, L02BA01, L02BA02, L02BG02, L02BG03, L02BG04, L02BG06, L02BB01, L02BB03, L02AE02, L02AE04, L02AB01
	4	Benign neoplasm and carcinoma in situ	210.x-234.x	
Endocrine, nutritional and metabolic diseases, and immunity disorders	5	Hypothyroidism	243, 244.x	H03A, H03B
	6	Hyper e hypoparathyroidism	252.0, 252.1	
	7	Diabetes without insulin therapy	250.x, 348.0x, 357.2, 362.0, 366.41	A10B
	8	Insulin therapy		A10A
	9	Dyslipidaemia	272.2, 272.4	C10
	10	Obesity	278.0x	
	11	Weight loss	260-263.x	
	12	Disorders of fluid, electrolyte, and acid-base balance	276.x	
	13	Gout	274.x	M04AC01, M04AA, M04AB
	14	Other disorders of endocrine, nutritional and metabolic diseases	240.x-242.x, 245.x, 246.x, 249.x, 251.x, 252.8, 252.9, 253.x-259.x, 264.x-269.x, 270.x, 271.x, 272.0, 272.1, 272.3, 272.5-272.9, 273.x, 275.x, 277.x, 278.1-278.8 (except 277.0)	
	15	Disorders involving the immune mechanisms	279.x	
Diseases of the blood and blood-forming organs	16	Coagulation defects	286.x	B02B
	17	Autoimmune haemolytic anaemias, Other anaemias, Anaemias only tracked from drug therapy	280.x-282.x, 283.1-283.9, 284.x-285.x	B03A, B03B, B03XA01, L03AA
	18	Other diseases of the blood and blood-forming organs	287.x-289.x	
Mental disorders	19	Dementia / Alzheimer	290.0-290.4x, 331.0x	N06DA, N06DX01

	20	Psychosis	295.x, 297.x, 298.2-298.9, 299.1x	N05AD, N05AA, N05AB, N05AC, N05AX, N05AE, N05AF, N05AG N05AH, N05AL
	21	Depression	296.2, 296.3, 296.82, 298.0, 300.4, 301.12, 309.0x, 309.1x, 311.x	N06A
	22	Bipolar disorders	296.0x, 296.1x, 296.4x, 296.5x, 296.6x, 296.7x, 296.80, 296.81, 296.89, 296.9x, 298.1x	N05AN
	23	Alcohol abuse	291.1, 291.2, 291.5, 291.8x, 291.9, 303.9, 305.0x, V11.3x	N07BB01
	24	Drug addition	292.0x, 292.82-292.89, 292.9x, 304.x, 305.2x-305.9x	N07BB04
	25	Anxiety	300.0x	N05BA, N05BB01, N05CD, N05BC01, N05BC51, N05BX, N05CF, N05CX01, N06BX
	26	Other mental disorders	290.8, 290.9, 291.0, 291.3, 291.4, 292.1x, 292.2, 292.81, 293.x, 294.x, 299.0x, 299.8x, 299.9x, 300.0x-300.2x, 300.3, 300.5-300.9, 301.0, 301.10, 301.11, 301.2x-301.9x, 302.x, 303.x, 305.1, 306.x-308.x, 309.2x-309.4x, 310.x, 312.x-319.x	
Diseases of the nervous system and sense organs	27	Parkinson's disease and parkinsonism	332.x	N04
	28	Multiple sclerosis	340	L03AB07, L03AB08, L04AA23, L04AA27, L03AX13, L04AA31, L04AA34, L03AB13, L04AX07
	29	Epilepsy and recurrent seizures	345.x	N03AF01, N03AB02, N03AA02, N03AA03, N03AA04, N03AE01, N03AD01, N03AG01, N05BA09, N03AG04, N03AX10, N03AG06, N03AF02, N03AX14, N03AX15
	30	Glaucoma	365.x	S01E
	31	Disorders of the eye and adnexa	360.x-379.x (except 365.x)	
	32	Diseases of the ear and mastoid process	380.x-389.x	
	33	Other diseases of the nervous system and sense organs	320.x-326.x, 330.x-331.x, 333.x-337.x, 340.x-344.x, 346.x-359.x	
Diseases of the circulatory system	34	Ischaemic Heart Disease/Angina	410.x – 414	C01DA, C01DX
	35	Heart failure	398.91, 402.11, 402.91, 404.11, 404.13, 404.91, 404.93, 428.x	
	36	Arrhythmia	426.10, 426.11, 426.13, 426.20-426.53, 426.60-426.89, 427.0, 427.2,	C01BA, C01BC, C01BD

			427.31, 427.60,427.9, 785.0x, V45.0x, V53.3x	
	37	Valvular diseases	093.20-093.24, 394.0x-397.1x, 424.00-424.91, 746.3x-746.6x, V42.2x, V43.3x	
	38	Vascular diseases	440.x, 441.2, 441.4, 441.7, 441.9, 443.1x-443.9x, 447.1, 557.1x, 557.9x, 785.4x, V43.4x	
	39	Cerebrovascular diseases	430.x-438.x	
	40	Hypertension	401.x-405.x	C03AA, C03AB, C03AH, C03AX01, C02CA04, C03BA02, C03BA03, C03BA04, C03BA05, C03BA07, C03BA08, C03BA09, C03BA10, C03BA11, C03DB01, C03DB02, C03EA, C09BA02, C09BA03, C09BA04, C09BA05, C09BA06, C09BA07, C09BA08, C09BA09, C09BB, C09DB, C09DA01, C09DA02, C09DA03, C09DA04, C09DA06, C09DA07, C09DA08, C02AB01, C02AB02, C02AC01, C02AC02, C02AC04, C02AC05, C02DB02, C02DB03, C02DB04, C02DC01, C02DD01, C02DG01, C02KA01, C02KB01, C02KC01, C02KD01, C02KX01, C09XA
	41	Coronary and peripheral vascular disease		B01AB, B01AX01, B01AD10, B01AD12, C04AD03, B01AC05
	42	Oral anticoagulant agents		B01AA, B01AE, B01AF
	43	Other diseases of the circulatory system	390.x-392.x, 393, 397.9, 398.90, 398.99, 411.8x, 412.x-417x, 420.x-423.x, 424.99, 425.x, 426.0, 426.12, 426.54, 426.9, 427.1, 427.32, 427.4x, 427.5, 427.61, 427.69, 427.8x, 429.x, 441.0x, 441.1, 441.3, 441.5, 441.6, 442.x, 443.0, 444.x-446.x, 447.0, 447.2-447.9, 448.x 451.x-459.x	
Diseases of the respiratory system	44	Chronic Obstructive Pulmonary Disease, Asthma, Chronic respiratory disease only tracked from drug therapy	490-492.x, 493.x, 494.x, 496	R03AA, R03AB, R03AC, R03DA, R03DB, R03DA20, R01AC01, R03BC01, R01AC51, S01GX01, S01GX51, R03BA

	45	Acute respiratory infections	460-466.x	
	46	Cystic Fibrosis	277.0	R05CB, R05FB01, R05FA01, A09AA02, R07AX02, R07AX30, R07AX31
	47	Other diseases of the respiratory system	470.x-478.x, 480.x-487.x, 495.x, 500.x-508.x, 510.x-519.x	
Diseases of the digestive system	48	Liver cirrhosis and other liver chronic diseases	571.x, 573.x	J05AP08, J05AP09, J05AP51, J05AP53, J05AP54, J05AP55, J05AP56, J05AP57, B05AA01
	49	Inflammatory bowel diseases (Ulcerative colitis and Chron's disease)	555.x-556.x	A07EC01, A07EC02, A07EC03, A07EC04
	50	Chronic and acute pancreatitis	577.0-577.1	
	51	Other diseases of the digestive system	520.x-553.x, 557.x-570, 572.x, 574.x-576.x, 577.2-577.9, 578.x, 579.x	
Diseases of the genitourinary system	52	Chronic kidney disease	585, V45.1, V56.x, V03AE	
	53	Other kidney disorders	580.x-584.x, 586, 587, 588.x-589.x	
	54	Other diseases of the genitourinary system	590.x-608.x, 610.x, 611.x, 614.x-629.x	
Diseases of the skin and subcutaneous tissues	55	Diseases of the skin and subcutaneous tissues, including No rheumatoid psoriasis	680.x-686.x, 690.x-695.x, 696.0, 696.2-696.5, 696.8, 697.x, 698.x, 700.x-709.x, 696.1	D05BB01, D05BB02, D05AX
Diseases of the musculoskeletal system and connective tissue	56	Autoimmune disease (Rheumatoid arthritis, Rheumatoid psoriasis, Anchylosing spondylitis, Systemic sclerosis, Systemic lupus erythematosus)	714.0, 696.0, 720.0, 710.1x, 710.0x	
	57	Other diseases of the musculoskeletal system and connective tissue	710.2-710.9, 711.x-713.x, 714.1x, 714.9x, 715.x-719.x, 720.1x-720.9x, 721.x-739.x	
Symptoms, signs and ill-defined conditions	58	Symptoms, signs and ill-defined conditions	780-799	
Other conditions	59	Transplantation	V42	L04AA01, L04AA02, L04AA03, L04AA04, L04AA05, L04AA06, L04AA08, L04AA09, L04AA10, L04AA11, L04AA12, L04AA14, L04AA15, L04AA16, L04AA17, L04AA18, L04AA19, L04AA21, L04AD01, L04AD02, L04AX01

60	Chronic pain	338.2, 338.4	N02AA01, N02AG01, N02AE01, N02AB03, N02AA05, N02AA55, N02AA03, N02AX06
61	Corticosteroids		H02

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Page 1-5
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			7
Objectives	3	State specific objectives, including any prespecified hypotheses			8
Methods					
Study Design	4	Present key elements of study design early in the paper			9-11
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			8

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	8-11
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	9
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			9

Bias	9	Describe any efforts to address potential sources of bias			9-11
Study size	10	Explain how the study size was arrived at			8-11
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			9,10
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			9-11
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	8-11

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	8,9
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	11,12
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			11,12
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			11,12

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			11,12
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			11,12
Discussion					
Key results	18	Summarise key results with reference to study objectives			12,13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	15,16
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			13-15

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			15
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			17
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	18

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Addressing vaccination priority by stratifying general population according with frailty: a large population-based cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-053281.R1
Article Type:	Original research
Date Submitted by the Author:	10-Oct-2021
Complete List of Authors:	Corrao, Giovanni; Università degli Studi di Milano-Bicocca Rea, Federico; University of Milan–Bicocca Carle, Flavia; Polytechnic University of Marche Scondotto, Salvatore; Department of Epidemiologic Observatory, Health Department of Sicily Allotta, Alessandra; Regional Health Authority, Department of Health Services and Epidemiological Observatory Lepore, Vito; Regional Health Agency of Puglia D'Ettorre, Antonio; Regional Health Agency of Puglia Tanzarella, Cinzia; Regional Health Agency of Puglia Vittori, Patrizia; Regional Health Authority Abena, Sabrina; Regional Health Authority Iommi, Marica; Polytechnic University of Marche Spazzafumo, Liana; Regional Health Agency of Marche Ercolanoni, Michele; Regional Welfare Service Blaco, Roberto; Regional Welfare Service Carbone, Simona; Italian Health Ministry, Department of Health Planning Giordani, Cristina; Italian Health Ministry, Department of Health Planning Manfellotto, Dario; Hospital Fatebenefratelli - AFaR, Department of Internal Medicine Galli, Massimo; University of Milan L. Sacco Hospital, Institute of Tropical and Infectious Diseases Mancia, Giuseppe; University of Milano-Bicocca, Clinical Medicine and Prevention; Ospedale San Gerardo, Clinia Medica
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, Health services research
Keywords:	COVID-19, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Addressing vaccination priority by stratifying general population according with frailty: a large population-based cohort study

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

Word count (manuscript): 3,094

Tables: 1

Figures: 3

Address for correspondence: Prof. Giovanni Corrao, Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi, 8, Edificio U7, 20126 Milano, Italy. Tel.: +39.02.64485854; E-mail: giovanni.corrao@ unimib.it

"Monitoring and Assessing care Pathways" MAP working group (Italian Health Ministry, Health Planning Dept):

- Italian Ministry of Health, Dept of Health Planning: Donata Bellentani, Simona Carbone (coordinator), Carla Ceccolini, Angela De Feo, Cristina Giordani, Rosanna Mariniello, Modesta Visca; Dept of health prevention: Natalia Magliocchetti, Giovanna Romano; External Expert: Antonio Lora, Paola Pisanti, Rinaldo Zanini.
- Polytechnic University of Marche (coordinator): Flavia Carle, Marica Iommi, Edlira Skrami.
- University of Milano-Bicocca, Laboratory of Healthcare Research & Pharmacoepidemiology: Anna Cantarutti, Giovanni Corrao, Matteo Monzio Compagnoni, Pietro Pugni, Federico Rea.
- Department of Epidemiology Lazio Region: Marina Davoli, Mirko Di Martino, Adele Lallo.
- Aosta Valley Region: Patrizia Vittori, Giuliana Vuillermin
- Campania Region: Alfonso Bernardo, Anna Frusciante
- Emilia Romagna Region: Laura Belotti, Rossana De Palma.
- Friuli Venezia Giulia Region: Andrea Di Lenarda, Marisa Prezza
- Lazio Region: Danilo Fusco, Chiara Marinacci
- Lombardy Region: Roberto Blaco, Olivia Leoni
- Marche Region: Liana Spazzafumo, Simone Pizzi
- Molise Region: Lolita Gallo
- Puglia Region: Ettore Attolini, Vito Lepore
- Sicily Region: Salvatore Scondotto, Giovanni De Luca
- Tuscany Region: Paolo Francesconi, Carla Rizzuti
- Veneto Region: Francesco Avossa, Silvia Vigna
- Research and Health Foundation (Fondazione ReS -Ricerca e Salute-): Letizia Dondi, Nello Martini, Antonella Pedrini, Carlo Piccinni
- National Agency for Regional Health Services: Mimma Cosentino, Maria Grazia Marvulli
- ANMCO (National Association of Hospital Cardiologists) Study Center: Aldo Maggioni

Abstract

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

Design. Retrospective observational cohort study.

Setting. Population-based study using the healthcare utilization database of five Italian Regions.

Participants. Beneficiaries of the National Health Service, resident in one of the five participating Regions, aged 18-79 years, and not resident in a nursing home. The model was built from the 7,655,502 beneficiaries of the Lombardy Region Health Service.

Main outcome measure. The score included gender, age and 29 conditions selected from a list of 61 candidates for independently predicting severe/fatal clinical manifestations of infection. The outcome was the severe (ICU admitted)/fatal manifestations of Covid-19 experienced during the first epidemic wave (until June 2020). CVS performance was validated by applying the model to several validation sets (populations from Lombardy, second epidemic wave, and other four Italian regions during 2020) for a total of about 15.4 million individuals and 7,031 outcomes. Predictive performance was assessed by discrimination (areas under the ROC curve) and calibration (plot of observed vs. predicted outcomes).

Results. A clear positive trend towards increasing outcome incidence as CVS increases was observed. Areas under the ROC curve of CVS ranged from 0.85 to 0.88, which were better than those of generic scores such as the Charlson Comorbidity Score (0.60). A remarkable calibration of observed and predicted outcome probability was observed.

Conclusions. A simple score from data used for public health management accurately predicted the occurrence of severe/fatal manifestations of Covid-19. Because of its performance, the use of CVS may help health decision-makers to achieve more accurate identification of high-risk citizens who need early preventive 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, may be easily drawn from electronic health databases covering beneficiaries of health systems (e.g., National Health Service, health insurance companies).
- The CVS was developed and validated on a large and unselected population of more than 15 million of Italian individuals.
- The CVS was validated across different temporal (first and second epidemic wave) and geographic (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 virus 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 Covid-19 patients at greater risk to progress to hospitalization 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 (e.g., monoclonal antibodies [1]) 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- and 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 [2-4], so opening the gate towards identifying multiple prognostic factors for the Covid-19 disease. However, although some factors have been accepted as "established" by the scientific community, their overall predictive value has not been robustly evaluated [5]. It should be in addition considered that basing the prediction 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 proposed and validated in hospital care settings [6,7], their use requiring specialized image acquisition or sophisticated laboratory examinations, which might be hardly applicable to the population at large. This implies that developing a score able to reliably predict the risk of progression of the Covid-19 disease to its most severe or lethal forms in the general population via simple and easily collectable information would be a valuable goal.

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 utilization 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) and Islands (Sicily) of Italy. 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 supplementary **Table S1**.

Data sources

All Italian citizens have equal access to healthcare services provided by the NHS. Computerized information systems on the provided services have been created within each of the 21 Italian Regions and autonomous Provinces, the related regional health care 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 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 starting of the Covid-19 pandemic, almost all Regions established, under the coordination of the National Health Institute, a population-based

registry of patients with a confirmed diagnosis of infection with the SARS-CoV-2 virus, 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, i.e. patient identification, was allowed only to the Regional Health Authority upon request from the judicial authority.

Candidate predictors

Taking into consideration the morbidity and mortality predictors reported in epidemiological studies [5,7-9], as well as comorbidity scores widely used worldwide or tuned to the Italian population (the Charlson index (CCI) [10] and the Multisource Comorbidity Score (MCS), developed for the general Italian population [11]), we identified 61 candidate predictors. Twenty-seven candidate predictors were traced from inpatients diagnostic codes, five from outpatients who were prescribed drugs, and the remaining twenty-nine 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 one 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 supplementary **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 elected to use the data from the first epidemic wave that hit Lombardy to develop the score.

We included all the NHS beneficiaries who at February 21, 2020 were resident in Lombardy since at least two years, were aged 18 to 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 hospitalization in an Intensive Care Unit or death with a Covid-19 diagnosis, up to June 30, 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, i.e., from January 1st, 2018 until February 20, 2020. The least absolute shrinkage and selection operator (LASSO) method was applied for selecting the conditions able to predict the outcome [12]. 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 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 degrees of freedom was used to represent the corresponding smoothed trends [13].

Score validation and performance

With the aim to validate the model across different temporal and geographic conditions (i.e., to assess the performance of CVS across different levels of treatment options, climatic characteristics, intensity of the epidemic spread, etc.), the score developed from the Lombardy cohort hit by the first pandemic wave (derivation set) was applied to several validation sets selected by using the same inclusion/exclusion criteria of the derivation one. One validation set consisted

of the cohort of Lombardy NHS beneficiaries who were free of Covid-19 up to July 1, 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 December 31, 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 March 1, 2020 and was censored at the outcome occurrence or at December 31, 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 (AUCs) [14]. Calibration plots displayed observed vs. predicted outcome probabilities. The Hosmer-Lemeshow goodness-of-fit test modified by Yu et al [15] was used for testing the null hypothesis of agreement between observed and predicted outcome probabilities.

Patient and Public involvement

No patients were involved in setting the research question or the outcome measures, nor were they 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 even 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, human immunodeficiency virus (HIV) infection, malignancies and anaemias were the most relevant contributors of the

outcome. However, other 19 clinical conditions (ranging across all major nosologic macrocategories) contributed to CVS.

Figure 1, upper box, 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 ≥80. Sixty-nine percent 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 1, lower box**).

Covid-19 Vulnerability Score performance

Figure 2, left box, shows that the area under the ROC curve of CVS was 0.89. This compared favourably with the AUC of the models based on scores not specifically addressing Covid-19 whose AUC values were 0.60 and 0.77 for the CCI and MCS, respectively. The 95% confidence intervals 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 2, right box**, the CVS AUC values were 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 calibration intercept close to the ideal value of 0 and 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 simple score based on demographic and clinical information derived from healthcare utilization data currently used by all Italian Regions for the management of NHS is able

to stratify NHS beneficiaries aged 18 to 79 years according to their risk to develop severe/fatal clinical manifestations of Covid-19. In a very large number of individuals from several Italian Regions the score we developed (termed Covid Vulnerability Score or CVS) exhibited a significantly better discriminating power than the CCI, i.e. the most worldwide used comorbidity score [10] which has been recently validated also for predicting mortality in Covid-19 patients hospitalised for pneumonia [16]. Furthermore, CVS outperformed a comorbidity score recently validated by our group for the general Italian population and found to outperform the CCI. This allows concluding that our CVS can reliably identify people in whom age, gender and a variety of comorbidities interact to make them frailer to the development of the severe and fatal clinical manifestations of the Covid-19 infection. This may provide a useful tool for establishing priority in the future vaccination programs for the general Italian population up to 79 years of age, with the exception of the priority given to individuals involved in specific job categories. The ongoing vaccination campaign was based in a descending fashion on age alone as well as on individually listed conditions or diseases that have shown a greater prevalence of severe or lethal Covid-19 infections in clinical studies. This tool may find a useful application for the establishment of priority access to the third dose, and to future treatment options, such as monoclonal antibodies, if their cost will be too high to allow NHS to plan an extended use.

Our study identified several prognostic factors that, in addition to age and gender, predict the severity of Covid-19 and are included among the medical illnesses and dispensed drugs retrievable in the healthcare utilization database. Consistently with a recent meta-analysis [4], diabetes (mainly when under insulin therapy), cardiovascular disease (mainly coronary and peripheral vascular disease), hypertension, malignancies, chronic respiratory and kidney diseases, dementia and obesity were all associated with the Covid-19 outcome. People with HIV [17], and those who had a recent history of severe clinical manifestations of an infectious disease, including tuberculosis [18], also showed a significant association. Additionally, and according with other

studies, we found that diseases of the neurological system (e.g., epilepsy, recurrent seizures [19] and Parkinson disease and parkinsonism [20]), of the gastrointestinal tract (e.g., liver cirrhosis and other liver chronic diseases [21]), of the metabolism (e.g., gout [22]), of the skin (e.g. psoriasis [23]), and of the blood and blood-forming organs (e.g. coagulation defects [24], anaemias [25]) contributed to the Covid-19 related clinical frailty. We also confirmed the involvement of frailty conditions related to mental disorders, such as psychosis and depression [26] as well as of recent dispensations of drugs with immunosuppressive properties (such as corticosteroids [27]) agents against chronic pain, e.g., narcotic analgesics [28]), and with an anticoagulant [29] action. This confirms the now established notion that alterations of the structure and function of virtually all organs and systems of the body may adversely affect resistance to the Covid-19 disease. It should be emphasized that the association between the severity of Covid-19 and the dispensed drugs we found in our study is not in contrast with the frequent use of some of these drugs for the treatment of Covid-19, because in our analysis previous drug therapies were searched for to further 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 and resistance to the virus [30].

Our study has implications for several aspects of the public health policy against Covid-19 to be considered in the future, in particular, as mentioned above, for the second vaccination program planned by the Italian Ministry of Health. As done in the first campaign, this program offers early cost-free priority immunization to people resident in a nursing home and those aged 80 years and older, while in people aged 79 years or below vaccination is planned in the later stages of the vaccination program possibly. The rationale for offering early vaccination to the oldest fraction of the population is strong because of the 24,575 severe/fatal cases of Covid-19 registered in

Lombardy during 2020, almost 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 have been recently reported to be 82 years, which means that search for and use of a risk score more complex than age alone in octogenarians and nonagenarians may carry a limited practical advantage. On the contrary, use of our score may offer the possibility of identifying accurately younger high-risk people to whom offer vaccination at an earlier time.

The present study has several strengths. Because in Italy a public funded healthcare system involves virtually all citizens, our sample of NHS beneficiaries had not only an extremely large size but also reflected an unselected population. It is also a strength that the Italian healthcare utilization 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. It should also be emphasized the excellent ability of CVS for discriminating individuals who will experience and those who will not experience Covid-19 related serious complications. Finally, it is also remarkable that, although built according to Lombardy data collected during the first epidemic wave (i.e. before the summer 2020), CVS performed similarly well during the second epidemic wave (i.e. after the summer 2020), in which knowledge on treatment options for Covid-19 improved, as well as in other areas of Italy, with 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 the Covid-19 complications extends across different temporal and geographic conditions.

Limitations are that the predictors of Covid-19 we searched for are restricted to those routinely collected and available in the administrative databases, which are the same for all Regions of Italy, i.e., hospital admissions and drug dispensed. Thus, education, functional status, socioeconomic information, and extra-clinical factors that can affect the prognosis of Covid-19 patients were not included. Our scoring system did not capture the severity of associated comorbidities, as well as

health services and treatments supplied by private providers. Misdiagnosis (due to poor accuracy in reporting diagnoses and comorbidities) and upcoding in hospital records (sometimes in pursuit of higher reimbursements) might have generated too conservative estimates of the CVS performance.

Finally, our approach may have failed to identify comorbidities that, albeit increase the risk of severe/fatal clinical manifestations, limit social contacts, thereby escaping infection by the SARS-CoV-2 virus. 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 (i.e., 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 simple 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 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.

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). 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.

Conflicts of Interest

Giovanni Corrao 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 to a variety of projects that were funded by pharmaceutical companies (i.e., Novartis, GSK, Roche, AMGEN and BMS). He also received honoraria as member of Advisory Board from Roche.

Giuseppe Mancia received honoraria for participation as speaker/chairman in national/international meetings from Boehringer Ingelheim, Ferrer, Medtronic, Menarini Int, Merck Serono, Recordati, and Servier.

For the remaining authors, nothing was declared.

Contributors

GC conceived the idea for this manuscript. GC, FR and FC designed the study. GC and GM drafted the manuscript. FR, AA, ADE, SA, MI and ME performed the data analysis. SS, VL, CT, PV, LS,

and RB extracted the data and authorised their utilisation. - GC, FC, SC, CG, DM, MG, and GM supervised the project. All authors assisted the results interpretation and manuscript revision. All authors read and approved the final manuscript.

Ethical approval

Under the rules of the Italian Drugs Agency (available at: http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf), retrospective studies using administrative databases do not require Ethics Committee protocol approval.

Data sharing

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.

Dissemination to participants and related patient and public communities

There are no plans to disseminate the results of the research to study participants or the relevant patient community.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have

been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

[1] Lloyd EC, Gandhi TN, Petty LA. Monoclonal Antibodies for COVID-19. JAMA. 2021 Feb 5. doi: 10.1001/jama.2021.1225. Epub ahead of print. PMID: 33544136

[2] Földi M, Farkas N, Kiss S, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. Obes Rev. 2020 Oct;21(10):e13095

[3] Mantovani A, Byrne CD, Zheng MH, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2020;30:1236-48

[4] Lippi G., Wong J., Henry B. M. Hypertension and its severity or mortality in Coronavirus Disease 2019 (COVID-19): a pooled analysis. Polish Archives of Internal Medicine 2020 doi: 10.20452/pamw.15272

[5] Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. PLoS One 2020;15:e0241955

[6] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients with COVID-19. JAMA Intern Med. 2020; 10.1001/jamainternmed.2020.2033

[7] Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328

[8] Gupta RK, Marks M, Samuels THA, et al; UCLH COVID-19 Reporting Group. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J 2020;56:2003498

[9] Ebrahimi M, Malehi AS, Rahim F. COVID-19 Patients: A Systematic Review and Meta-Analysis of Laboratory Findings, Comorbidities, and Clinical Outcomes Comparing Medical Staff versus the General Population. Osong Public Health Res Perspect 2020;11:269-79

[10] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83

[11] Corrao G, Rea F, Di Martino M, et al. Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. BMJ Open 2017;7:e019503

[12] Tibshirani R. The LASSO method for variable selection in the Cox model. Stat Med 1997;16:385-95

[13] Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. Bone Marrow Transplantation 2020;55:675-80

[14] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72

[15] Yu W, Xu W, Zhu L. A modified Hosmer–Lemeshow test for large data sets. Communications in Statistics - Theory and Methods 2017;46:11813-25

[16] Christensen DM, Strange JE, Gislason G, et al. Charlson Comorbidity Index Score and Risk of Severe Outcome and Death in Danish COVID-19 Patients. J Gen Intern Med 2020;35:2801-3

[17] Shareef MA, Bashaiwth HM, AlAkbari AO, et al. A systematic review of contemporary evidence on SARS-CoV-2 and HIV coinfection: What does it look like up to date? Avicenna J Med 2020;10:189-97

[18] Liu Y., Bi L., Chen Y. Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity. medRxiv. March 2020 doi: 10.1101/2020.03.10.20033795. 2020.03.10.20033795

[19] Cabezudo-Garcia P, Ciano-Petersen NL, Mena-Vazquez N, et al. Incidence and case fatality rate of COVID-19 in patients with active epilepsy. Neurology 2020;95:e1417-e1425

[20] Vignatelli L, Zenesini C, Belotti LMB, et al. Risk of hospitalization and death for COVID-19 in people with Parkinson's disease or parkinsonism. Mov Disord. 2020 Nov 16:10.1002/mds.28408

[21] Bajaj JS, Garcia-Tsao G, Wong F, et al. Cirrhosis is associated with high mortality and readmissions over 90 days regardless of COVID-19: A multi-center cohort. Liver Transpl 2021 Jan 11. doi: 10.1002/lt.25981

[22] Safdarian AR, Momenzadeh K, Kahe F, Farhangnia P, Rezaei N. Death due to COVID-19 in a patient with diabetes, epilepsy, and gout comorbidities. Clin Case Rep 2020 Nov 25:10.1002/ccr3.3557

[23] Mahil SK, Dand N, Mason KJ, et al. Factors associated with adverse COVID-19 outcomes in patients with psoriasis-insights from a global registry-based study. J Allergy Clin Immunol 2021;147:60-71

[24] Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020;18:844-7

[25] Roy NBA, Telfer P, Eleftheriou P, et al. Protecting vulnerable patients with inherited anaemias from unnecessary death during the COVID-19 pandemic. Br J Haematol 2020;189:635-9

[26] Druss BG. Addressing the COVID-19 pandemic in populations with serious mental illness. JAMA Psychiatry 2020;77:891-2

[27] Suissa S, Patenaude V, Lapi F, Erns P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. Thorax 2013;68:1029-36

[28] Wiese AD, Griffin MR, Schaffner W, et al. Long-acting Opioid Use and the Risk of Serious Infections: A Retrospective Cohort Study. Clin Infect Dis 2019;68:1862-9

[29] Kollias A, Kyriakoulis KG, Dimakakos E, et al. Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action. Br J Haematol 2020;189:846-7

[30] Rentsch CT, Beckman JA, Tomlinson L, et al. Early initiation of prophylactic anticoagulation for prevention of COVID-19 mortality: a nationwide cohort study of hospitalized patients in the United States. medRxiv. 2020 Dec 11:2020.12.09.20246579

Table 1. Prevalence of male gender, age categories and 29 conditions/diseases contributing to the Covid Vulnerability Score (CVS). For each listed contributor, the outcome incidence among the exposed people, the odds ratio (and 90% confidence interval), and the corresponding weight of the contribution to CVS, are reported

	No. (%)	No. outcome events	Incidence every 10,000	Odds ratio†	(90% confidence interval†)	Weight‡
Male gender	3,797,636 (49.6%)	6,849	18.0	3.07	2.95 to 3.19	11
Age ≤ 45	3,111,426 (40.6%)	271	0.9	1.00	(reference)	0
Age 46-59	2,305,062 (30.1%)	1,435	6.2	5.95	(5.36 to 6.62)	18
Age 60-69	1,222,310 (16.0%)	2,506	20.5	15.62	(14.09 to 17.32)	27
Age 70-79	1,016,704 (13.3%)	4,948	48.7	27.64	(24.96 to 30.61)	33
HIV infection	31,300 (0.4%)	154	49.2	1.52	(1.33 to 1.74)	4
Other infectious and parasitic diseases	42,422 (0.6%)	443	104.4	1.37	(1.26 to 1.49)	3
Malignancies	177,024 (2.3%)	1,073	60.6	1.42	(1.35 to 1.50)	4
Diabetes without insulin therapy	278,785 (3.6%)	1,419	50.9	1.60	(1.53 to 1.68)	5
Insulin therapy	101,996 (1.3%)	973	95.4	2.35	(2.21 to 2.49)	9
Obesity	16,571 (0.2%)	103	62.2	1.34	(1.13 to 1.58)	3
Disorders of fluid, electrolyte, and acid-base balance	8,576 (0.1%)	135	157.4	1.29	(1.11 to 1.49)	3
Gout	164,428 (2.2%)	1,518	92.3	1.57	(1.50 to 1.66)	5
Coagulation defects	3,603 (0.1%)	36	99.9	1.41	(1.07 to 1.85)	3
Anaemias	613,430 (8.0%)	2,228	36.3	1.51	(1.45 to 1.58)	4
Dementia / Alzheimer	12,671 (0.2%)	145	114.4	1.26	(1.09 to 1.46)	2
Psychosis	138,034 (1.8%)	684	49.6	1.94	(1.80 to 2.08)	7
Depression	588,688 (7.7%)	1,729	29.4	1.35	(1.29 to 1.42)	3
Parkinson's disease and parkinsonism	40,885 (0.5%)	274	67.0	1.21	(1.09 to 1.34)	2
Epilepsy and recurrent seizures	122,171 (1.6%)	510	41.7	1.37	(1.26 to 1.48)	3
Other diseases of the nervous system and sense organs	35,495 (0.5%)	253	71.3	1.26	(1.13 to 1.40)	2
Ischaemic Heart Disease/Angina	91,539 (1.2%)	845	92.3	1.18	(1.11 to 1.26)	2
Heart failure	21,840 (0.3%)	428	196.0	1.30	(1.18 to 1.43)	3
Vascular diseases	14,936 (0.2%)	217	145.3	1.17	(1.04 to 1.32)	2
Cerebrovascular diseases	35,205 (0.5%)	333	94.6	1.12	(1.02 to 1.23)	1
Hypertension	796,044 (10.4%)	3,136	39.4	1.20	(1.15 to 1.25)	2

Coronary and peripheral vascular disease	658,737 (8.6%)	2,668	40.5	1.75	(1.68 to 1.82)	6
Oral anticoagulant agents	144,713 (1.9%)	1,221	84.4	1.39	(1.32 to 1.47)	3
COPD/Asthma	20,034 (0.3%)	268	133.8	1.15	(1.03 to 1.28)	1
Liver cirrhosis and other liver chronic diseases	29,484 (0.4%)	177	60.0	1.31	(1.16 to 1.49)	3
Chronic kidney disease	17,109 (0.2%)	371	216.8	1.32	(1.20 to 1.46)	3
Diseases of the skin and subcutaneous tissues	106,747 (1.4%)	353	33.1	1.10	(1.00 to 1.20)	1
Chronic pain	191,442 (2.5%)	1,007	52.6	1.28	(1.21 to 1.36)	2
Corticosteroids	935,246 (12.2%)	2,588	27.7	1.62	(1.55 to 1.68)	5
Individuals without any of the 29 conditions above listed	4,600,012 (60.1%)	1,350	2.9	-	-	-

HIV, Human immunodeficiency virus; COPD, Chronic obstructive pulmonary disease

The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 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 (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk

† Odds ratio, and 90% confidence interval, estimated by multivariable logistic regression. Odds ratios 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

Legend of Figures

Figure 1. Relationship between categories of CoViD-19 Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles

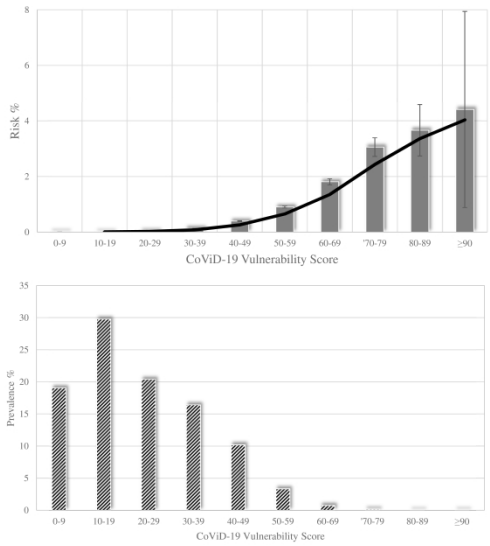
Footnote. The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk.

Figure 2. Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of CoViD-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of CoViD-19 Vulnerability Score (CVS) from several validation sets (right box)

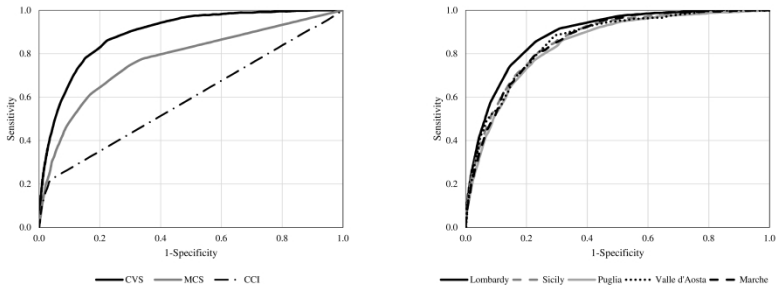
Footnote. Derivation set (left box) was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (ICU admitted and mechanically ventilated via intubation) and / or fatal outcomes. Validation sets (right box) were based on: (i) 7,575,924 resident in Lombardy whom observation started on July 1, 2020 and who experienced 2,822 severe/fatal outcomes within December 31, 2020; (ii) 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 March 1, 2020 and who respectively experienced 173, 542, 1,953 and 1,541 severe/fatal outcomes within December 31, 2020

Figure 3. Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

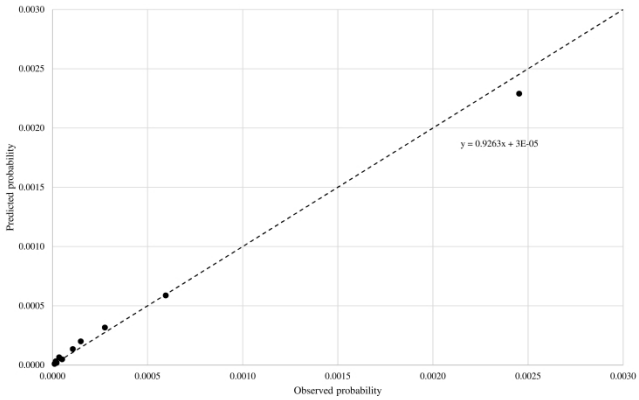
Footnote. The analysis was based on the pooled validation sets of 15,441,033 from Lombardy, Valle d'Aosta, Marche, Puglia and Sicily who experienced 7,031 severe/fatal outcomes rom starting (July 1, 2020 in Lombardy, or March 1, 2020 in the other regions) until December 31, 2020



Relationship between categories of CoViD-19 Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles



Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of CoViD-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of CoViD-19 Vulnerability Score (CVS) from several validation sets (right box)



Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

Addressing vaccination priority by stratifying general population according with frailty: a large population-based cohort study

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Selected features of regional populations included into the validation study in comparison with entire Italian population

		Italian 2020 population census†		Indicators of Covid-19 Epidemic Spread (March-December 2020)‡	
Region	Location	Whole population	Population aged 18 – 79 years	Ascertained cases	Deaths
Lombardy	Norther Italy	10,027,602	7,663,864	478,903	25,123
Valle d'Aoste	North Italy	125,034	95,914	7,273	379
Marche	Central Italy	1,512,672	1,150,809	41,624	1,571
Puglia	Southern Italy	3,953,305	3,055,720	90,964	2,472
Sicily	Island	4,875,290	3,744,848	93,644	2,412
	Total	20,493,903	15,711,155	712,408	31,957
	Italy	59,641,488	45,788,626	2,107,166	74,159

† source: http://demo.istat.it/popres/index.php?anno=2020&lingua=ita

‡ source: Protezione Civile. Dati COVID-19 Italia (available at https://github.com/pcm-dpc/COVID-19)

Supplementary Table S2. List of candidate conditions for predicting the risk of experiencing severe fatal forms of COVID-19 infection

Diagnostic categories	#	Disease / condition	ICD-9 CM	ATC
Infectious and parasitic diseases	1	HIV infection	042.x, V08	J05AB14, J05AE, J05AF01, J05AF02, J05AF04, J05AF05, J05AF06, J05AF09. J05AG, J05AR, J05AX07, J05AX08, J05AX09, J05AX12
	2	Tuberculosis and Other infectious and parasitic diseases	010.x - 018.x, 001.x-009.x, 020.x-027.x, 030.x-0.41.x, 045.x–057.x, 060.x-066.x, 070.x-088.x, 090.x-104.x, 110.x-118.x, 120.x-139.x	J04AB
Neoplasms	3	Solid malignancies and Neoplasm of lymphatic and haematopoietic tissue	140.x-165.x, 170.x-176.x, 179.x-199.x, V58.0, 92.2, 200.x-208-x	L01, L03AC, L02BA01, L02BA02, L02BG02, L02BG03, L02BG04, L02BG06, L02BB01, L02BB03, L02AE02, L02AE04, L02AB01
	4	Benign neoplasm and carcinoma in situ	210.x-234.x	
Endocrine, nutritional and metabolic diseases, and immunity disorders	5	Hypothyroidism	243, 244.x	H03A, H03B
	6	Hyper e hypoparathyroidism	252.0, 252.1	
	7	Diabetes without insulin therapy	250.x, 348.0x, 357.2, 362.0, 366.41	A10B
	8	Insulin therapy		A10A
	9	Dyslipidaemia	272.2, 272.4	C10
	10	Obesity	278.0x	
	11	Weight loss	260-263.x	
	12	Disorders of fluid, electrolyte, and acid-base balance	276.x	
	13	Gout	274.x	M04AC01, M04AA, M04AB
	14	Other disorders of endocrine, nutritional and metabolic diseases	240.x-242.x, 245.x, 246.x, 249.x, 251.x, 252.8, 252.9, 253.x-259.x, 264.x-269.x, 270.x, 271.x, 272.0, 272.1, 272.3, 272.5-272.9, 273.x, 275.x, 277.x, 278.1-278.8 (except 277.0)	
	15	Disorders involving the immune mechanisms	279.x	
Diseases of the blood and blood-forming organs	16	Coagulation defects	286.x	B02B
	17	Autoimmune haemolytic anaemias, Other anaemias, Anaemias only tracked from drug therapy	280.x-282.x, 283.1-283.9, 284.x-285.x	B03A, B03B, B03XA01, L03AA
	18	Other diseases of the blood and blood-forming organs	287.x-289.x	
Mental disorders	19	Dementia / Alzheimer	290.0-290.4x, 331.0x	N06DA, N06DX01

	20	Psychosis	295.x, 297.x, 298.2-298.9, 299.1x	N05AD, N05AA, N05AB, N05AC, N05AX, N05AE, N05AF, N05AG N05AH, N05AL
	21	Depression	296.2, 296.3, 296.82, 298.0, 300.4, 301.12, 309.0x, 309.1x, 311.x	N06A
	22	Bipolar disorders	296.0x, 296.1x, 296.4x, 296.5x, 296.6x, 296.7x, 296.80, 296.81, 296.89, 296.9x, 298.1x	N05AN
	23	Alcohol abuse	291.1, 291.2, 291.5, 291.8x, 291.9, 303.9, 305.0x, V11.3x	N07BB01
	24	Drug addition	292.0x, 292.82-292.89, 292.9x, 304.x, 305.2x-305.9x	N07BB04
	25	Anxiety	300.0x	N05BA, N05BB01, N05CD, N05BC01, N05BC51, N05BX, N05CF, N05CX01, N06BX
	26	Other mental disorders	290.8, 290.9, 291.0, 291.3, 291.4, 292.1x, 292.2, 292.81, 293.x, 294.x, 299.0x, 299.8x, 299.9x, 300.0x-300.2x, 300.3, 300.5-300.9, 301.0, 301.10, 301.11, 301.2x-301.9x, 302.x, 303.x, 305.1, 306.x-308.x, 309.2x-309.4x, 310.x, 312.x-319.x	
Diseases of the nervous system and sense organs	27	Parkinson's disease and parkinsonism	332.x	N04
	28	Multiple sclerosis	340	L03AB07, L03AB08, L04AA23, L04AA27, L03AX13, L04AA31, L04AA34, L03AB13, L04AX07
	29	Epilepsy and recurrent seizures	345.x	N03AF01, N03AB02, N03AA02, N03AA03, N03AA04, N03AE01, N03AD01, N03AG01, N05BA09, N03AG04, N03AX10, N03AG06, N03AF02, N03AX14, N03AX15
	30	Glaucoma	365.x	S01E
	31	Disorders of the eye and adnexa	360.x-379.x (except 365.x)	
	32	Diseases of the ear and mastoid process	380.x-389.x	
	33	Other diseases of the nervous system and sense organs	320.x-326.x, 330.x-331.x, 333.x-337.x, 340.x-344.x, 346.x-359.x	
Diseases of the circulatory system	34	Ischaemic Heart Disease/Angina	410.x – 414	C01DA, C01DX
	35	Heart failure	398.91, 402.11, 402.91, 404.11, 404.13, 404.91, 404.93, 428.x	
	36	Arrhythmia	426.10, 426.11, 426.13, 426.20-426.53, 426.60-426.89, 427.0, 427.2,	C01BA, C01BC, C01BD

			427.31, 427.60,427.9, 785.0x, V45.0x, V53.3x	
	37	Valvular diseases	093.20-093.24, 394.0x-397.1x, 424.00-424.91, 746.3x-746.6x, V42.2x, V43.3x	
	38	Vascular diseases	440.x, 441.2, 441.4, 441.7, 441.9, 443.1x-443.9x, 447.1, 557.1x, 557.9x, 785.4x, V43.4x	
	39	Cerebrovascular diseases	430.x-438.x	
	40	Hypertension	401.x-405.x	C03AA, C03AB, C03AH, C03AX01, C02CA04, C03BA02, C03BA03, C03BA04, C03BA05, C03BA07, C03BA08, C03BA09, C03BA10, C03BA11, C03DB01, C03DB02, C03EA, C09BA02, C09BA03, C09BA04, C09BA05, C09BA06, C09BA07, C09BA08, C09BA09, C09BB, C09DB, C09DA01, C09DA02, C09DA03, C09DA04, C09DA06, C09DA07, C09DA08, C02AB01, C02AB02, C02AC01, C02AC02, C02AC04, C02AC05, C02DB02, C02DB03, C02DB04, C02DC01, C02DD01, C02DG01, C02KA01, C02KB01, C02KC01, C02KD01, C02KX01, C09XA
	41	Coronary and peripheral vascular disease		B01AB, B01AX01, B01AD10, B01AD12, C04AD03, B01AC05
	42	Oral anticoagulant agents		B01AA, B01AE, B01AF
	43	Other diseases of the circulatory system	390.x-392.x, 393, 397.9, 398.90, 398.99, 411.8x, 412.x-417x, 420.x-423.x, 424.99, 425.x, 426.0, 426.12, 426.54, 426.9, 427.1, 427.32, 427.4x, 427.5, 427.61, 427.69, 427.8x, 429.x, 441.0x, 441.1, 441.3, 441.5, 441.6, 442.x, 443.0, 444.x-446.x, 447.0, 447.2-447.9, 448.x 451.x-459.x	
Diseases of the respiratory system	44	Chronic Obstructive Pulmonary Disease, Asthma, Chronic respiratory disease only tracked from drug therapy	490-492.x, 493.x, 494.x, 496	R03AA, R03AB, R03AC, R03DA, R03DB, R03DA20, R01AC01, R03BC01, R01AC51, S01GX01, S01GX51, R03BA

	45	Acute respiratory infections	460-466.x	
	46	Cystic Fibrosis	277.0	R05CB, R05FB01, R05FA01, A09AA02, R07AX02, R07AX30, R07AX31
	47	Other diseases of the respiratory system	470.x-478.x, 480.x-487.x, 495.x, 500.x-508.x, 510.x-519.x	
Diseases of the digestive system	48	Liver cirrhosis and other liver chronic diseases	571.x, 573.x	J05AP08, J05AP09, J05AP51, J05AP53, J05AP54, J05AP55, J05AP56, J05AP57, B05AA01
	49	Inflammatory bowel diseases (Ulcerative colitis and Chron's disease)	555.x-556.x	A07EC01, A07EC02, A07EC03, A07EC04
	50	Chronic and acute pancreatitis	577.0-577.1	
	51	Other diseases of the digestive system	520.x-553.x, 557.x-570, 572.x, 574.x-576.x, 577.2-577.9, 578.x, 579.x	
Diseases of the genitourinary system	52	Chronic kidney disease	585, V45.1, V56.x, V03AE	
	53	Other kidney disorders	580.x-584.x, 586, 587, 588.x-589.x	
	54	Other diseases of the genitourinary system	590.x-608.x, 610.x, 611.x, 614.x-629.x	
Diseases of the skin and subcutaneous tissues	55	Diseases of the skin and subcutaneous tissues, including No rheumatoid psoriasis	680.x-686.x, 690.x-695.x, 696.0, 696.2-696.5, 696.8, 697.x, 698.x, 700.x-709.x, 696.1	D05BB01, D05BB02, D05AX
Diseases of the musculoskeletal system and connective tissue	56	Autoimmune disease (Rheumatoid arthritis, Rheumatoid psoriasis, Anchylosing spondylitis, Systemic sclerosis, Systemic lupus erythematosus)	714.0, 696.0, 720.0, 710.1x, 710.0x	
	57	Other diseases of the musculoskeletal system and connective tissue	710.2-710.9, 711.x-713.x, 714.1x, 714.9x, 715.x-719.x, 720.1x-720.9x, 721.x-739.x	
Symptoms, signs and ill-defined conditions	58	Symptoms, signs and ill-defined conditions	780-799	
Other conditions	59	Transplantation	V42	L04AA01, L04AA02, L04AA03, L04AA04, L04AA05, L04AA06, L04AA08, L04AA09, L04AA10, L04AA11, L04AA12, L04AA14, L04AA15, L04AA16, L04AA17, L04AA18, L04AA19, L04AA21, L04AD01, L04AD02, L04AX01

60	Chronic pain	338.2, 338.4	N02AA01, N02AG01, N02AE01, N02AB03, N02AA05, N02AA55, N02AA03, N02AX06
61	Corticosteroids		H02

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Page 1-5
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			7
Objectives	3	State specific objectives, including any prespecified hypotheses			8
Methods					
Study Design	4	Present key elements of study design early in the paper			9-11
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			8

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	8-11
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	9
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			9

Bias	9	Describe any efforts to address potential sources of bias			9-11
Study size	10	Explain how the study size was arrived at			8-11
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			9,10
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			9-11
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	8-11

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	8,9
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	11,12
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			11,12
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			11,12

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			11,12
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			11,12
Discussion					
Key results	18	Summarise key results with reference to study objectives			12,13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	15,16
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			13-15

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			15
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			17
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	18

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Stratification of the risk of developing severe or lethal Covid-19 by a new score from a large Italian population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-053281.R2
Article Type:	Original research
Date Submitted by the Author:	21-Oct-2021
Complete List of Authors:	Corrao, Giovanni; Università degli Studi di Milano-Bicocca Rea, Federico; University of Milan–Bicocca Carle, Flavia; Polytechnic University of Marche Scondotto, Salvatore; Department of Epidemiologic Observatory, Health Department of Sicily Allotta, Alessandra; Regional Health Authority, Department of Health Services and Epidemiological Observatory Lepore, Vito; Regional Health Agency of Puglia D'Ettorre, Antonio; Regional Health Agency of Puglia Tanzarella, Cinzia; Regional Health Agency of Puglia Vittori, Patrizia; Regional Health Authority Abena, Sabrina; Regional Health Authority Iommi, Marica; Polytechnic University of Marche Spazzafumo, Liana; Regional Health Agency of Marche Ercolanoni, Michele; Regional Welfare Service Blaco, Roberto; Regional Welfare Service Carbone, Simona; Italian Health Ministry, Department of Health Planning Giordani, Cristina; Italian Health Ministry, Department of Health Planning Manfellotto, Dario; Hospital Fatebenefratelli - AFaR, Department of Internal Medicine Galli, Massimo; University of Milan L. Sacco Hospital, Institute of Tropical and Infectious Diseases Mancia, Giuseppe; University of Milano-Bicocca, Clinical Medicine and Prevention; Ospedale San Gerardo, Clinia Medica
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, Health services research
Keywords:	COVID-19, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Stratification of the risk of developing severe or lethal Covid-19 by a new score from a large Italian population

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

Word count (manuscript): 3,165

Tables: 1

Figures: 3

Address for correspondence: Prof. Giovanni Corrao, Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi, 8, Edificio U7, 20126 Milano, Italy. Tel.: +39.02.64485854; E-mail: giovanni.corrao@ unimib.it

"Monitoring and Assessing care Pathways" MAP working group (Italian Health Ministry, Health Planning Dept):

- Italian Ministry of Health, Dept of Health Planning: Donata Bellentani, Simona Carbone (coordinator), Carla Ceccolini, Angela De Feo, Cristina Giordani, Rosanna Mariniello, Modesta Visca; Dept of health prevention: Natalia Magliocchetti, Giovanna Romano; External Expert: Antonio Lora, Paola Pisanti, Rinaldo Zanini.
- Polytechnic University of Marche (coordinator): Flavia Carle, Marica Iommi, Edlira Skrami.
- University of Milano-Bicocca, Laboratory of Healthcare Research & Pharmacoepidemiology: Anna Cantarutti, Giovanni Corrao, Matteo Monzio Compagnoni, Pietro Pugni, Federico Rea.
- Department of Epidemiology Lazio Region: Marina Davoli, Mirko Di Martino, Adele Lallo.
- Aosta Valley Region: Patrizia Vittori, Giuliana Vuillermin
- Campania Region: Alfonso Bernardo, Anna Frusciante
- Emilia Romagna Region: Laura Belotti, Rossana De Palma.
- Friuli Venezia Giulia Region: Andrea Di Lenarda, Marisa Prezza
- Lazio Region: Danilo Fusco, Chiara Marinacci
- Lombardy Region: Roberto Blaco, Olivia Leoni
- Marche Region: Liana Spazzafumo, Simone Pizzi
- Molise Region: Lolita Gallo
- Puglia Region: Ettore Attolini, Vito Lepore
- Sicily Region: Salvatore Scondotto, Giovanni De Luca
- Tuscany Region: Paolo Francesconi, Carla Rizzuti
- Veneto Region: Francesco Avossa, Silvia Vigna
- Research and Health Foundation (Fondazione ReS -Ricerca e Salute-): Letizia Dondi, Nello Martini, Antonella Pedrini, Carlo Piccinni
- National Agency for Regional Health Services: Mimma Cosentino, Maria Grazia Marvulli
- ANMCO (National Association of Hospital Cardiologists) Study Center: Aldo Maggioni

Abstract

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

Design. Retrospective observational cohort study.

Setting. Population-based study using the healthcare utilization 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, i.e., 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, i.e. Lombardy population (second epidemic wave), and the other four Italian regions (entire 2020) for a total of about 15.4 million individuals and 7,031 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-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-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 geographic (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 virus 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 Covid-19 patients at greater risk to progress to hospitalization 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 (e.g., monoclonal antibodies [1]) 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- and 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 [2-4], which potentially helps to identify the multiple prognostic factors that are involved in the Covid-19 disease. However, although some factors have been accepted as “established” by the scientific community, their overall predictive value has not been robustly evaluated [5]. 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 [6,7], their use requiring specialized 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 utilization 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 supplementary **Table S1**.

Data sources

All Italian citizens have equal access to healthcare services provided by the NHS. Computerized information systems on the provided services have been created within each of the 21 Italian regions and autonomous Provinces, the related regional health care 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, 9th 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 starting 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 the SARS-CoV-2 virus, 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, i.e. 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 [5,7-9], as well as comorbidity scores widely used worldwide or tuned to the Italian population (the Charlson comorbidity index [10] and the Multisource Comorbidity Score (MCS), developed for the general Italian population [11]), we identified 61 candidate predictors. Twenty-seven candidate predictors were traced from inpatients diagnostic codes, five from outpatients who were prescribed drugs, and the remaining twenty-nine 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 one 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 supplementary **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 February 21, 2020 were resident in Lombardy for at least two years, were aged 18 to 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 hospitalization in an Intensive Care Unit or death with a Covid-19 diagnosis, up to June 30, 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, i.e., from January 1st, 2018 until February 20, 2020. The least absolute shrinkage and selection operator (LASSO) method was applied for selecting the conditions able to predict the outcome [12]. 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-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 degrees of freedom was used to represent the corresponding smoothed trends [13].

Score validation and performance

To validate the model across different temporal and geographic conditions (i.e., 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 July 1, 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

December 31th, 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 March 1th, 2020 and were censored at the outcome occurrence or at December 31, 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 (AUCs) [14]. Calibration plots displayed observed vs. predicted outcome probabilities. The Hosmer-Lemeshow goodness-of-fit test modified by Yu et al [15] 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-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, human immunodeficiency virus (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 1, **upper box**, 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 ≥80. Sixty-nine percent 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 1**, **lower box**).

Covid-Vulnerability Score performance

Figure 2, **left box**, shows that the area under the ROC curve 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% confidence intervals 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 2**, **right box**, the CVS AUC values were almost superimposable between the different regions participating in the study, i.e., 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 utilization data currently used throughout Italy for the management of NHS is able to stratify NHS beneficiaries aged 18 to 79 years for their risk to develop severe/fatal clinical manifestations of Covid-19. The score (developed in a very large number of individuals from

several Italian regions) exhibited a significantly better discriminating power than the Charlson comorbidity index, i.e. the most worldwide used comorbidity score [10] which has been recently validated also for predicting mortality in Covid-19 patients hospitalised for pneumonia [16]. It also outperformed a comorbidity score validated by our group for the general Italian population and also found to be better than the Charlson comorbidity index. This allows to conclude that the score we developed (termed Covid-Vulnerability Score or CVS) can reliably identify people in whom age, gender and a variety of comorbidities interact to make them more at risk for the clinically severe and fatal manifestations of SARS-CoV-2 infection. This makes CVS a potentially useful tool for establishing priority in the future vaccination programs for the general Italian population up to 79 years of age which has so far been based in a descending fashion on age alone as well as on individually listed conditions or diseases that have shown a greater prevalence of severe or lethal Covid-19 in clinical studies. CVS may also find a useful future application to the determination of priority access to the third dose of vaccine, or to the delivery of future treatment options, such as new antiviral agents and monoclonal antibodies, if their cost will be too high to allow an extended use.

Our study identified several prognostic factors that, in addition to age and gender, predict the severity of Covid-19 and are included among the medical illnesses and dispensed drugs retrievable in the healthcare utilization database. Consistently with a recent meta-analysis [4], diabetes (mainly when under insulin therapy), cardiovascular disease (mainly coronary and peripheral vascular disease), hypertension, malignancies, chronic respiratory and kidney diseases, dementia and obesity were all associated with the Covid-19 outcome. People with HIV [17], and those who had a recent history of severe clinical manifestations of an infectious disease, including tuberculosis [18], also showed a significant association with the severity of Covid-19. Additionally, and according to other studies, we found that diseases of the neurological system (e.g., epilepsy, recurrent seizures [19] and Parkinson disease and parkinsonism [20]), of the

gastrointestinal tract (e.g., liver cirrhosis and other liver chronic diseases [21]), of metabolism (e.g., gout [22]), of the skin (e.g. psoriasis [23]), and of the blood and blood-forming organs (e.g. coagulation defects [24] and anaemias [25]) contributed to the Covid-19 related clinical frailty. We also confirmed the involvement in a greater risk of severe or lethal forms of Covid-19 of mental disorders, such as psychosis and depression [26] as well as of recent dispensations of drugs with immunosuppressive properties (e.g., corticosteroids [27]) agents against chronic pain (e.g., narcotic analgesics [28]), or with an anticoagulant [29] action. This confirms the now established notion that alterations of the structure and function of virtually all organs and systems of the body may adversely affect resistance to the Covid-19 disease. It should be emphasized 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 [30].

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 offers an early cost-free priority third dose to people resident 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 have 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 program 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 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 utilization 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 (i.e. before the summer 2020), CVS performed similarly well during the second epidemic wave (i.e. after the summer 2020), despite differences in treatment options for in and outpatients as well as hospitalization criteria compared to 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 Covid-19 complications extends across different temporal and geographic 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), i.e., hospital admissions and drug dispensed. Thus, educational factors, functional status, socioeconomic characteristics, and other extra-clinical variables that can affect the prognosis of Covid-19 patients 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 up-coding of hospital records.

Finally, our approach may have failed to identify comorbidities that, albeit 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 (i.e., 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.

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). 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.

Conflicts of Interest

Giovanni Corrao 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 to a variety of projects that were funded by pharmaceutical companies (i.e., Novartis, GSK, Roche, AMGEN and BMS). He also received honoraria as member of Advisory Board from Roche.

Giuseppe Mancia received honoraria for participation as speaker/chairman in national/international meetings from Boehringer Ingelheim, Ferrer, Medtronic, Menarini Int, Merck Serono, Recordati, and Servier.

For the remaining authors, nothing was declared.

Contributors

GC conceived the idea for this manuscript. GC, FR and FC designed the study. GC and GM drafted the manuscript. FR, AA, ADE, SA, MI and ME performed the data analysis. SS, VL, CT, PV, LS,

and RB extracted the data and authorised their utilisation. - GC, FC, SC, CG, DM, MG, and GM supervised the project. All authors assisted the results interpretation and manuscript revision. All authors read and approved the final manuscript.

Ethical approval

Under the rules of the Italian Drugs Agency (available at: http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf), retrospective studies using administrative databases do not require Ethics Committee protocol approval.

Data sharing

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.

Dissemination to participants and related patient and public communities

There are no plans to disseminate the results of the research to study participants or the relevant patient community.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have

been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

[1] Lloyd EC, Gandhi TN, Petty LA. Monoclonal Antibodies for COVID-19. JAMA. 2021 Feb 5. doi: 10.1001/jama.2021.1225. Epub ahead of print. PMID: 33544136

[2] Földi M, Farkas N, Kiss S, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. Obes Rev. 2020 Oct;21(10):e13095

[3] Mantovani A, Byrne CD, Zheng MH, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2020;30:1236-48

[4] Lippi G., Wong J., Henry B. M. Hypertension and its severity or mortality in Coronavirus Disease 2019 (COVID-19): a pooled analysis. Polish Archives of Internal Medicine 2020 doi: 10.20452/pamw.15272

[5] Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. PLoS One 2020;15:e0241955

[6] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients with COVID-19. JAMA Intern Med. 2020; 10.1001/jamainternmed.2020.2033

[7] Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328

[8] Gupta RK, Marks M, Samuels THA, et al; UCLH COVID-19 Reporting Group. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J 2020;56:2003498

[9] Ebrahimi M, Malehi AS, Rahim F. COVID-19 Patients: A Systematic Review and Meta-Analysis of Laboratory Findings, Comorbidities, and Clinical Outcomes Comparing Medical Staff versus the General Population. Osong Public Health Res Perspect 2020;11:269-79

[10] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83

[11] Corrao G, Rea F, Di Martino M, et al. Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. BMJ Open 2017;7:e019503

[12] Tibshirani R. The LASSO method for variable selection in the Cox model. Stat Med 1997;16:385-95

[13] Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. Bone Marrow Transplantation 2020;55:675-80

[14] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72

[15] Yu W, Xu W, Zhu L. A modified Hosmer–Lemeshow test for large data sets. Communications in Statistics - Theory and Methods 2017;46:11813-25

[16] Christensen DM, Strange JE, Gislason G, et al. Charlson Comorbidity Index Score and Risk of Severe Outcome and Death in Danish COVID-19 Patients. J Gen Intern Med 2020;35:2801-3

[17] Shareef MA, Bashaiwth HM, AlAkbari AO, et al. A systematic review of contemporary evidence on SARS-CoV-2 and HIV coinfection: What does it look like up to date? Avicenna J Med 2020;10:189-97

[18] Liu Y., Bi L., Chen Y. Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity. medRxiv. March 2020 doi: 10.1101/2020.03.10.20033795. 2020.03.10.20033795

[19] Cabezudo-Garcia P, Ciano-Petersen NL, Mena-Vazquez N, et al. Incidence and case fatality rate of COVID-19 in patients with active epilepsy. Neurology 2020;95:e1417-e1425

[20] Vignatelli L, Zenesini C, Belotti LMB, et al. Risk of hospitalization and death for COVID-19 in people with Parkinson's disease or parkinsonism. Mov Disord. 2020 Nov 16:10.1002/mds.28408

[21] Bajaj JS, Garcia-Tsao G, Wong F, et al. Cirrhosis is associated with high mortality and readmissions over 90 days regardless of COVID-19: A multi-center cohort. Liver Transpl 2021 Jan 11. doi: 10.1002/lt.25981

[22] Safdarian AR, Momenzadeh K, Kahe F, Farhangnia P, Rezaei N. Death due to COVID-19 in a patient with diabetes, epilepsy, and gout comorbidities. Clin Case Rep 2020 Nov 25:10.1002/ccr3.3557

[23] Mahil SK, Dand N, Mason KJ, et al. Factors associated with adverse COVID-19 outcomes in patients with psoriasis-insights from a global registry-based study. J Allergy Clin Immunol 2021;147:60-71

[24] Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020;18:844-7

[25] Roy NBA, Telfer P, Eleftheriou P, et al. Protecting vulnerable patients with inherited anaemias from unnecessary death during the COVID-19 pandemic. Br J Haematol 2020;189:635-9

[26] Druss BG. Addressing the COVID-19 pandemic in populations with serious mental illness. JAMA Psychiatry 2020;77:891-2

[27] Suissa S, Patenaude V, Lapi F, Erns P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. Thorax 2013;68:1029-36

[28] Wiese AD, Griffin MR, Schaffner W, et al. Long-acting Opioid Use and the Risk of Serious Infections: A Retrospective Cohort Study. Clin Infect Dis 2019;68:1862-9

[29] Kollias A, Kyriakoulis KG, Dimakakos E, et al. Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action. Br J Haematol 2020;189:846-7

[30] Rentsch CT, Beckman JA, Tomlinson L, et al. Early initiation of prophylactic anticoagulation for prevention of COVID-19 mortality: a nationwide cohort study of hospitalized patients in the United States. medRxiv. 2020 Dec 11:2020.12.09.20246579

Table 1. Prevalence of male gender, age categories and 29 conditions/diseases contributing to the Covid-Vulnerability Score (CVS). For each listed contributor, the outcome incidence among the exposed people, the odds ratio (and 90% confidence interval), and the corresponding weight of the contribution to CVS, are reported

	No. (%)	No. outcome events	Incidence every 10,000	Odds ratio†	(90% confidence interval†)	Weight‡
Male gender	3,797,636 (49.6%)	6,849	18.0	3.07	2.95 to 3.19	11
Age ≤ 45	3,111,426 (40.6%)	271	0.9	1.00	(reference)	0
Age 46-59	2,305,062 (30.1%)	1,435	6.2	5.95	(5.36 to 6.62)	18
Age 60-69	1,222,310 (16.0%)	2,506	20.5	15.62	(14.09 to 17.32)	27
Age 70-79	1,016,704 (13.3%)	4,948	48.7	27.64	(24.96 to 30.61)	33
HIV infection	31,300 (0.4%)	154	49.2	1.52	(1.33 to 1.74)	4
Other infectious and parasitic diseases	42,422 (0.6%)	443	104.4	1.37	(1.26 to 1.49)	3
Malignancies	177,024 (2.3%)	1,073	60.6	1.42	(1.35 to 1.50)	4
Diabetes without insulin therapy	278,785 (3.6%)	1,419	50.9	1.60	(1.53 to 1.68)	5
Insulin therapy	101,996 (1.3%)	973	95.4	2.35	(2.21 to 2.49)	9
Obesity	16,571 (0.2%)	103	62.2	1.34	(1.13 to 1.58)	3
Disorders of fluid, electrolyte, and acid-base balance	8,576 (0.1%)	135	157.4	1.29	(1.11 to 1.49)	3
Gout	164,428 (2.2%)	1,518	92.3	1.57	(1.50 to 1.66)	5
Coagulation defects	3,603 (0.1%)	36	99.9	1.41	(1.07 to 1.85)	3
Anaemias	613,430 (8.0%)	2,228	36.3	1.51	(1.45 to 1.58)	4
Dementia / Alzheimer	12,671 (0.2%)	145	114.4	1.26	(1.09 to 1.46)	2
Psychosis	138,034 (1.8%)	684	49.6	1.94	(1.80 to 2.08)	7
Depression	588,688 (7.7%)	1,729	29.4	1.35	(1.29 to 1.42)	3
Parkinson's disease and parkinsonism	40,885 (0.5%)	274	67.0	1.21	(1.09 to 1.34)	2
Epilepsy and recurrent seizures	122,171 (1.6%)	510	41.7	1.37	(1.26 to 1.48)	3
Other diseases of the nervous system and sense organs	35,495 (0.5%)	253	71.3	1.26	(1.13 to 1.40)	2
Ischaemic Heart Disease/Angina	91,539 (1.2%)	845	92.3	1.18	(1.11 to 1.26)	2
Heart failure	21,840 (0.3%)	428	196.0	1.30	(1.18 to 1.43)	3
Vascular diseases	14,936 (0.2%)	217	145.3	1.17	(1.04 to 1.32)	2
Cerebrovascular diseases	35,205 (0.5%)	333	94.6	1.12	(1.02 to 1.23)	1
Hypertension	796,044 (10.4%)	3,136	39.4	1.20	(1.15 to 1.25)	2

Coronary and peripheral vascular disease	658,737 (8.6%)	2,668	40.5	1.75	(1.68 to 1.82)	6
Oral anticoagulant agents	144,713 (1.9%)	1,221	84.4	1.39	(1.32 to 1.47)	3
COPD/Asthma	20,034 (0.3%)	268	133.8	1.15	(1.03 to 1.28)	1
Liver cirrhosis and other liver chronic diseases	29,484 (0.4%)	177	60.0	1.31	(1.16 to 1.49)	3
Chronic kidney disease	17,109 (0.2%)	371	216.8	1.32	(1.20 to 1.46)	3
Diseases of the skin and subcutaneous tissues	106,747 (1.4%)	353	33.1	1.10	(1.00 to 1.20)	1
Chronic pain	191,442 (2.5%)	1,007	52.6	1.28	(1.21 to 1.36)	2
Corticosteroids	935,246 (12.2%)	2,588	27.7	1.62	(1.55 to 1.68)	5
Individuals without any of the 29 conditions above listed	4,600,012 (60.1%)	1,350	2.9	-	-	-

HIV, Human immunodeficiency virus; COPD, Chronic obstructive pulmonary disease

The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 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. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk

† Odds ratio, and 90% confidence interval, estimated by multivariable logistic regression. Odds ratios 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

Legend of Figures

Figure 1. Relationship between categories of Covid-Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles

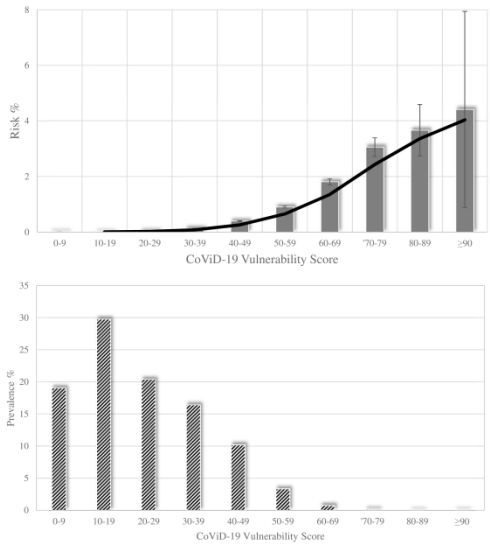
Footnote. The analysis was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), this cohort experienced 9,160 severe (intensive care unit admitted and mechanically ventilated via intubation) and / or fatal outcomes. The average incidence rate during the first wave was therefore 12.0 cases per 10,000 people at risk.

Figure 2. Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of Covid-Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of Covid-Vulnerability Score (CVS) from several validation sets (right box)

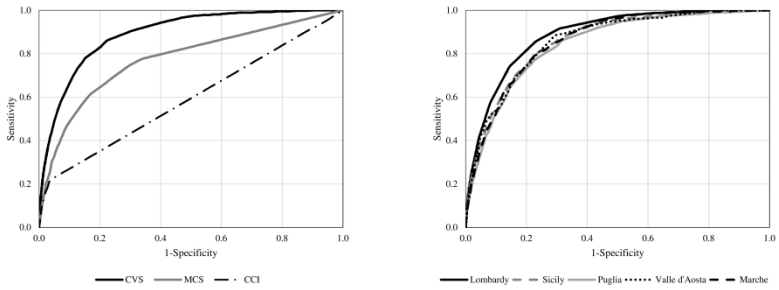
Footnote. Derivation set (left box) was based on the cohort of 7,655,502 beneficiaries of the Lombardy Region Health Service for at least two years, who on February 21, 2020 were alive, aged between 18 and 79 years and did not reside in a nursing home. During the first epidemic wave (until June 2020), 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: (i) 7,575,924 resident in Lombardy whom observation started on July 1, 2020 and who experienced 2,822 severe/fatal outcomes within December 31, 2020; (ii) 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 March 1, 2020 and who respectively experienced 173, 542, 1,953 and 1,541 severe/fatal outcomes within December 31, 2020

Figure 3. Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

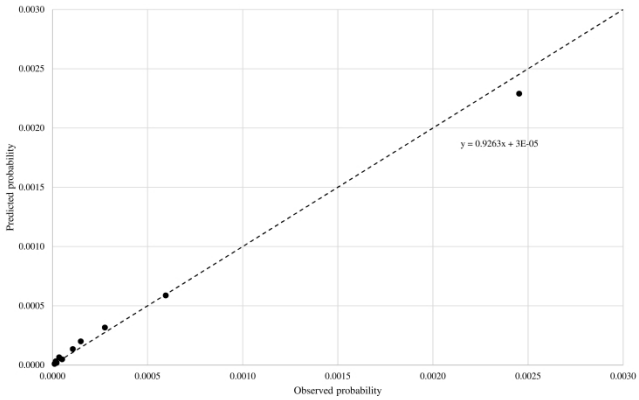
Footnote. The analysis was based on the pooled validation sets of 15,441,033 from Lombardy, Valle d'Aosta, Marche, Puglia and Sicily who experienced 7,031 severe/fatal outcomes rom starting (July 1, 2020 in Lombardy, or March 1, 2020 in the other regions) until December 31, 2020



Relationship between categories of CoViD-19 Vulnerability Score and (i) the risk of occurrence of severe/fatal forms of Covid-19 (upper box), (ii) its distribution among NHS beneficiaries (bottom box). Columns indicate the observed values (of risk and prevalence respectively). Solid and dashed lines respectively represent the fitted cubic spline with the corresponding 5th and 95th percentiles



Receiver Operating Characteristics (ROC) curves comparing discriminant power (i) of CoViD-19 Vulnerability Score (CVS), Charlson Comorbidity Index (CCI) and Multisource Comorbidity Score (MCS) from the derivation set (left box); (ii) of CoViD-19 Vulnerability Score (CVS) from several validation sets (right box)



Calibration plot of observed (X-axis) vs predicted (Y-axis) risk of severe/fatal outcomes

Addressing vaccination priority by stratifying general population according with frailty: a large population-based cohort study

Giovanni Corrao, PhD[1,2], Federico Rea, PhD[1,2], Flavia Carle, PhD[1,3], Salvatore Scondotto, PhD[1,4], Alessandra Allotta, MSc[4], Vito Lepore, MD[5], Antonio D'Ettorre, MSc[5], Cinzia Tanzarella, MSc[5], Patrizia Vittori, MSc[6], Sabrina Abena, MSc[6], Marica Iommi, PhD[3], Liana Spazzafumo, MSc[1,7], Michele Ercolanoni, MSc[8,9], Roberto Blaco, MSc[9], Simona Carbone, MSc[10], Cristina Giordani, MSc[10], Dario Manfellotto, MD[11], Massimo Galli, MD[12,13], Giuseppe Mancia, MD[14,15] on behalf of the "Monitoring and Assessing care Pathways (MAP)" working group of the Italian Ministry of Health

[1] National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy
[2] Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3] Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy
[4] Department of Health Services and Epidemiological Observatory, Regional Health Authority, Sicily Region, Palermo, Italy
[5] Regional Health Agency of Puglia (Agenzia regionale socio-sanitaria), Bari, Italy
[6] Regional Unit of Epidemiology and Social Policies, Regional Health Authority, Valle D'Aosta Region, Aosta, Italy
[7] Regional Health Agency of Marche, Ancona, Italy
[8] ARIA S.p.a., Milan, Italy
[9] Epidemiologic Observatory, Regional Welfare Service, Lombardy Region, Milan, Italy
[10] Department of Health Planning, Italian Health Ministry, Rome, Italy
[11] Department of Internal Medicine, Hospital Fatebenefratelli - AFaR, Isola Tiberina, Rome, Italy
[12] Infectious Diseases Unit, Luigi Sacco Hospital, Milan, Italy
[13] Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy
[14] Emeritus Professor, University of Milano-Bicocca, Milan, Italy
[15] Policlinico di Monza, Monza, Italy

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Selected features of regional populations included into the validation study in comparison with entire Italian population

		Italian 2020 population census†		Indicators of Covid-19 Epidemic Spread (March-December 2020)‡	
Region	Location	Whole population	Population aged 18 – 79 years	Ascertained cases	Deaths
Lombardy	Norther Italy	10,027,602	7,663,864	478,903	25,123
Valle d'Aoste	North Italy	125,034	95,914	7,273	379
Marche	Central Italy	1,512,672	1,150,809	41,624	1,571
Puglia	Southern Italy	3,953,305	3,055,720	90,964	2,472
Sicily	Island	4,875,290	3,744,848	93,644	2,412
	Total	20,493,903	15,711,155	712,408	31,957
	Italy	59,641,488	45,788,626	2,107,166	74,159

† source: http://demo.istat.it/popres/index.php?anno=2020&lingua=ita

‡ source: Protezione Civile. Dati COVID-19 Italia (available at https://github.com/pcm-dpc/COVID-19)

Supplementary Table S2. List of candidate conditions for predicting the risk of experiencing severe fatal forms of COVID-19 infection

Diagnostic categories	#	Disease / condition	ICD-9 CM	ATC
Infectious and parasitic diseases	1	HIV infection	042.x, V08	J05AB14, J05AE, J05AF01, J05AF02, J05AF04, J05AF05, J05AF06, J05AF09. J05AG, J05AR, J05AX07, J05AX08, J05AX09, J05AX12
	2	Tuberculosis and Other infectious and parasitic diseases	010.x - 018.x, 001.x-009.x, 020.x-027.x, 030.x-0.41.x, 045.x–057.x, 060.x-066.x, 070.x-088.x, 090.x-104.x, 110.x-118.x, 120.x-139.x	J04AB
Neoplasms	3	Solid malignancies and Neoplasm of lymphatic and haematopoietic tissue	140.x-165.x, 170.x-176.x, 179.x-199.x, V58.0, 92.2, 200.x-208-x	L01, L03AC, L02BA01, L02BA02, L02BG02, L02BG03, L02BG04, L02BG06, L02BB01, L02BB03, L02AE02, L02AE04, L02AB01
	4	Benign neoplasm and carcinoma in situ	210.x-234.x	
Endocrine, nutritional and metabolic diseases, and immunity disorders	5	Hypothyroidism	243, 244.x	H03A, H03B
	6	Hyper e hypoparathyroidism	252.0, 252.1	
	7	Diabetes without insulin therapy	250.x, 348.0x, 357.2, 362.0, 366.41	A10B
	8	Insulin therapy		A10A
	9	Dyslipidaemia	272.2, 272.4	C10
	10	Obesity	278.0x	
	11	Weight loss	260-263.x	
	12	Disorders of fluid, electrolyte, and acid-base balance	276.x	
	13	Gout	274.x	M04AC01, M04AA, M04AB
	14	Other disorders of endocrine, nutritional and metabolic diseases	240.x-242.x, 245.x, 246.x, 249.x, 251.x, 252.8, 252.9, 253.x-259.x, 264.x-269.x, 270.x, 271.x, 272.0, 272.1, 272.3, 272.5-272.9, 273.x, 275.x, 277.x, 278.1-278.8 (except 277.0)	
	15	Disorders involving the immune mechanisms	279.x	
Diseases of the blood and blood-forming organs	16	Coagulation defects	286.x	B02B
	17	Autoimmune haemolytic anaemias, Other anaemias, Anaemias only tracked from drug therapy	280.x-282.x, 283.1-283.9, 284.x-285.x	B03A, B03B, B03XA01, L03AA
	18	Other diseases of the blood and blood-forming organs	287.x-289.x	
Mental disorders	19	Dementia / Alzheimer	290.0-290.4x, 331.0x	N06DA, N06DX01

	20	Psychosis	295.x, 297.x, 298.2-298.9, 299.1x	N05AD, N05AA, N05AB, N05AC, N05AX, N05AE, N05AF, N05AG N05AH, N05AL
	21	Depression	296.2, 296.3, 296.82, 298.0, 300.4, 301.12, 309.0x, 309.1x, 311.x	N06A
	22	Bipolar disorders	296.0x, 296.1x, 296.4x, 296.5x, 296.6x, 296.7x, 296.80, 296.81, 296.89, 296.9x, 298.1x	N05AN
	23	Alcohol abuse	291.1, 291.2, 291.5, 291.8x, 291.9, 303.9, 305.0x, V11.3x	N07BB01
	24	Drug addition	292.0x, 292.82-292.89, 292.9x, 304.x, 305.2x-305.9x	N07BB04
	25	Anxiety	300.0x	N05BA, N05BB01, N05CD, N05BC01, N05BC51, N05BX, N05CF, N05CX01, N06BX
	26	Other mental disorders	290.8, 290.9, 291.0, 291.3, 291.4, 292.1x, 292.2, 292.81, 293.x, 294.x, 299.0x, 299.8x, 299.9x, 300.0x-300.2x, 300.3, 300.5-300.9, 301.0, 301.10, 301.11, 301.2x-301.9x, 302.x, 303.x, 305.1, 306.x-308.x, 309.2x-309.4x, 310.x, 312.x-319.x	
Diseases of the nervous system and sense organs	27	Parkinson's disease and parkinsonism	332.x	N04
	28	Multiple sclerosis	340	L03AB07, L03AB08, L04AA23, L04AA27, L03AX13, L04AA31, L04AA34, L03AB13, L04AX07
	29	Epilepsy and recurrent seizures	345.x	N03AF01, N03AB02, N03AA02, N03AA03, N03AA04, N03AE01, N03AD01, N03AG01, N05BA09, N03AG04, N03AX10, N03AG06, N03AF02, N03AX14, N03AX15
	30	Glaucoma	365.x	S01E
	31	Disorders of the eye and adnexa	360.x-379.x (except 365.x)	
	32	Diseases of the ear and mastoid process	380.x-389.x	
	33	Other diseases of the nervous system and sense organs	320.x-326.x, 330.x-331.x, 333.x-337.x, 340.x-344.x, 346.x-359.x	
Diseases of the circulatory system	34	Ischaemic Heart Disease/Angina	410.x – 414	C01DA, C01DX
	35	Heart failure	398.91, 402.11, 402.91, 404.11, 404.13, 404.91, 404.93, 428.x	
	36	Arrhythmia	426.10, 426.11, 426.13, 426.20-426.53, 426.60-426.89, 427.0, 427.2,	C01BA, C01BC, C01BD

			427.31, 427.60,427.9, 785.0x, V45.0x, V53.3x	
	37	Valvular diseases	093.20-093.24, 394.0x-397.1x, 424.00-424.91, 746.3x-746.6x, V42.2x, V43.3x	
	38	Vascular diseases	440.x, 441.2, 441.4, 441.7, 441.9, 443.1x-443.9x, 447.1, 557.1x, 557.9x, 785.4x, V43.4x	
	39	Cerebrovascular diseases	430.x-438.x	
	40	Hypertension	401.x-405.x	C03AA, C03AB, C03AH, C03AX01, C02CA04, C03BA02, C03BA03, C03BA04, C03BA05, C03BA07, C03BA08, C03BA09, C03BA10, C03BA11, C03DB01, C03DB02, C03EA, C09BA02, C09BA03, C09BA04, C09BA05, C09BA06, C09BA07, C09BA08, C09BA09, C09BB, C09DB, C09DA01, C09DA02, C09DA03, C09DA04, C09DA06, C09DA07, C09DA08, C02AB01, C02AB02, C02AC01, C02AC02, C02AC04, C02AC05, C02DB02, C02DB03, C02DB04, C02DC01, C02DD01, C02DG01, C02KA01, C02KB01, C02KC01, C02KD01, C02KX01, C09XA
	41	Coronary and peripheral vascular disease		B01AB, B01AX01, B01AD10, B01AD12, C04AD03, B01AC05
	42	Oral anticoagulant agents		B01AA, B01AE, B01AF
	43	Other diseases of the circulatory system	390.x-392.x, 393, 397.9, 398.90, 398.99, 411.8x, 412.x-417x, 420.x-423.x, 424.99, 425.x, 426.0, 426.12, 426.54, 426.9, 427.1, 427.32, 427.4x, 427.5, 427.61, 427.69, 427.8x, 429.x, 441.0x, 441.1, 441.3, 441.5, 441.6, 442.x, 443.0, 444.x-446.x, 447.0, 447.2-447.9, 448.x 451.x-459.x	
Diseases of the respiratory system	44	Chronic Obstructive Pulmonary Disease, Asthma, Chronic respiratory disease only tracked from drug therapy	490-492.x, 493.x, 494.x, 496	R03AA, R03AB, R03AC, R03DA, R03DB, R03DA20, R01AC01, R03BC01, R01AC51, S01GX01, S01GX51, R03BA

	45	Acute respiratory infections	460-466.x	
	46	Cystic Fibrosis	277.0	R05CB, R05FB01, R05FA01, A09AA02, R07AX02, R07AX30, R07AX31
	47	Other diseases of the respiratory system	470.x-478.x, 480.x-487.x, 495.x, 500.x-508.x, 510.x-519.x	
Diseases of the digestive system	48	Liver cirrhosis and other liver chronic diseases	571.x, 573.x	J05AP08, J05AP09, J05AP51, J05AP53, J05AP54, J05AP55, J05AP56, J05AP57, B05AA01
	49	Inflammatory bowel diseases (Ulcerative colitis and Chron's disease)	555.x-556.x	A07EC01, A07EC02, A07EC03, A07EC04
	50	Chronic and acute pancreatitis	577.0-577.1	
	51	Other diseases of the digestive system	520.x-553.x, 557.x-570, 572.x, 574.x-576.x, 577.2-577.9, 578.x, 579.x	
Diseases of the genitourinary system	52	Chronic kidney disease	585, V45.1, V56.x, V03AE	
	53	Other kidney disorders	580.x-584.x, 586, 587, 588.x-589.x	
	54	Other diseases of the genitourinary system	590.x-608.x, 610.x, 611.x, 614.x-629.x	
Diseases of the skin and subcutaneous tissues	55	Diseases of the skin and subcutaneous tissues, including No rheumatoid psoriasis	680.x-686.x, 690.x-695.x, 696.0, 696.2-696.5, 696.8, 697.x, 698.x, 700.x-709.x, 696.1	D05BB01, D05BB02, D05AX
Diseases of the musculoskeletal system and connective tissue	56	Autoimmune disease (Rheumatoid arthritis, Rheumatoid psoriasis, Anchylosing spondylitis, Systemic sclerosis, Systemic lupus erythematosus)	714.0, 696.0, 720.0, 710.1x, 710.0x	
	57	Other diseases of the musculoskeletal system and connective tissue	710.2-710.9, 711.x-713.x, 714.1x, 714.9x, 715.x-719.x, 720.1x-720.9x, 721.x-739.x	
Symptoms, signs and ill-defined conditions	58	Symptoms, signs and ill-defined conditions	780-799	
Other conditions	59	Transplantation	V42	L04AA01, L04AA02, L04AA03, L04AA04, L04AA05, L04AA06, L04AA08, L04AA09, L04AA10, L04AA11, L04AA12, L04AA14, L04AA15, L04AA16, L04AA17, L04AA18, L04AA19, L04AA21, L04AD01, L04AD02, L04AX01

60	Chronic pain	338.2, 338.4	N02AA01, N02AG01, N02AE01, N02AB03, N02AA05, N02AA55, N02AA03, N02AX06
61	Corticosteroids		H02

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Page 1-5
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			7
Objectives	3	State specific objectives, including any prespecified hypotheses			8
Methods					
Study Design	4	Present key elements of study design early in the paper			9-11
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			8

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	8-11
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	9
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			9

Bias	9	Describe any efforts to address potential sources of bias			9-11
Study size	10	Explain how the study size was arrived at			8-11
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			9,10
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			9-11
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	8-11

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	8,9
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	11,12
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			11,12
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			11,12

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			11,12
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			11,12
Discussion					
Key results	18	Summarise key results with reference to study objectives			12,13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	15,16
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			13-15

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			15
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			17
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	18

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.